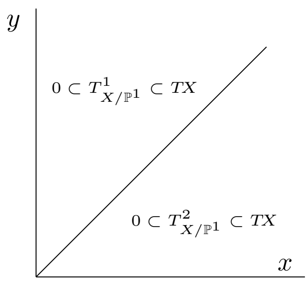
Documenta Mathematica

Journal der
Deutschen Mathematiker-Vereinigung
Gegründet 1996



See page 163

Band 14 · 2009

DOCUMENTA MATHEMATICA, Journal der Deutschen Mathematiker-Vereinigung, veröffentlicht Forschungsarbeiten aus allen mathematischen Gebieten und wird in traditioneller Weise referiert. Es wird indiziert durch Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

Artikel können als TEX-Dateien per E-Mail bei einem der Herausgeber eingereicht werden. Hinweise für die Vorbereitung der Artikel können unter der unten angegebenen WWW-Adresse gefunden werden.

DOCUMENTA MATHEMATICA, Journal der Deutschen Mathematiker-Vereinigung, publishes research manuscripts out of all mathematical fields and is refereed in the traditional manner. It is indexed in Mathematical Reviews, Science Citation Index Expanded, Zentralblatt für Mathematik.

Manuscripts should be submitted as TEX -files by e-mail to one of the editors. Hints for manuscript preparation can be found under the following web address.

`http://www.math.uni-bielefeld.de/documenta`

ISSN 1431-0635 (Print), ISSN 1431-0643 (Internet)

SPARC
LEADING EDGE

DOCUMENTA MATHEMATICA is a Leading Edge Partner of SPARC, the Scholarly Publishing and Academic Resource Coalition of the Association of Research Libraries (ARL), Washington DC, USA.

Address of Technical Managing Editor: Ulf Rehmann, Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Typesetting in TEX.

Documenta Mathematica

Band 14, 2009

Chow–Künneth Decomposition for Some Moduli Spaces[1]

Jaya NN Iyer, Stefan Müller–Stach

Received: February 2, 2008
Revised: July 7, 2008

Communicated by Thomas Peternell

Abstract. In this paper we investigate Murre's conjecture on the Chow–Künneth decomposition for universal families of smooth curves over spaces which dominate the moduli space $\mathcal{M}_g$, in genus at most 8 and show existence of a Chow–Künneth decomposition. This is done in the setting of equivariant cohomology and equivariant Chow groups to get equivariant Chow–Künneth decompositions.

2000 Mathematics Subject Classification: 14C25, 14D05, 14D20, 14D21
Keywords and Phrases: Equivariant Chow groups, orthogonal projectors.

Contents

[1]This work was supported by DFG Sonderforschungsbereich/Transregio 45.

1. Introduction

Suppose X is a nonsingular projective variety defined over the complex numbers. We consider the rational Chow group $CH^i(X)_{\mathbb{Q}} = CH^i(X) \otimes \mathbb{Q}$ of algebraic cycles of codimension i on X. The conjectures of S. Bloch and A. Beilinson predict a finite descending filtration $\{F^jCH^i(X)_{\mathbb{Q}}\}$ on $CH^i(X)_{\mathbb{Q}}$ and satisfying certain compatibility conditions. A candidate for such a filtration has been proposed by J. Murre and he has made the following conjecture [Mu2],
Murre's conjecture: The motive (X, Δ) of X has a Chow-Künneth decomposition:

$$\Delta = \sum_{i=0}^{2d} \pi_i \in CH^d(X \times X) \otimes \mathbb{Q}$$

such that π_i are orthogonal projectors, lifting the Künneth projectors in $H^{2d-i}(X) \otimes H^i(X)$. Furthermore, these algebraic projectors act trivially on the rational Chow groups in a certain range.
These projectors give a candidate for a filtration of the rational Chow groups, see §2.1.
This conjecture is known to be true for curves, surfaces and a product of a curve and surface [Mu1], [Mu3]. A variety X is known to have a Chow–Künneth decomposition if X is an abelian variety/scheme [Sh],[De-Mu], a uniruled threefold [dA-Mü1], universal families over modular varieties [Go-Mu], [GHM2] and the universal family over one Picard modular surface [MMWYK], where a partial set of projectors are found. Finite group quotients (maybe singular) of an abelian variety also satisfy the above conjecture [Ak-Jo]. Furthermore, for some varieties with a nef tangent bundle, Murre's conjecture is proved in [Iy]. A criterion for existence of such a decomposition is also given in [Sa]. Some other examples are also listed in [Gu-Pe].
Gordon-Murre-Hanamura [GHM2], [Go-Mu] obtained Chow–Künneth projectors for universal families over modular varieties. Hence it is natural to ask if the universal families over the moduli space of curves of higher genus also admit a Chow–Künneth decomposition. In this paper, we investigate the existence of Chow–Künneth decomposition for families of smooth curves over spaces which closely approximate the moduli spaces of curves $\mathcal{M}_g$ of genus at most 8, see §5.
In this example, we take into account the non-trivial action of a linear algebraic group G acting on the spaces. This gives rise to the equivariant cohomology and equivariant Chow groups, which were introduced and studied by Borel, Totaro, Edidin-Graham [Bo], [To], [Ed-Gr]. Hence it seems natural to formulate Murre's conjecture with respect to the cycle class maps between the rational equivariant Chow groups and the rational equivariant cohomology, see §4.5. Since in concrete examples, good quotients of non-compact varieties exist, it became necessary to extend Murre's conjecture for non-compact smooth varieties, by taking only the bottom weight cohomology $W_iH^i(X, \mathbb{Q})$ (see [D]), into consideration. This is weaker than the formulation done in [BE]. For our purpose though, it suffices to look at this weaker formulation. We then

construct a category of equivariant Chow motives, fixing an algebraic group G (see [dB-Az], [Ak-Jo], for a category of motives of quotient varieties, under a finite group action).
With this formalism, we show (see §5.2);

Theorem 1.1. *The equivariant Chow motive of a universal family of smooth curves $\mathcal{X} \to U$ over spaces U which dominate the moduli space of curves $\mathcal{M}_g$, for $g \leq 8$, admits an equivariant Chow–Künneth decomposition, for a suitable linear algebraic group G acting non-trivially on $\mathcal{X}$.*

Whenever smooth good quotients exist under the action of G, then the equivariant Chow-Künneth projectors actually correspond to the absolute Chow–Künneth projectors for the quotient varieties. In this way, we get orthogonal projectors for universal families over spaces which closely approximate the moduli spaces $\mathcal{M}_g$, when g is at most 8.
One would like to try to prove a Chow–Künneth decomposition for $\mathcal{M}_g$ and $\mathcal{M}_{g,n}$ (which parametrizes curves with marked points) and we consider our work a step forward. However since we only work on an open set U one has to refine projectors after taking closures a bit in a way we don't yet know.
Other examples that admit a Chow–Künneth decomposition are Fano varieties of r-dimensional planes contained in a general complete intersection in a projective space, see Corollary 5.3.
The proofs involve classification of curves in genus at most 8 by Mukai [Muk],[Muk2] with respect to embeddings as complete intersections in homogeneous spaces. This allows us to use Lefschetz theorem and construct orthogonal projectors.
Acknowledgements: The first named author thanks the Math Department of Mainz, for its hospitality during the visits in 2007 and 2008, when this work was carried out. We also thank a referee for a useful remark concerning our definition of the weight filtration.

2. Preliminaries

The category of nonsingular projective varieties over $\mathbb{C}$ will be denoted by $\mathcal{V}$. Let $CH^i(X)_{\mathbb{Q}} = CH^i(X) \otimes \mathbb{Q}$ denote the rational Chow group of codimension i algebraic cycles modulo rational equivalence.
Suppose $X, Y \in Ob(\mathcal{V})$ and $X = \cup X_i$ be a decomposition into connected components X_i and $d_i = \dim X_i$. Then $\mathrm{Corr}^r(X,Y) = \oplus_i CH^{d_i+r}(X_i \times Y)_{\mathbb{Q}}$ is called a space of correspondences of degree r from X to Y.
A category $\mathcal{M}$ of Chow motives is constructed in [Mu2]. Suppose X is a nonsingular projective variety over $\mathbb{C}$ of dimension d. Let $\Delta \subset X \times X$ be the diagonal. Consider the Künneth decomposition of the class of Δ in the Betti Cohomology:

$$[\Delta] = \oplus_{i=0}^{2d} \pi_i^{hom}$$

where $\pi_i^{hom} \in H^{2d-i}(X,\mathbb{Q}) \otimes H^i(X,\mathbb{Q})$.

Definition 2.1. *The motive of X is said to have Künneth decomposition if each of the classes π_i^{hom} is algebraic, i.e., π_i^{hom} is the image of an algebraic* *cycle π_i under the cycle class map from the rational Chow groups to the Betti cohomology.*

Definition 2.2. *The motive of X is said to have a Chow–Künneth decomposition if each of the classes π_i^{hom} is algebraic and they are orthogonal projectors, i.e., $\pi_i \circ \pi_j = \delta_{i,j}\pi_i$.*

Lemma 2.3. *If X and Y have a Chow–Künneth decomposition then $X \times Y$ also has a Chow–Künneth decomposition.*

Proof. If π_i^X and π_j^Y are the Chow–Künneth components for $h(X)$ and $h(Y)$ respectively then

$$\pi_i^{X\times Y} = \sum_{p+q=i} \pi_p^X \times \pi_q^Y \ \in CH^*(X \times Y \times X \times Y)_{\mathbb{Q}}$$

are the Chow–Künneth components for $X \times Y$. Here the product $\pi_p^X \times \pi_q^Y$ is taken after identifying $X \times Y \times X \times Y \simeq X \times X \times Y \times Y$. □

2.1. Murre's conjectures. J. Murre [Mu2], [Mu3] has made the following conjectures for any smooth projective variety X.
(A) The motive $h(X) := (X, \Delta_X)$ of X has a Chow-Künneth decomposition:

$$\Delta_X = \sum_{i=0}^{2n} \pi_i \in CH^n(X \times X) \otimes \mathbb{Q}$$

such that π_i are orthogonal projectors.
(B) The correspondences $\pi_0, \pi_1, ..., \pi_{j-1}, \pi_{2j+1}, ..., \pi_{2n}$ act as zero on $CH^j(X)\otimes \mathbb{Q}$.
(C) Suppose

$$F^r CH^j(X) \otimes \mathbb{Q} = \mathrm{Ker}\pi_{2j} \cap \mathrm{Ker}\pi_{\mathrm{2j-1}} \cap ... \cap \mathrm{Ker}\pi_{\mathrm{2j-r+1}}.$$

Then the filtration $F^\bullet$ of $CH^j(X) \otimes \mathbb{Q}$ is independent of the choice of the projectors π_i.
(D) Further, $F^1 CH^i(X) \otimes \mathbb{Q} = (CH^i(X) \otimes \mathbb{Q})_{hom}$, the cycles which are homologous to zero.

In §4, we will extend (A) in the setting of equivariant Chow groups.

3. Equivariant Chow groups and equivariant Chow motives

In this section, we recall some preliminary facts on the equivariant groups to formulate Murre's conjectures for a smooth variety X of dimension d, which is equipped with an action by a linear reductive algebraic group G. The equivariant groups and their properties that we recall below were defined by Borel, Totaro, Edidin-Graham, Fulton [Bo],[To],[Ed-Gr], [Fu2].

3.1. Equivariant cohomology $H^i_G(X,\mathbb{Z})$ of X. Suppose X is a variety with an action on the left by an algebraic group G. Borel defined the equivariant cohomology $H^*_G(X)$ as follows. There is a contractible space EG on which G acts freely (on the right) with quotient $BG := EG/G$. Then form the space

$$EG\times_G X := EG\times X/(e.g,x)\sim(e,g.x).$$

In other words, $EG\times_G X$ represents the (topological) quotient stack $[X/G]$.

Definition 3.1. *The equivariant cohomology of X with respect to G is the ordinary singular cohomology of $EG\times_G X$:*

$$H^i_G(X) = H^i(EG\times_G X).$$

For the special case when X is a point, we have

$$H^i_G(point) = H^i(BG)$$

For any X, the map $X\to point$ induces a pullback map $H^i(BG)\to H^i_G(X)$. Hence the equivariant cohomology of X has the structure of a $H^i(BG)$-algebra, at least when $H^i(BG)=0$ for odd i.

3.2. Equivariant Chow groups $CH^i_G(X)$ of X. [Ed-Gr]
As in the previous subsection, let X be a smooth variety of dimension n, equipped with a left G-action. Here G is an affine algebraic group of dimension g. Choose an l-dimensional representation V of G such that V has an open subset U on which G acts freely and whose complement has codimension more than $n-i$. The diagonal action on $X\times U$ is also free, so there is a quotient in the category of algebraic spaces. Denote this quotient by $X_G := (X\times U)/G$.

Definition 3.2. *The i-th equivariant Chow group $CH^G_i(X)$ is the usual Chow group $CH_{i+l-g}(X_G)$. The codimension i equivariant Chow group $CH^i_G(X)$ is the usual codimension i Chow group $CH^i(X_G)$.*

Note that if X has pure dimension n then

$$\begin{aligned} CH^i_G(X) &= CH^i(X_G)\\ &= CH_{n+l-g-i}(X_G)\\ &= CH^G_{n-i}(X). \end{aligned}$$

Proposition 3.3. *The equivariant Chow group $CH^G_i(X)$ is independent of the representation V, as long as $V-U$ has codimension more than $n-i$.*

Proof. See [Ed-Gr, Definition-Proposition 1]. □

If $Y\subset X$ is an m-dimensional subvariety which is invariant under the G-action, and compatible with the G-action on X, then it has a G-equivariant fundamental class $[Y]_G\in CH^G_m(X)$. Indeed, we can consider the product $(Y\times U)\subset X\times U$, where U is as above and the corresponding quotient $(Y\times U)/G$ canonically embeds into X_G. The fundamental class of $(Y\times U)/G$ defines the class $[Y]_G\in CH^G_m(X)$. More generally, if V is an l-dimensional representation of G and $S \subset X \times V$ is an $m+l$-dimensional subvariety which is invariant under the G-action, then the quotient $(S \cap (X \times U))/G \subset (X \times U)/G$ defines the G-equivariant fundamental class $[S]_G \in CH^G_m(X)$ of S.

Proposition 3.4. *If $\alpha \in CH^G_m(X)$ then there exists a representation V such that $\alpha = \sum a_i [S_i]_G$, for some G-invariant subvarieties S_i of $X \times V$.*

Proof. See [Ed-Gr, Proposition 1]. □

3.3. Functoriality properties. Suppose $f : X \to Y$ is a G-equivariant morphism. Let $\mathcal{S}$ be one of the following properties of schemes or algebraic spaces: proper, flat, smooth, regular embedding or l.c.i.

Proposition 3.5. *If $f : X \to Y$ has property $\mathcal{S}$, then the induced map $f_G : X_G \to Y_G$ also has property $\mathcal{S}$.*

Proof. See [Ed-Gr, Proposition 2]. □

Proposition 3.6. *Equivariant Chow groups have the same functoriality as ordinary Chow groups for equivariant morphisms with property $\mathcal{S}$.*

Proof. See [Ed-Gr, Proposition 3]. □

If X and Y have G-actions then there are exterior products

$$CH^G_i(X) \otimes CH^G_j(Y) \to CH^G_{i+j}(X \times Y).$$

In particular, if X is smooth then there is an intersection product on the equivariant Chow groups which makes $\oplus_j CH^G_j(X)$ into a graded ring.

3.4. Cycle class maps. [Ed-Gr, §2.8]
Suppose X is a complex algebraic variety and G is a complex algebraic group. The equivariant Borel-Moore homology $H^G_{BM,i}(X)$ is the Borel-Moore homology $H_{BM,i}(X_G)$, for $X_G = X \times_G U$. This is independent of the representation as long as $V - U$ has sufficiently large codimension. This gives a cycle class map,

$$cl_i : CH^G_i(X) \to H^G_{BM,2i}(X, \mathbb{Z})$$

compatible with usual operations on equivariant Chow groups. Suppose X is smooth of dimension d then X_G is also smooth. In this case the Borel-Moore cohomology $H^G_{BM,2i}(X, \mathbb{Z})$ is dual to $H^{2d-i}(X_G) = H^{2d-i}(X \times_G U)$.
This gives the cycle class maps

$$cl^i : CH^i_G(X) \to H^{2i}_G(X, \mathbb{Z}). \tag{1}$$

There are also maps from the equivariant groups to the usual groups:

$$H^i_G(X, \mathbb{Z}) \to H^i(X, \mathbb{Z}) \tag{2}$$

and

$$CH^i_G(X) \to CH^i(X). \tag{3}$$

3.5. Weight filtration $W_{\cdot}$ on $H^i_G(X,\mathbb{Z})$. In this paper, we assign only the bottom weight W_i of the equivariant cohomology in the simplest situation. Consider a smooth variety X equipped with a left G action as above.
We can define

$$W_iH^i_G(X,\mathbb{Q}) := W_iH^i((X\times U)/G,\mathbb{Q}),$$

for $U\subset V$ an open subset with a free G-action, where codim $V-U$ is at least $n-i$.

Lemma 3.7. *The group $W_iH^i_G(X,\mathbb{Q})$ is independent of the choice of the G-representation V as long as codim $V-U$ is at least $n-i$.*

Proof. The proof of independence of V in the case of equivariant Chow groups [Ed-Gr, Definition-Proposition 1] applies directly in the case of the bottom weight equivariant cohomology. □

3.6. Equivariant Chow motives and the category of equivariant Chow motives. When G is a finite group then a category of Chow motives for (maybe singular) quotients of varieties under the G-action was constructed in [dB-Az], [Ak-Jo]. More generally, we consider the following situation, taking into account the equivariant cohomology and the equivariant rational Chow groups, which does not seem to have been considered before.
Fix an affine complex algebraic group G. Let $\mathcal{V}_G$ be the category whose objects are complex smooth projective varieties with a G-action and the morphisms are G-equivariant morphisms.
For any $X,Y,Z\in Ob(\mathcal{V}_G)$, consider the projections

$$\begin{aligned} X\times Y\times Z &\xrightarrow{p_{XY}} X\times Y,\\ X\times Y\times Z &\xrightarrow{p_{YZ}} Y\times Z,\\ X\times Y\times Z &\xrightarrow{p_{XZ}} X\times Z. \end{aligned}$$

which are G-equivariant.
Let d be the dimension of X. The group of correspondences from X to Y of degree r is defined as

$$\mathrm{Corr}^r_G(X\times Y) := CH^{r+d}_G(X\times Y).$$

Every G-equivariant morphism $X\to Y$ defines an element in $\mathrm{Corr}^0_G(X\times Y)$, by taking the graph cycle.
For any $f\in\mathrm{Corr}^r_G(X,Y)$ and $g\in\mathrm{Corr}^e_G(Y,Z)$ define their composition

$$g\circ f\in\mathrm{Corr}^{r+e}_G(X,Z)$$

by the prescription

$$g\circ f = p_{XZ*}(p^*_{XY}(f).p^*_{YZ}(g)).$$

This gives a linear action of correspondences on the equivariant Chow groups

$$\mathrm{Corr}^r_G(X,Y)\times CH^s_G(X)_{\mathbb{Q}} \longrightarrow CH^{r+s}_G(Y)_Q$$

$$(\gamma,\alpha)\mapsto p_{Y*}(p^*_X\alpha.\gamma)$$

for the projections $p_X: X\times Y\longrightarrow X$, $p_Y: X\times Y\longrightarrow Y$.

The category of pure equivariant G-motives with rational coefficients is denoted by $\mathcal{M}_G^+$. The objects of $\mathcal{M}_G^+$ are triples $(X,p,m)_G$, for $X \in Ob(\mathcal{V}_G)$, $p \in \mathrm{Corr}_G^0(X,X)$ is a projector, i.e., $p \circ p = p$ and $m \in \mathbb{Z}$. The morphisms between the objects $(X,p,m)_G, (Y,q,n)_G$ in $\mathcal{M}_G^+$ are given by the correspondences $f \in \mathrm{Corr}_G^{n-m}(X,Y)$ such that $f \circ p = q \circ f = f$. The composition of the morphisms is the composition of correspondences. This category is pseudoabelian and $\mathbb{Q}$-linear [Mu2]. Furthermore, it is a tensor category defined by

$$(X,p,m)_G \otimes (Y,q,n)_G = (X \times Y, p \otimes q, m+n)_G.$$

The object $(\mathrm{Spec}\,\mathbb{C}, id, 0)_G$ is the unit object and the Lefschetz motive $\mathbb{L}$ is the object $(\mathrm{Spec}\,\mathbb{C}, id, -1)_G$. Here $\mathrm{Spec}\,\mathbb{C}$ is taken with a trivial G-action. The Tate twist of a G-motive M is $M(r) := M \otimes \mathbb{L}^{\otimes -r} = (X,p,m+r)_G$.

Definition 3.8. *The theory of equivariant Chow motives (*[Sc]*) provides a functor*

$$h : \mathcal{V}_G \longrightarrow \mathcal{M}_G^+.$$

For each $X \in Ob(\mathcal{V}_G)$ the object $h(X) = (X, \Delta, 0)_G$ is called the equivariant Chow motive of X. Here Δ is the class of the diagonal in $CH^*(X \times X)_{\mathbb{Q}}$, which is G-invariant for the diagonal action on $X \times X$ and hence lies in $Corr_G^0(X,X) = CH_G^*(X \times X)_{\mathbb{Q}}$.

4. Murre's conjectures for the equivariant Chow motives

Suppose X is a complex smooth variety of dimension d, equipped with a G-action. Consider the product variety $X \times X$ together with the diagonal action of the group G.

The cycle class map

$$cl^d : CH^d(X \times X)_{\mathbb{Q}} \to H^{2d}(X \times X, \mathbb{Q}). \tag{4}$$

actually maps to the weight $2d$ piece $W_{2d}H^{2d}(X \times X, \mathbb{Q})$ of the ordinary cohomology group.

Applying this to the spaces $X \times U$, for open subset $U \subset V$ as in §3.2, (4) holds for the equivariant groups as well and there are cycle class maps:

$$cl^d : CH_G^d(X \times X)_{\mathbb{Q}} \to W_{2d}H_G^{2d}(X \times X, \mathbb{Q}). \tag{5}$$

Lemma 4.1. *The image of the diagonal cycle $[\Delta_X]$ under the cycle class map cl^d lies in the subspace*

$$\bigoplus_i W_{2d-i}H_G^{2d-i}(X) \otimes W_i H_G^i(X)$$

of $W_{2d}H_G^{2d}(X \times X, \mathbb{Q})$.

Proof. First we prove the assertion for the ordinary cohomology of non-compact smooth varieties and next apply it to the product spaces $X \times U$, which is equipped with a free G-action and the quotient space X_G.

If X is a compact smooth variety then we notice that the weight $2d$ piece coincides with the cohomology group $H^{2d}(X \times X, \mathbb{Q})$ and by the Künneth formula for products the statement follows in the usual cohomology. Suppose X is not compact. Using (4), notice that the image of the diagonal cycle $[\Delta_X]$ lies in $W_{2d}H^{2d}(X \times X, \mathbb{Q})$. Choose a smooth compactification $\overline{X}$ of X and consider the commutative diagram:

$$\begin{array}{ccc} \bigoplus_i H^{2d-i}(\overline{X}) \otimes H^i(\overline{X}) & \stackrel{\simeq}{\to} & H^{2d}(\overline{X} \times \overline{X}, \mathbb{Q}) \\ \downarrow & & \downarrow \\ \bigoplus_i W_{2d-i}H^{2d-i}(X) \otimes W_i H^i(X) & \stackrel{k}{\to} & W_{2d}H^{2d}(X \times X, \mathbb{Q}). \end{array}$$

The vertical arrows are surjective maps, defined by the localization. Hence the map k is surjective. The injectivity follows because this is the Künneth product map, restricted to the bottom weight cohomology. This shows that k is an isomorphism.

In particular, the isomorphism k can be applied to the bottom weights of the ordinary cohomology groups of the smooth variety $X \times U$, for any open subset $U \subset V$ of large complementary codimension and V is a G-representation. But this is essentially the bottom weight of the equivariant cohomology group of X. To conclude, we need to observe that the diagonal cycle $[\Delta_X]$ is G-invariant. □

Denote the decomposition of the G-invariant diagonal cycle

$$\Delta_X = \oplus_{i=0}^{2d} \pi_i^G \in W_{2d}H_G^{2d}(X \times X, \mathbb{Q}) \tag{6}$$

such that π_i^G lies in the space $W_{2d-i}H_G^{2d-i}(X) \otimes W_i H_G^i(X)$.

We defined the equivariant Chow motive of a smooth projective variety with a G-action in §3.6. We extend the notion of orthogonal projectors on a smooth variety equipped with a G-action, as follows.

Definition 4.2. *Suppose X is a smooth variety equipped with a G-action. The equivariant Chow motive $(X, \Delta_X)_G$ of X is said to have an* equivariant Künneth decomposition *if the classes π_i^G are algebraic, i.e., they have a lift in the equivariant Chow group $CH_G^d(X \times X)_{\mathbb{Q}}$. Furthermore, if X admits a smooth compactification $X \subset \overline{X}$ such that the action of G extends on $\overline{X}$ and the Künneth projectors extend to orthogonal projectors on $\overline{X}$ then we say that X has an* equivariant Chow–Künneth decomposition.

Remark 4.3. *When G is a linear algebraic group, using the results of Sumihiro* [Su]*, Bierstone-Milman* [Bi-Mi, Theorem 13.2]*, Reichstein-Youssin* [Re-Yo]*, one can always choose a smooth compactification $\overline{X} \supset X$ such that action of G extends to $\overline{X}$. Since any affine algebraic group is linear, we can always find smooth G-equivariant compactifications in our set-up.*

Suppose X is a smooth variety with a free G-action so that we can form the quotient variety $Y := X/G$. Using [Ed-Gr], we have the identification of the rational Chow groups

$$CH^*(Y)_{\mathbb{Q}} = CH^*_G(X)_{\mathbb{Q}}$$

and

$$CH^*(Y \times Y)_{\mathbb{Q}} = CH^*_G(X \times X)_{\mathbb{Q}}.$$

Furthermore, these identifications respect the ring structure on the above rational Chow groups. A similar identification also holds for the rational cohomology groups. In view of this, we make the following definition.

Definition 4.4. *Suppose X is a smooth variety with a G-action and G acts freely on X. Denote the quotient space $Y := X/G$. The absolute Chow–Künneth decomposition of Y is defined to be the equivariant Chow–Künneth decomposition of X.*

We can now extend Murre's conjecture to smooth varieties with a G-action, as follows.

Conjecture 4.5. *Suppose X is a smooth variety with a G-action. Then X has an equivariant Chow–Künneth decomposition.*

In particular, if the action of G is trivial then we can extend Murre's conjecture to a (not necessarily compact) smooth variety, by taking only the bottom weight cohomology $W_iH^i(X)$ of the ordinary cohomology. This is weaker than obtaining projectors for the ordinary cohomology. We remark a projector π_1 in the case of quasi–projective varieties has been constructed by Bloch and Esnault [BE].

5. Families of curves

Our goal in this paper is to find an (explicit) absolute Chow–Künneth decomposition for the universal families of curves over close approximations of the moduli space of smooth curves of small genus. We begin with the following situation which motivates the statements on universal curves.

Lemma 5.1. *Any smooth hypersurface $X \subset \mathbb{P}^n$ of degree d has an absolute Chow–Künneth decomposition. If $L \subset X$ is any line, then the blow-up $X' \to X$ also has a Chow–Künneth decomposition.*

Proof. Notice that the cohomology of X is algebraic except in the middle dimension $H^{n-1}(X, \mathbb{Q})$. By the Lefschetz Hyperplane section theorem, the algebraic cohomology $H^{2j}(X, Q)$, $j \neq n-1$, is generated by the hyperplane section H^j. So the projectors are simply

$$\pi_r := \frac{1}{d}.H^{n-1-r} \times H^r \in CH^{n-1}(X \times X)_{\mathbb{Q}}$$

for $r \neq n-1$. We can now take $\pi_{n-1} := \Delta_X - \sum_{r, r\neq n-1} \pi_r$. This gives a complete set of orthogonal projectors and a Chow–Künneth decomposition for X. Since $X' \to X$ is a blow-up along a line, the new cohomology is again algebraic, by the blow-up formula. Similarly we get a Chow–Künneth decomposition for X' (see also [dA-Mü2, Lemma 2] for blow-ups). □

The above lemma can be generalized to the following situation.

Lemma 5.2. *Suppose Y is a smooth projective variety of dimension r over $\mathbb{C}$ which has only algebraic cohomology groups $H^i(Y)$ for all $0 \leq i \leq m$ for some $m < r$. Then we can construct orthogonal projectors*

$$\pi_0, \pi_1, ..., \pi_m, \pi_{2r-m}, \pi_{2r-m+1}, ..., \pi_{2r}$$

in the usual Chow group $CH^r(Y \times Y)_{\mathbb{Q}}$, and where π_{2i} acts as $\delta_{i,p}$ on $H^{2p}(Y)$ and $\pi_{2i-1} = 0$. Moreover, if there is an affine complex algebraic group G acting on Y, then we can lift the above projectors in the equivariant Chow group $CH^r_G(Y \times Y)_{\mathbb{Q}}$ as orthogonal projectors.

Proof. See also [dA-Mü1, dA-Mü2]. Let $H^{2p}(Y)$ be generated by cohomology classes of cycles $C_1, \ldots, C_s$ and $H^{2r-2p}(Y)$ be generated by cohomology classes of cycles $D_1, \ldots, D_s$. We denote by M the intersection matrix with entries

$$M_{ij} = C_i \cdot D_j \in \mathbb{Z}.$$

After base change and passing to $\mathbb{Q}$–coefficients we may assume that M is diagonal, since the cup–product $H^{2p}(Y, \mathbb{Q}) \otimes H^{2r-2p}(Y, \mathbb{Q}) \to \mathbb{Q}$ is non–degenerate. We define the projector π_{2p} as

$$\pi_{2p} = \sum_{k=1}^{s} \frac{1}{M_{kk}} D_k \times C_k.$$

It is easy to check that $\pi_{2p\,*}(C_k) = D_k$. Define π_{2r-2p} as the adjoint, i.e., transpose of π_{2p}. Via the Gram–Schmidt process from linear algebra we can successively make all projectors orthogonal. $\square$

Suppose $X \subset \mathbb{P}^n$ is a smooth complete intersection of multidegree $d_1 \leq d_2 \leq ... \leq d_s$. Let $F_r(X)$ be the variety of r-dimensional planes contained in X. Let $\delta := \min\{(r+1)(n-r) - \binom{d+r}{r}, n - 2r - s\}$.

Corollary 5.3. *If X is general then $F_r(X)$ is a smooth projective variety of dimension δ and it has an absolute Chow–Künneth decomposition.*

Proof. The first assertion on the smoothness of the variety $F_r(X)$ is well–known, see [Al-Kl], [ELV], [De-Ma]. For the second assertion, notice that $F_r(X)$ is a subvariety of the Grassmanian $G(r, \mathbb{P}^n)$ and is the zero set of a section of a vector bundle. Indeed, let S be the tautological bundle on $G(r, \mathbb{P}^n)$. Then a section of $\oplus_{i=1}^s Sym^{d_i} H^0(\mathbb{P}^n, \mathcal{O}(1))$ induces a section of the vector bundle $\oplus_{i=1}^s Sym^{d_i} S^*$ on $G(r, \mathbb{P}^n)$. Thus, $F_r(X)$ is the zero locus of the section of the $\bigoplus_{i=1}^s Sym^{d_i} S^*$ induced by the equations defining the complete intersection X. A Lefschetz theorem is proved in [De-Ma, Theorem 3.4]:

$$H^i(G(r, \mathbb{P}^n), \mathbb{Q}) \to H^i(F_r(X), \mathbb{Q})$$

is bijective, for $i \leq \delta - 1$. We can apply Lemma 5.2 to get the orthogonal projectors in all degrees except in the middle dimension. The projector corresponding to the middle dimension can be gotten by subtracting the sum of these projectors from the diagonal class.

□

Corollary 5.4. *Suppose $X \subset \mathbb{P}^n$ is a smooth projective variety of dimension d. Let $r = 2d - n$. Then we can construct orthogonal projectors*

$$\pi_0, \pi_1, ..., \pi_r, \pi_{2d-r}, \pi_{2d-r+1}, ..., \pi_{2d}.$$

Proof. Barth [Ba] has proved a Lefschetz theorem for higher codimensional subvarieties in projective spaces:

$$H^i(\mathbb{P}^n, \mathbb{Q}) \to H^i(X, \mathbb{Q})$$

is bijective if $i \leq 2d-n$ and is injective if $i = 2d-n+1$. The claim now follows from Lemma 5.2. □

Remark 5.5. *The above corollary says that if we can embed a variety X in a low dimensional projective space then we get at least a partial set of orthogonal projectors. A conjecture of Hartshorne's says that any codimension two subvariety of $\mathbb{P}^n$ for $n \geq 6$ is a complete intersection. This gives more examples for subvarieties with several algebraic cohomology groups.*

5.1. Chow–Künneth decomposition for the universal plane curve. We want to find explicit equivariant Chow–Künneth projectors for the universal plane curve of degree d. Let $d \geq 1$ and consider the linear system $\mathbb{P} = |\mathcal{O}_{\mathbb{P}^2}(d)|$ and the universal plane curve

$$\begin{array}{ccc} \mathcal{C} & \subset & \mathbb{P}^2 \times \mathbb{P} \\ \downarrow & & \\ \mathbb{P}. & & \end{array}$$

Furthermore, we notice that the general linear group $G := GL_3(\mathbb{C})$ acts on $\mathbb{P}^2$ and hence acts on the projective space $\mathbb{P} = |\mathcal{O}_{\mathbb{P}^2}(d)|$. This gives an action on the product space $\mathbb{P}^2 \times \mathbb{P}$ and leaves the universal smooth plane curve $\mathcal{C} \subset \mathbb{P}^2 \times \mathbb{P}$ invariant under the G-action.

Lemma 5.6. *The variety $\mathcal{C}$ has an absolute Chow–Künneth decomposition and an absolute equivariant Chow–Künneth decomposition.*

Proof. We observe that $\mathcal{C} \subset \mathbb{P}^2 \times \mathbb{P}$ is a smooth hypersurface of bi-degree $(d, 1)$ with variables in $\mathbb{P}^2$ whose coefficients are polynomial functions on $\mathbb{P}$. Notice that $\mathbb{P}^2 \times \mathbb{P}$ has a Chow–Künneth decomposition and Lefschetz theorems hold for the embedding $\mathcal{C} \subset \mathbb{P}^2 \times \mathbb{P}$, since $\mathcal{O}(d, 1)$ is very ample. Now we can repeat the arguments from Lemma 5.2 to get an absolute Chow–Künneth decomposition and absolute equivariant Chow–Künneth decomposition, for the variety $\mathcal{C}$. □

5.2. Families of curves contained in homogeneous spaces. We notice that when $d = 3$ in the previous subsection, the family of plane cubics restricted to the loci of stable curves is a complete family of genus one stable curves. If $d \geq 4$, then the above family of plane curves is no longer a complete family of genus g curves. Hence to find families which closely approximate over the moduli spaces of stable curves, we need to look for curves embedded as complete intersections in other simpler looking varieties. For this purpose, we look at the curves embedded in special Fano varieties of small genus $g \leq 8$, which was studied by S. Mukai [Muk], [Muk2], [Muk3], [Muk5] and Ide-Mukai [IdMuk]. We recall the main result that we need.

Theorem 5.7. *Suppose C is a generic curve of genus $g \leq 8$. Then C is a complete intersection in a smooth projective variety which has only algebraic cohomology.*

Proof. This is proved in [Muk], [Muk2], [Muk3], [IdMuk] and [Muk5]. The below classification is for the generic curve.
When $g \leq 5$ then it is well-known that the generic curve is a linear section of a Grassmanian.
When $g = 6$ then a curve has finitely many g^1_4 if and only if it is a complete intersection of a Grassmanian and a smooth quadric , see [Muk3, Theorem 5.2].
When $g = 7$ then a curve is a linear section of a 10-dimensional spinor variety $X \subset \mathbb{P}^{15}$ if and only if it is non-tetragonal, see [Muk3, Main theorem].
When $g = 8$ then it is classically known that the generic curve is a linear section of the grassmanian $G(2, 6)$ in its Plücker embedding.

□

Suppose $\mathbb{P}(g)$ is the parameter space of linear sections of a Grassmanian or of a spinor variety, which depends on the genus, as in the proof of above Theorem 5.7. $\mathbb{P}(g)$ is a product of projective spaces on which an algebraic group G (copies of PGL_N) acts. Generic curves are isomorphic, if they are in the same orbit of G.

Proposition 5.8. *Suppose $\mathbb{P}(g)$ is as above, for $g \leq 8$. Then there is a universal curve*

$$\mathcal{C}_g \to \mathbb{P}(g)$$

such that the classifying (rational) map $\mathbb{P}(g) \to \mathcal{M}_g$ is dominant. The smooth projective variety $\mathcal{C}_g$ has an absolute Chow–Künneth decomposition and an absolute equivariant Chow-Künneth decomposition for the natural G–action mentioned above.

Proof. The first assertion follows from Theorem 5.7. For the second assertion notice that the universal curve, when $g \leq 8$, is a complete intersection in $\mathbb{P}(g) \times V$ where V is either a Grassmanian or a spinor variety, which are homogeneous varieties. In other words, $\mathcal{C}_g$ is a complete intersection in a space which has only algebraic cohomology. Hence, by Lemma 5.2, $\mathcal{C}_g$ has orthogonal projectors $\pi_0, \pi_1, ..., \pi_m, \pi_{2r-m}, \pi_{2r-m+1}, ..., \pi_{2r}$, where $r := \dim\mathcal{C}_g$ and $m = \dim\mathcal{C}_g - 1$, using Lefschetz hyplerplane section theorem. Taking $\pi_{m+1} = \Delta_{\mathcal{C}_g} - \sum_{i\neq m+1} \pi_i$, gives an absolute Chow–Künneth decomposition for $\mathcal{C}_g$. Now a homogeneous variety looks like $V = G/P$ where G is an (linear) algebraic group and P is a parabolic subgroup. Hence the group G acts on the variety V. This induces an action on the linear system $\mathbb{P}(g)$ and hence G acts on the ambient variety $\mathbb{P}(g)\times V$ and leaves the universal curve $\mathcal{C}_g$ invariant. Hence we can again apply Lemma 5.2 to obtain absolute equivariant Chow–Künneth decomposition for $\mathcal{C}_g$. □

Consider the universal family of curves $\mathcal{C}_g \to \mathbb{P}(g)$ as obtained above, which are equipped with an action of a linear algebraic group G.
Suppose there is an open subset $U_g \subset \mathbb{P}(g)$, with the universal family $\mathcal{C}_{U_g} \to U_g$, on which G acts freely to form a good quotient family

$$Y_g := \mathcal{C}_{U_g}/G \to S_g := U_g/G.$$

Notice that the classifying map $S_g \to \mathcal{M}_g$ is dominant.

Corollary 5.9. *The smooth variety Y_g has an absolute Chow–Künneth decomposition.*

Proof. Consider the localization sequence, for the embedding $j : \mathcal{C}_{U_g} \times \mathcal{C}_{U_g} \hookrightarrow \mathcal{C}_g \times \mathcal{C}_g$,

$$CH^d_G(\mathcal{C}_g \times \mathcal{C}_g)_{\mathbb{Q}} \stackrel{j^*}{\to} CH^d_G(\mathcal{C}_{U_g} \times \mathcal{C}_{U_g})_{\mathbb{Q}} \to 0.$$

Here d is the dimension of $\mathcal{C}_g$. Then the map j^* is an equivariant ring homomorphism and transforms orthogonal projectors to orthogonal projectors. Similarly there is a commuting diagram between the equivariant cohomologies:

$$\begin{array}{ccc} \bigoplus_i H^{2d-i}_G(\mathcal{C}_g) \otimes H^i_G(\mathcal{C}_g) & \stackrel{\simeq}{\to} & H^{2d}_G(\mathcal{C}_g, \mathbb{Q}) \\ \downarrow & & \downarrow \\ \bigoplus_i W_{2d-i}H^{2d-i}_G(\mathcal{C}_{U_g}) \otimes W_i H^i_G(\mathcal{C}_{U_g}) & \stackrel{\simeq}{\to} & W_{2d}H^{2d}_G(\mathcal{C}_{U_g}, \mathbb{Q}) \end{array}$$

The vertical arrows are surjective maps mapping onto the bottom weights of the equivariant cohomology groups. By Proposition 5.8, the variety $\mathcal{C}_g$ has an absolute equivariant Chow–Künneth decomposition. Hence the images of the equivariant Chow–Künneth projectors for the complete smooth variety $\mathcal{C}_g$, under the morphism j^* give equivariant Chow–Künneth projectors for the smooth variety $\mathcal{C}_{U_g}$.
Using [Ed-Gr], we have the identification of the rational Chow groups

$$CH^*(Y_g)_{\mathbb{Q}} = CH^*_G(\mathcal{C}_{U_g})_{\mathbb{Q}}$$

and

$$CH^*(Y_g \times Y_g)_{\mathbb{Q}} = CH^*_G(\mathcal{C}_{U_g} \times \mathcal{C}_{U_g})_{\mathbb{Q}}.$$

Furthermore, this respects the ring structure on the above rational Chow groups. A similar identification also holds for the rational cohomology groups. This means that the equivariant Chow–Künneth projectors for the variety $\mathcal{C}_{U_g}$

correspond to a complete set of absolute Chow–Künneth projectors for the quotient variety Y_g. $\square$

Remark 5.10. *Since Mukai has a similar classification for the non-generic curves in genus ≤ 8, one can obtain absolute equivariant Chow–Künneth decomposition for these special families of smooth curves, by applying the proof of Proposition 5.8. There is also a classification for $K3$-surfaces and in many cases the generic $K3$-surface is obtained as a linear section of a Grassmanian* [Muk]. *Hence we can apply the above results to families of $K3$-surfaces over spaces which dominate the moduli space of $K3$-surfaces.*

References

[Al-Kl] A.B. Altman, S.L. Kleiman, *Foundations of the theory of Fano schemes.* Compositio Math. 34 (1977), no. 1, 3–47.

[Ak-Jo] R. Akhtar, R. Joshua, *Künneth decomposition for quotient varieties*, Indag. Math. (N.S.) 17 (2006), no. 3, 319–344.

[Ba] W. Barth *Transplanting cohomology classes in complex-projective space.* Amer. J. Math. 92 1970 951–967.

[BBD] A.A. Beilinson, J. Bernstein, P. Deligne, *Faisceaux pervers* (French) [Perverse sheaves], Analysis and topology on singular spaces, I (Luminy, 1981), 5–171, Astérisque, 100, Soc. Math. France, Paris, 1982.

[Bi-Mi] E. Bierstone, P.D. Milman, *Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant.* Invent. Math. 128 (1997), no. 2, 207–302.

[BE] S. Bloch, H. Esnault, *Künneth projectors for open varieties*, preprint arXiv:math/0502447.

[Bo] A. Borel *Linear algebraic groups*. Second edition. Graduate Texts in Mathematics, 126. Springer-Verlag, New York, 1991. xii+288 pp.

[CH] A. Corti, M. Hanamura, *Motivic decomposition and intersection Chow groups. I*, Duke Math. J. 103 (2000), no. 3, 459–522.

[De-Ma] O. Debarre, L. Manivel, *Sur la varièté des espaces linéaires contenus dans une intersection complète.* (French) [The variety of linear spaces contained in a complete intersection] Math. Ann. 312 (1998), no. 3, 549–574.

[dA-Mü1] P. del Angel, S. Müller–Stach, *Motives of uniruled* 3*-folds*, Compositio Math. 112 (1998), no. 1, 1–16.

[dA-Mü2] P. del Angel, S. Müller-Stach, *On Chow motives of 3-folds*, Trans. Amer. Math. Soc. 352 (2000), no. 4, 1623–1633.

[dB-Az] S. del Baño Rollin, V. Navarro Aznar, *On the motive of a quotient variety.* Dedicated to the memory of Fernando Serrano. Collect. Math. 49 (1998), no. 2-3, 203–226.

[D] P. Deligne, *Thèorie de Hodge. III*, Inst. Hautes Ètudes Sci. Publ. Math. No. 44 (1974), 5–77.

[De-Mu] Ch. Deninger, J. Murre, *Motivic decomposition of abelian schemes and the Fourier transform*, J. Reine Angew. Math. 422 (1991), 201–219.

[Del-Mu] P. Deligne, D. Mumford, *The irreducibility of the space of curves of given genus*, Inst. Hautes Études Sci. Publ. Math. No. 36 1969 75–109.

[Di-Ha] S. Diaz, J. Harris, *Geometry of the Severi variety.* Trans. Amer. Math. Soc. 309 (1988), no. 1, 1–34.

[Ed-Gr] D. Edidin, W. Graham, *Equivariant intersection theory.* Invent. Math. 131 (1998), no. 3, 595–634.

[ELV] H. Esnault, M. Levine, E. Viehweg, *Chow groups of projective varieties of very small degree.* Duke Math. J. 87 (1997), no. 1, 29–58.

[Fu] W. Fulton, *Intersection theory*, Second edition. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., 2. Springer-Verlag, Berlin, 1998. xiv+470 pp.

[Fu2] W. Fulton, *Equivariant cohomology in Algebraic geometry*, http://www.math.lsa.umich.edu/∼ dandersn/eilenberg/index.html.

[Go-Mu] B. Gordon, J. P. Murre, *Chow motives of elliptic modular threefolds*, J. Reine Angew. Math. 514 (1999), 145–164.

[GHM1] B. Gordon, M. Hanamura, J. P. Murre, *Relative Chow-Künneth projectors for modular varieties*, J. Reine Angew. Math. 558 (2003), 1–14.

[GHM2] B. Gordon, M. Hanamura, J. P. Murre, *Absolute Chow-Künneth projectors for modular varieties*, J. Reine Angew. Math. 580 (2005), 139–155.

[Gu-Pe] V. Guletskiĭ, C. Pedrini, *Finite-dimensional motives and the conjectures of Beilinson and Murre* Special issue in honor of Hyman Bass on his seventieth birthday. Part III. K-Theory 30 (2003), no. 3, 243–263.

[Ha] J. Harris, *On the Severi problem*, Invent. Math. 84 (1986), no. 3, 445–461.

[Iy] J. N. Iyer, *Murre's conjectures and explicit Chow Künneth projectors for varieties with a nef tangent bundle*, to appear in Trans. of Amer. Math. Soc.

[Ja] U. Jannsen, *Motivic sheaves and filtrations on Chow groups*, Motives (Seattle, WA, 1991), 245–302, Proc. Sympos. Pure Math., 55, Part 1, Amer. Math. Soc., Providence, RI, 1994.

[Kol] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3., 32. Springer-Verlag, Berlin, 1996. viii+320 pp.

[KMM] J. Kollár, Y. Miyaoka, S. Mori, *Rationally connected varieties*, Jour. Alg. Geom. 1 (1992) 429–448.

[Man] Yu. Manin, *Correspondences, motifs and monoidal transformations* (in Russian), Mat. Sb. (N.S.) 77 (119) (1968), 475–507.

[MMWYK] A. Miller, S. Müller-Stach, S. Wortmann, Y.H. Yang, K. Zuo. *Chow-Künneth decomposition for universal families over Picard modular surfaces*, arXiv:math.AG/0505017.

[Muk] S. Mukai, *Biregular classification of Fano 3-folds and Fano manifolds of coindex 3.* Proc. Nat. Acad. Sci. U.S.A. 86 (1989), no. 9, 3000–3002.

[Muk2] S. Mukai, *Curves and Grassmannians.* Algebraic geometry and related topics (Inchon, 1992), 19–40, Conf. Proc. Lecture Notes Algebraic Geom., I, Int. Press, Cambridge, MA, 1993

[Muk3] S. Mukai, *Curves and symmetric spaces. I.* Amer. J. Math. 117 (1995), no. 6, 1627–1644.

[IdMuk] M. Ide, S. Mukai, *Canonical curves of genus eight.* Proc. Japan Acad. Ser. A Math. Sci. 79 (2003), no. 3, 59–64.

[Muk5] S. Mukai, *Curves, $K3$ surfaces and Fano 3-folds of genus* $\leq 10.$, Algebraic geometry and commutative algebra, Vol. I, 357–377, Kinokuniya, Tokyo, 1988.

[Md] Mumford, D. *Projective invariants of projective structures and applications* 1963 Proc. Internat. Congr. Mathematicians (Stockholm, 1962) pp. 526–530 Inst. Mittag-Leffler, Djursholm.

[Md2] D. Mumford, *Geometric invariant theory.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 34 Springer-Verlag, Berlin-New York 1965 vi+145 pp.

[Mu1] J. P. Murre, *On the motive of an algebraic surface*, J. Reine Angew. Math. 409 (1990), 190–204.

[Mu2] J. P. Murre, *On a conjectural filtration on the Chow groups of an algebraic variety. I. The general conjectures and some examples*, Indag. Math. (N.S.) 4 (1993), no. 2, 177–188.

[Mu3] J. P. Murre, *On a conjectural filtration on the Chow groups of an algebraic variety. II. Verification of the conjectures for threefolds which are the product on a surface and a curve*, Indag. Math. (N.S.) 4 (1993), no. 2, 189–201.

[Re-Yo] Z. Reichstein, B. Youssin, *Equivariant resolution of points of indeterminacy.* Proc. Amer. Math. Soc. 130 (2002), no. 8, 2183–2187.

[Sa] M. Saito, *Chow–Künneth decomposition for varieties with low cohomological level*, arXiv:math.AG/0604254.

[Sc] A. J. Scholl, *Classical motives*, Motives (Seattle, WA, 1991), 163–187, Proc. Sympos. Pure Math., 55, Part 1, Amer. Math. Soc., Providence, RI, 1994.

[Sh] A. M. Shermenev, *The motive of an abelian variety*, Funct. Analysis, 8 (1974), 55–61.

[Su] Sumihiro, H. *Equivariant completion.* J. Math. Kyoto Univ. 14 (1974), 1–28.

[To] B. Totaro, *The Chow ring of a classifying space.* Algebraic K-theory (Seattle, WA, 1997), 249–281, Proc. Sympos. Pure Math., 67, Amer. Math. Soc., Providence, RI, 1999.

Jaya NN Iyer
The Institute of
Mathematical Sciences,
CIT Campus
Taramani
Chennai 600113
India
jniyer@imsc.res.in

Stefan Müller–Stach
Mathematisches Institut der
Johannes Gutenberg Universität
Mainz
Staudingerweg 9
55099 Mainz
Germany
mueller-stach@mathematik.uni-mainz.de

Hecke Operators on Quasimaps into Horospherical Varieties

Dennis Gaitsgory and David Nadler

Received: November 10, 2006
Revised: January 30, 2009

Communicated by Edward Frenkel

Abstract. Let G be a connected reductive complex algebraic group. This paper and its companion [GN] are devoted to the space Z of meromorphic quasimaps from a curve into an affine spherical G-variety X. The space Z may be thought of as an algebraic model for the loop space of X. The theory we develop associates to X a connected reductive complex algebraic subgroup $\check{H}$ of the dual group $\check{G}$. The construction of $\check{H}$ is via Tannakian formalism: we identify a certain tensor category $\mathrm{Q}(Z)$ of perverse sheaves on Z with the category of finite-dimensional representations of $\check{H}$.

In this paper, we focus on horospherical varieties, a class of varieties closely related to flag varieties. For an affine horospherical G-variety X_{horo}, the category $\mathrm{Q}(Z_{\mathrm{horo}})$ is equivalent to a category of vector spaces graded by a lattice. Thus the associated subgroup $\check{H}_{\mathrm{horo}}$ is a torus. The case of horospherical varieties may be thought of as a simple example, but it also plays a central role in the general theory. To an arbitrary affine spherical G-variety X, one may associate a horospherical variety X_{horo}. Its associated subgroup $\check{H}_{\mathrm{horo}}$ turns out to be a maximal torus in the subgroup $\check{H}$ associated to X.

2000 Mathematics Subject Classification: 22E67, 14M17
Keywords and Phrases: Loop spaces, spherical varieties, Langlands duality

1. Introduction

Let G be a connected reductive complex algebraic group. In this paper and its companion [GN], we study the space Z of meromorphic quasimaps from a curve into an affine spherical G-variety X. A G-variety X is said to be spherical if a Borel subgroup of G acts on X with a dense orbit. Examples include flag varieties, symmetric spaces, and toric varieties. A meromorphic quasimap consists of a point of the curve, a G-bundle on the curve, and a meromorphic section of the associated X-bundle with a pole only at the distinguished point. The space Z may be thought of as an algebraic model for the loop space of X. The theory we develop identifies a certain tensor category $\mathrm{Q}(Z)$ of perverse sheaves on Z with the category of finite-dimensional representations of a connected reductive complex algebraic subgroup $\check{H}$ of the dual group $\check{G}$. Our method is to use Tannakian formalism: we endow $\mathrm{Q}(Z)$ with a tensor product, a fiber functor to vector spaces, and the necessary compatibility constraints so that it must be equivalent to the category of representations of such a group. Under this equivalence, the fiber functor corresponds to the forgetful functor which assigns to a representation of $\check{H}$ its underlying vector space. In the paper [GN], we define the category $\mathrm{Q}(Z)$, and endow it with a tensor product and fiber functor. This paper provides a key technical result needed for the construction of the fiber functor.

Horospherical G-varieties form a special class of G-varieties closely related to flag varieties. A subgroup $S \subset G$ is said to be horospherical if it contains the unipotent radical of a Borel subgroup of G. A G-variety X is said to be horospherical if for each point $x \in X$, its stabilizer $S_x \subset G$ is horospherical. When X is an affine horospherical G-variety, the subgroup $\check{H}$ we associate to it turns out to be a torus. To see this, we explicitly calculate the functor which corresponds to the restriction of representations from $\check{G}$. Representations of $\check{G}$ naturally act on the category $\mathrm{Q}(Z)$ via the geometric Satake correspondence. The restriction of representations is given by applying this action to the object of $\mathrm{Q}(Z)$ corresponding to the trivial representation of $\check{H}$. The main result of this paper describes this action in the horospherical case. The statement does not mention $\mathrm{Q}(Z)$, but rather what is needed in [GN] where we define and study $\mathrm{Q}(Z)$.

In the remainder of the introduction, we first describe a piece of the theory of geometric Eisenstein series which the main result of this paper generalizes. This may give the reader some context from which to approach the space Z and our main result. We then define Z and state our main result. Finally, we collect notation and preliminary results needed in what follows. Throughout the introduction, we use the term space for objects which are strictly speaking stacks and ind-stacks.

1.1. Background.

One way to approach the results of this paper is to interpret them as a generalization of a theorem of Braverman-Gaitsgory [BG, Theorem 3.1.4] from the theory of geometric Eisenstein series. Let C be a

smooth complete complex algebraic curve. The primary aim of the geometric Langlands program is to construct sheaves on the moduli space Bun_G of G-bundles on C which are eigensheaves for Hecke operators. These are the operators which result from modifying G-bundles at prescribed points of the curve C. Roughly speaking, the theory of geometric Eisenstein series constructs sheaves on Bun_G starting with local systems on the moduli space Bun_T, where T is the universal Cartan of G. When the original local system is sufficiently generic, the resulting sheaf is an eigensheaf for the Hecke operators.
At first glance, the link between Bun_T and Bun_G should be the moduli stack Bun_B of B-bundles on C, where $B \subset G$ is a Borel subgroup with unipotent radical $U \subset B$ and reductive quotient $T = B/U$. Unfortunately, naively working with the natural diagram

$$\begin{array}{ccc} \mathrm{Bun}_B & \to & \mathrm{Bun}_G \\ \downarrow & & \\ \mathrm{Bun}_T & & \end{array}$$

leads to difficulties: the fibers of the horizontal map are not compact. The eventual successful construction depends on V. Drinfeld's relative compactification of Bun_B along the fibers of the map to Bun_G. The starting point for the compactification is the observation that Bun_B also classifies data

$$(\mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_T \in \mathrm{Bun}_T, \sigma : \mathcal{P}_T \to \mathcal{P}_G \overset{G}{\times} G/U)$$

where σ is a T-equivariant bundle map to the $\mathcal{P}_G$-twist of G/U. From this perspective, it is natural to be less restrictive and allow maps into the $\mathcal{P}_G$-twist of the fundamental affine space

$$\overline{G/U} = \mathrm{Spec}(\mathbb{C}[G]^U).$$

Here $\mathbb{C}[G]$ denotes the ring of regular functions on G, and $\mathbb{C}[G]^U \subset \mathbb{C}[G]$ the (right) U-invariants. Following V. Drinfeld, we define the compactification $\overline{\mathrm{Bun}}_B$ to be that classifying quasimaps

$$(\mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_T \in \mathrm{Bun}_T, \sigma : \mathcal{P}_T \to \mathcal{P}_G \overset{G}{\times} \overline{G/U})$$

where σ is a T-equivariant bundle map which factors

$$\sigma|_{C'} : \mathcal{P}_T|_{C'} \to \mathcal{P}_G \overset{G}{\times} G/U|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overline{G/U}|_{C'},$$

for some open curve $C' \subset C$. Of course, the quasimaps that satisfy

$$\sigma : \mathcal{P}_T \to \mathcal{P}_G \overset{G}{\times} G/U$$

form a subspace canonically isomorphic to Bun_B.
Since the Hecke operators on Bun_G do not lift to $\overline{\mathrm{Bun}}_B$, it is useful to introduce a version of $\overline{\mathrm{Bun}}_B$ on which they do. Following [BG, Section 4], we define the space ${}_\infty\overline{\mathrm{Bun}}_B$ to be that classifying meromorphic quasimaps

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_T \in \mathrm{Bun}_T, \sigma : \mathcal{P}_T|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} \overline{G/U}|_{C\setminus c})$$

where σ is a T-equivariant bundle map which factors

$$\sigma|_{C'} : \mathcal{P}_T|_{C'} \to \mathcal{P}_G \overset{G}{\times} G/U|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overline{G/U}|_{C'},$$

for some open curve $C' \subset C \setminus c$. We call $c \in C$ the pole point of the quasimap. Given a meromorphic quasimap with G-bundle $\mathcal{P}_G$ and pole point $c \in C$, we may modify $\mathcal{P}_G$ at c and obtain a new meromorphic quasimap. In this way, the Hecke operators on Bun_G lift to ${}_\infty\overline{\mathrm{Bun}}_B$.

Now the result we seek to generalize [BG, Theorem 3.1.4] describes how the Hecke operators act on a distinguished object of the category $\mathrm{P}({}_\infty\overline{\mathrm{Bun}}_B)$ of perverse sheaves with $\mathbb{C}$-coefficients on ${}_\infty\overline{\mathrm{Bun}}_B$. Let $\Lambda_G = \mathrm{Hom}(\mathbb{C}^\times, T)$ be the coweight lattice, and let $\Lambda_G^+ \subset \Lambda$ be the semigroup of dominant coweights of G. For $\lambda \in \Lambda_G^+$, we have the Hecke operator

$$H_G^\lambda : \mathrm{P}({}_\infty\overline{\mathrm{Bun}}_B) \to \mathrm{P}({}_\infty\overline{\mathrm{Bun}}_B)$$

given by convolving with the simple spherical modification of coweight λ. (See [BG, Section 4] or Section 5 below for more details.) For $\mu \in \Lambda_G$, we have the locally closed subspace ${}_\infty\overline{\mathrm{Bun}}_B^\mu \subset {}_\infty\overline{\mathrm{Bun}}_B$ that classifies data for which the map

$$\mathcal{P}_T(\mu \cdot c)|_{C\setminus c} \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} \overline{G/U}|_{C \setminus c}$$

extends to a holomorphic map

$$\mathcal{P}_T(\mu \cdot c) \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} \overline{G/U}$$

which factors

$$\mathcal{P}_T(\mu \cdot c) \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} G/U \to \mathcal{P}_G \overset{G}{\times} \overline{G/U}.$$

We write ${}_\infty\overline{\mathrm{Bun}}_B^{\leq\mu} \subset {}_\infty\overline{\mathrm{Bun}}_B$ for the closure of ${}_\infty\overline{\mathrm{Bun}}_B^\mu \subset {}_\infty\overline{\mathrm{Bun}}_B$, and

$$\mathrm{IC}^{\leq\mu}_{{}_\infty\overline{\mathrm{Bun}}_B} \in \mathrm{P}({}_\infty\overline{\mathrm{Bun}}_B)$$

for the intersection cohomology sheaf of ${}_\infty\overline{\mathrm{Bun}}_B^{\leq\mu} \subset {}_\infty\overline{\mathrm{Bun}}_B$.

Theorem 1.1.1. [BG, Theorem 3.1.4] *For $\lambda \in \Lambda_G^+$, there is a canonical isomorphism*

$$H_G^\lambda(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_B}) \simeq \sum_{\mu \in \Lambda_T} \mathrm{IC}^{\leq\mu}_{{}_\infty\overline{\mathrm{Bun}}_B} \otimes \mathrm{Hom}_{\check{T}}(V_{\check{T}}^\mu, V_{\check{G}}^\lambda)$$

Here we write $V_{\check{G}}^\lambda$ for the irreducible representation of the dual group $\check{G}$ with highest weight $\lambda \in \Lambda_G^+$, and $V_{\check{T}}^\mu$ for the irreducible representation of the dual torus $\check{T}$ of weight $\mu \in \Lambda_G$.

In the same paper of Braverman-Gaitsgory [BG, Section 4], there is a generalization [BG, Theorem 4.1.5] of this theorem from the Borel subgroup $B \subset G$ to other parabolic subgroups $P \subset G$. We recall and use this generalization in Section 5 below. It is the starting point for the results of this paper.

1.2. Main result. The main result of this paper is a version of [BG, Theorem 3.1.4] for X an arbitrary affine horospherical G-variety with a dense G-orbit $\mathring{X} \subset X$. For any point in the dense G-orbit $\mathring{X} \subset X$, we refer to its stabilizer $S \subset G$ as the generic stabilizer of X. All such subgroups are conjugate to each other. By choosing such a point, we obtain an identification $\mathring{X} \simeq G/S$.
To state our main theorem, we first introduce some more notation. Satz 2.1 of [Kn] states that the normalizer of a horospherical subgroup $S \subset G$ is a parabolic subgroup $P \subset G$ with the same derived group $[P,P] = [S,S]$. We write A for the quotient torus P/S, and $\Lambda_A = \mathrm{Hom}(\mathbb{C}^\times, A)$ for its coweight lattice. Similarly, for the identity component $S^0 \subset S$, we write A_0 for the quotient torus P/S_0, and $\Lambda_{A_0} = \mathrm{Hom}(\mathbb{C}^\times, A_0)$ for its coweight lattice. The natural maps $T \to A_0 \to A$ induce maps of coweight lattices

$$\Lambda_T \xrightarrow{q} \Lambda_{A_0} \xrightarrow{i} \Lambda_A,$$

where q is a surjection, and i is an injection. For a conjugate of S, the associated tori are canonically isomorphic to those associated to S. Thus when S is the generic stabilizer of a horospherical G-variety X, the above tori, lattices and maps are canonically associated to X.
For an affine horospherical G-variety X with dense G-orbit $\mathring{X} \subset X$, we define the space Z to be that classifying mermorphic quasimaps into X. Such a quasimap consists of data

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \sigma : C \setminus c \to \mathcal{P}_G \overset{G}{\times} X|_{C\setminus c})$$

where σ is a section which factors

$$\sigma|_{C'} : C' \to \mathcal{P}_G \overset{G}{\times} \mathring{X}|_{C'} \to \mathcal{P}_G \overset{G}{\times} X|_{C'},$$

for some open curve $C' \subset C \setminus c$.
Given a meromorphic quasimap into X with G-bundle $\mathcal{P}_G$ and pole point $c \in C$, we may modify $\mathcal{P}_G$ at c and obtain a new meromorphic quasimap. But in this context the resulting Hecke operators on Z do not in general preserve the category of perverse sheaves. Instead, we must consider the bounded derived category $\mathrm{Sh}(Z)$ of sheaves of $\mathbb{C}$-modules on Z. For $\lambda \in \Lambda_G^+$, we have the Hecke operator

$$H_G^\lambda : \mathrm{Sh}(Z) \to \mathrm{Sh}(Z)$$

given by convolving with the simple spherical modification of coweight λ. (See Section 5 below for more details.) For $\kappa \in \Lambda_{A_0}$, we have a locally closed subspace $Z^\kappa \subset Z$ consisting of meromorphic quasimaps that factor

$$\sigma : C \setminus c \to \mathcal{P}_G \overset{G}{\times} \mathring{X}|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} X|_{C\setminus c}$$

and have a singularity of type κ at $c \in C$. (See Section 3.5 below for more details.) We write $Z^{\leq\kappa} \subset Z$ for the closure of $Z^\kappa \subset Z$, and

$$\mathrm{IC}_Z^{\leq\kappa} \in \mathrm{Sh}(Z)$$

for its intersection cohomology sheaf.
Our main result is the following.

Theorem 1.2.1. *For $\lambda \in \Lambda_G^+$, there is an isomorphism*

$$H_G^\lambda(\mathrm{IC}_{\overline{Z}}^{\leq 0}) \simeq \sum_{\kappa \in \Lambda_{A_0}} \sum_{\mu \in \Lambda_T, q(\mu)=\kappa} \mathrm{IC}_{\overline{Z}}^{\leq \kappa} \otimes \mathrm{Hom}_{\check{T}}(V_{\check{T}}^\mu, V_{\check{G}}^\lambda)[\langle 2\check{\rho}_M, \mu\rangle].$$

Here the torus A_0 and its coweight lattice Λ_{A_0} are those associated to the generic stabilizer $S \subset G$. We write M for the Levi quotient of the normalizer $P \subset G$ of the generic stabilizer $S \subset G$, and $2\check{\rho}_M$ for the sum of the positive roots of M.

In the context of the companion paper [GN], the theorem translates into the following fundamental statement. The tensor category $\mathrm{Q}(Z)$ associated to X is the category of semisimple perverse sheaves with simple summands $\mathrm{IC}_{\overline{Z}}^{\leq \kappa}$, for $\kappa \in \Lambda_{A_0}$, and the dual subgroup $\check{H}$ associated to X is the subtorus $\mathrm{Spec}\, \mathbb{C}[\Lambda_{A_0}] \subset \check{T}$.

1.3. Notation. Throughout this paper, let G be a connected reductive complex algebraic group, let $B \subset G$ be a Borel subgroup with unipotent radical $U(B)$, and let $T = B/U(B)$ be the abstract Cartan.

Let $\check{\Lambda}_G$ denote the weight lattice $\mathrm{Hom}(T, \mathbb{C}^\times)$, and $\check{\Lambda}_G^+ \subset \check{\Lambda}_G$ the semigroup of dominant weights. For $\lambda \in \check{\Lambda}_G^+$, we write V_G^λ for the irreducible representation of G of highest weight λ.

Let Λ_G denote the coweight lattice $\mathrm{Hom}(\mathbb{C}^\times, T)$, and $\Lambda_G^+ \subset \Lambda_G$ the semigroup of dominant coweights. For $\lambda \in \Lambda_G^+$, let $V_{\check{G}}^\lambda$ denote the irreducible representation of the dual group $\check{G}$ of highest weight λ.

Let $\Lambda_G^{\mathrm{pos}} \subset \Lambda_G$ denote the semigroup of coweights in Λ_G which are non-negative on $\check{\Lambda}_G^+$, and let $R_G^{\mathrm{pos}} \subset \Lambda_G^{\mathrm{pos}}$ denote the semigroup of positive coroots.

Let $P \subset G$ be a parabolic subgroup with unipotent radical $U(P)$, and let M be the Levi factor $P/U(P)$.

We have the natural map

$$\check{r} : \check{\Lambda}_{M/[M,M]} \to \check{\Lambda}_G$$

of weights, and the dual map

$$r : \Lambda_G \to \Lambda_{M/[M,M]}$$

of coweights.

Let $\check{\Lambda}_{G,P}^+ \subset \check{\Lambda}_{M/[M,M]}$ denote the inverse image $\check{r}^{-1}(\check{\Lambda}_G^+)$. Let $\Lambda_{G,P}^{\mathrm{pos}} \subset \Lambda_{M/[M,M]}$ denote the semigroup of coweights in $\Lambda_{M/[M,M]}$ which are non-negative on $\check{\Lambda}_{G,P}^+$. Let $R_{G,P}^{\mathrm{pos}} \subset \Lambda_{G,P}^{\mathrm{pos}}$ denote the image $r(R_G^{\mathrm{pos}})$.

Let $\mathcal{W}_M$ denote the Weyl group of M, and let $\mathcal{W}_M\check{\Lambda}_G^+ \subset \check{\Lambda}_G$ denote the union of the translates of $\check{\Lambda}_G^+$ by $\mathcal{W}_M$. Let $\tilde{\Lambda}_{G,P}^{\mathrm{pos}} \subset \Lambda_M^+$ denote the semigroup of dominant coweights of M which are nonnegative on $\mathcal{W}_M\check{\Lambda}_G^+$.

Finally, let $\langle\cdot,\cdot\rangle : \check{\Lambda}_G \times \Lambda_G \to \mathbb{Z}$ denote the natural pairing, and let $\check{\rho}_M \in \check{\Lambda}_G$ denote half the sum of the positive roots of M.

1.4. Bundles and Hecke correspondences. Let C be a smooth complete complex algebraic curve.

For a connected complex algebraic group H, let Bun_H be the moduli stack of H-bundles on C. Objects of Bun_H will be denoted by $\mathcal{P}_H$.

Let $\mathcal{H}_H$ be the Hecke ind-stack that classifies data

$$(c \in C, \mathcal{P}^1_H, \mathcal{P}^2_H \in \mathrm{Bun}_H, \alpha : \mathcal{P}^1_H|_{C\setminus c} \xrightarrow{\sim} \mathcal{P}^2_H|_{C\setminus c})$$

where α is an isomorphism of H-bundles. We have the maps

$$\mathrm{Bun}_H \xleftarrow{h^{\leftarrow}_H} \mathcal{H}_H \xrightarrow{h^{\rightarrow}_H} \mathrm{Bun}_H$$

defined by

$$h^{\leftarrow}_H(c, \mathcal{P}^1_H, \mathcal{P}^2_H, \alpha) = \mathcal{P}^1_H \qquad h^{\rightarrow}_H(c, \mathcal{P}^1_H, \mathcal{P}^2_H, \alpha) = \mathcal{P}^2_H,$$

and the map

$$\pi : \mathcal{H}_H \to C$$

defined by

$$\pi(c, \mathcal{P}^1_H, \mathcal{P}^2_H, \alpha) = c.$$

It is useful to have another description of the Hecke ind-stack $\mathcal{H}_H$ for which we introduce some more notation. Let $\mathcal{O}$ be the ring of formal power series $\mathbb{C}[[t]]$, let $\mathcal{K}$ be the field of formal Laurent series $\mathbb{C}((t))$, and let D be the formal disk $\mathrm{Spec}(\mathcal{O})$. For a point $c \in C$, let $\mathcal{O}_c$ be the completed local ring of C at c, and let D_c be the formal disk $\mathrm{Spec}(\mathcal{O}_c)$. Let $\mathrm{Aut}(\mathcal{O})$ be the group-scheme of automorphisms of the ring $\mathcal{O}$. Let $H(\mathcal{O})$ be the group of $\mathcal{O}$-valued points of H, and let $H(\mathcal{K})$ be the group of $\mathcal{K}$-valued points of H. Let Gr_H be the affine Grassmannian of H. It is an ind-scheme whose set of $\mathbb{C}$-points is the quotient $H(\mathcal{K})/H(\mathcal{O})$.

Now consider the $(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))$-torsor

$$\widehat{\mathrm{Bun}_H \times} C \to \mathrm{Bun}_H \times C$$

that classifies data

$$(c \in C, \mathcal{P}_H \in \mathrm{Bun}_H, \beta : D \times H \xrightarrow{\sim} \mathcal{P}_H|_{D_c}, \gamma : D \xrightarrow{\sim} D_c)$$

where β is an isomorphism of H-bundles, and γ is an identification of formal disks. We have an identification

$$\mathcal{H}_H \simeq \widehat{\mathrm{Bun}_H \times} C \overset{(H(\mathcal{O})\rtimes \mathrm{Aut}(\mathcal{O}))}{\times} \mathrm{Gr}_H$$

such that the projection $h^{\rightarrow}_H$ corresponds to the obvious projection from the twisted product to Bun_H.

For H reductive, the $(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))$-orbits $\mathrm{Gr}^{\lambda}_H \subset \mathrm{Gr}_H$ are indexed by $\lambda \in \Lambda^+_H$. For $\lambda \in \Lambda^+_H$, we write $\mathcal{H}^{\lambda}_H \subset \mathcal{H}_H$ for the substack

$$\mathcal{H}^{\lambda}_H \simeq \widehat{\mathrm{Bun}_H \times} C \overset{(H(\mathcal{O})\rtimes \mathrm{Aut}(\mathcal{O}))}{\times} \mathrm{Gr}^{\lambda}_H .$$

For a parabolic subgroup $P \subset H$, the connected components $S_{P,\theta} \subset \mathrm{Gr}_P$ are indexed by $\theta \in \Lambda_P/\Lambda_{[P,P]^{sc}}$, where $[P,P]^{sc}$ denotes the simply connected cover of $[P,P]$. For $\theta \in \Lambda_P/\Lambda_{[P,P]^{sc}}$, we write $\mathcal{S}_{P,\theta} \subset \mathcal{H}_P$ for the ind-substack

$$\mathcal{S}_{P,\theta} \simeq \widehat{\mathrm{Bun}_P \times C} \overset{(P(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))}{\times} S_{P,\theta}.$$

For $\theta \in \Lambda_P/\Lambda_{[P,P]^{sc}}$, and $\lambda \in \Lambda_H^+$, we write $\mathcal{S}^\lambda_{P,\theta} \subset \mathcal{H}_P$ for the ind-substack

$$\mathcal{S}^\lambda_{P,\theta} \simeq \widehat{\mathrm{Bun}_P \times C} \overset{(P(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))}{\times} S^\lambda_{P,\theta}$$

where $S^\lambda_{P,\theta}$ denotes the intersection $S_{P,\theta} \cap \mathrm{Gr}^\lambda_H$.
For any ind-stack $\mathcal{Z}$ over $\mathrm{Bun}_H \times C$, we have the $(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))$-torsor

$$\widehat{\mathcal{Z}} \to \mathcal{Z}$$

obtained by pulling back the $(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))$-torsor

$$\widehat{\mathrm{Bun}_H \times C} \to \mathrm{Bun}_H \times C.$$

We also have the Cartesian diagram

$$\begin{array}{ccc} \mathcal{H}_H \underset{\mathrm{Bun}_H \times C}{\times} \mathcal{Z} & \overset{h^{\to}_H}{\to} & \mathcal{Z} \\ \downarrow & & \downarrow \\ \mathcal{H}_H & \overset{h^{\to}_H}{\to} & \mathrm{Bun}_H \end{array}$$

and an identification

$$\mathcal{H}_H \underset{\mathrm{Bun}_H \times C}{\times} \mathcal{Z} \simeq \widehat{\mathcal{Z}} \overset{(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))}{\times} \mathrm{Gr}_H$$

such that the projection $h^{\to}_H$ corresponds to the obvious projection from the twisted product to $\mathcal{Z}$. For $\mathcal{F} \in \mathrm{Sh}(\mathcal{Z})$, and $\mathcal{P} \in \mathrm{P}_{(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))}(\mathrm{Gr}_H)$, we may form the twisted product

$$(\mathcal{F} \widetilde{\boxtimes} \mathcal{P})^r \in \mathrm{Sh}(\mathcal{H}_H \underset{\mathrm{Bun}_H \times C}{\times} \mathcal{Z}).$$

with respect to the map $h^{\to}_H$. In particular, for $\lambda \in \Lambda^+_H$, we may take $\mathcal{P}$ to be the intersection cohomology sheaf $\mathcal{A}^\lambda_G$ of the closure $\overline{\mathrm{Gr}}^\lambda_H \subset \mathrm{Gr}_H$ of the $(H(\mathcal{O}) \rtimes \mathrm{Aut}(\mathcal{O}))$-orbit $\mathrm{Gr}^\lambda_H \subset \mathrm{Gr}_H$.

2. Affine horospherical G-varieties

A subgroup $S \subset G$ is said to be horospherical if it contains the unipotent radical of a Borel subgroup of G. A G-variety X is said to be horospherical if for each point $x \in X$, its stabilizer $S_x \subset G$ is horospherical. A G-variety X is said to be spherical if a Borel subgroup of G acts on X with a dense orbit. Note that a horospherical G-variety contains a dense G-orbit if and only if it is spherical.

Let X be an affine G-variety. As a representation of G, the ring of regular functions $\mathbb{C}[X]$ decomposes into isotypic components

$$\mathbb{C}[X] \simeq \sum_{\lambda \in \check{\Lambda}_G^+} \mathbb{C}[X]_\lambda.$$

We say that $\mathbb{C}[X]$ is graded if

$$\mathbb{C}[X]_\lambda \mathbb{C}[X]_\mu \subset \mathbb{C}[X]_{\lambda+\mu},$$

for all $\lambda, \mu \in \check{\Lambda}_G^+$. We say that $\mathbb{C}[X]$ is simple if the irreducible representation V^λ of highest weight λ occurs in $\mathbb{C}[X]_\lambda$ with multiplicity 0 or 1, for all $\lambda \in \check{\Lambda}_G^+$.

Proposition 2.0.1. *Let X be an affine G-variety.*
(1) [Pop, Proposition 8, (3)] *X is horospherical if and only if $\mathbb{C}[X]$ is graded.*
(2) [Pop, Theorem 1] *X is spherical if and only if $\mathbb{C}[X]$ is simple.*

We see by the proposition that affine horospherical G-varieties containing a dense G-orbit are classified by finitely-generated subsemigroups of $\check{\Lambda}_G^+$. To such a variety X, one associates the subsemigroup

$$\check{\Lambda}_X^+ \subset \check{\Lambda}_G^+$$

of dominant weights λ with $\dim \mathbb{C}[X]_\lambda > 0$.

2.1. Structure of generic stabilizer.

Theorem 2.1.1. [Kn, Satz 2.2] *If X is an irreducible horospherical G-variety, then there is an open G-invariant subset $W \subset X$, and a G-equivariant isomorphism $W \simeq G/S \times Y$, where $S \subset G$ is a horospherical subgroup, and Y is a variety on which G acts trivially.*

Note that for any two such open subsets $W \subset X$ and isomorphisms $W \simeq G/S \times Y$, the subgroups $S \subset G$ are conjugate. We refer to such a subgroup $S \subset G$ as the generic stabilizer of X.

Lemma 2.1.2. [Kn, Satz 2.1] *If $S \subset G$ is a horospherical subgroup, then its normalizer is a parabolic subgroup $P \subset G$ with the same derived group $[P,P] = [S,S]$ and unipotent radical $U(P) = U(S)$.*

Note that the identity component $S^0 \subset S$ is also horospherical with the same derived group $[S^0, S^0] = [S,S]$ and unipotent radical $U(S^0) = U(S)$.
Let $S \subset G$ be a horospherical subgroup with identity component $S^0 \subset S$, and normalizer $P \subset G$. We write A for the quotient torus P/S, and Λ_A for its coweight lattice $\mathrm{Hom}(\mathbb{C}^\times, A)$. Similarly, we write A_0 for the quotient torus P/S^0, and Λ_{A_0} for its coweight lattice $\mathrm{Hom}(\mathbb{C}^\times, A_0)$. The natural maps

$$T \to A_0 \to A$$

induce maps of coweight lattices

$$\Lambda_T \xrightarrow{q} \Lambda_{A_0} \xrightarrow{i} \Lambda_A,$$

where q is a surjection, and i is an injection. For a conjugate of S, the associated tori, lattices, and maps are canonically isomorphic to those associated to S.

Thus when S is the generic stabilizer of a horospherical G-variety X, the tori, lattices and maps are canonically associated to X.

We shall need the following finer description of which subgroups $S \subset G$ may appear as the generic stabilizer of an affine horospherical G-variety. To state it, we introduce some more notation used throughout the paper. For a horospherical subgroup $S \subset G$ with identity component $S^0 \subset S$, and normalizer $P \subset G$, let M be the Levi quotient $P/U(P)$, let M_S be the Levi quotient $S/U(S)$, and let M_S^0 be the identity component of M_S. The natural maps

$$S^0 \to S \to P$$

induce isomorphisms of derived groups

$$[M_S^0, M_S^0] \xrightarrow{\sim} [M_S, M_S] \xrightarrow{\sim} [M, M].$$

We write $\Lambda_{M/[M,M]}$ for the coweight lattice of the torus $M/[M,M]$, and $\Lambda_{M_S^0/[M_S,M_S]}$ for the coweight lattice of the torus $M_S^0/[M_S, M_S]$. The natural maps

$$M_S^0/[M_S, M_S] \to M/[M, M] \to A_0$$

induce a short exact sequence of coweight lattices

$$0 \to \Lambda_{M_S^0/[M_S,M_S]} \to \Lambda_{M/[M,M]} \to \Lambda_{A_0} \to 0.$$

Proposition 2.1.3. *Let $S \subset G$ be a horospherical subgroup. Then S is the generic stabilizer of an affine horospherical G-variety containing a dense G-orbit if and only if*

$$\Lambda_{M_S^0/[M_S,M_S]} \cap \Lambda_{G,P}^{\mathrm{pos}} = \langle 0 \rangle.$$

Proof. The proof of the proposition relies on the following lemma. Let $\check{V}$ be a finite-dimensional real vector space, and let $\check{V}^+$ be an open set in $\check{V}$ which is preserved by the action of $\mathbb{R}^{>0}$. Let V be the dual of $\check{V}$, and let V^{pos} be the closed cone of covectors in V that are nonnegative on all vectors in $\check{V}^+$. For a linear subspace $\check{W} \subset \check{V}$, we write $\check{W}^\perp \subset V$ for its orthogonal.

Lemma 2.1.4. *The map $\check{W} \mapsto \check{W}^\perp$ provides a bijection from the set of all linear subspaces $\check{W} \subset \check{V}$ such that $\check{W} \cap \check{V}^+ \neq \emptyset$ to the set of all linear subspaces $W \subset V$ such that $W \cap V^{\mathrm{pos}} = \langle 0 \rangle$.*

Proof. If $\check{W} \cap \check{V}^+ \neq \emptyset$, then clearly $\check{W}^\perp \cap V^{\mathrm{pos}} = \langle 0 \rangle$. Conversely, if $W \cap V^{\mathrm{pos}} = \langle 0 \rangle$, then since $\check{V}^+$ is open, there is a hyperplane $H \subset V$ such that $W \subset H$, and $H \cap V^{\mathrm{pos}} = \langle 0 \rangle$. Thus $H^\perp \subset W^\perp$, and $H^\perp \cap \check{V}^+ \neq \emptyset$, and so $W^\perp \cap \check{V}^+ \neq \emptyset$. $\square$

Now suppose X is an affine horospherical G-variety with an open G-orbit and generic stabilizer $S \subset G$ with normalizer $P \subset G$. Then we have $\check{\Lambda}_X^+ \subset \check{\Lambda}_{G,P}^+$, since otherwise $[S, S]$ would be smaller. We also have that $\check{\Lambda}_X^+$ intersects the interior of $\check{\Lambda}_{G,P}^+$, since otherwise $[S, S]$ would be larger. Applying Lemma 2.1.4, we conclude

$$\Lambda_{M_S^0/[M_S,M_S]} \cap \Lambda_{G,P}^{\mathrm{pos}} = \langle 0 \rangle.$$

Conversely, suppose $S \subset G$ is a horospherical subgroup with normalizer $P \subset G$. We define X to be the spectrum of the ring $\mathbb{C}[X]$ of (right) S-invariants in the ring of regular functions $\mathbb{C}[G]$. Then $\mathbb{C}[X]$ is finitely-generated, since S contains the unipotent radical of a Borel subgroup of G. We have $\check{\Lambda}^+_X \subset \check{\Lambda}^+_{G,P}$, since otherwise $[S,S]$ would be smaller. Suppose

$$\Lambda_{M_S^0/[M_S,M_S]} \cap \Lambda^{\mathrm{pos}}_{G,P} = \langle 0 \rangle.$$

Applying Lemma 2.1.4, we conclude that $\check{\Lambda}^+_X$ intersects the interior of $\check{\Lambda}^+_{G,P}$. Therefore $S/[S,S]$ consists of exactly those elements of $P/[P,P]$ annhilated by $\check{\Lambda}^+_X$, and so S is the generic stabilizer of X. □

2.2. Canonical affine closure. Let $S \subset G$ be the generic stabilizer of an affine horospherical G-variety X containing a dense G-orbit. Let $\mathbb{C}[G]$ be the ring of regular functions on G, and let $\mathbb{C}[G]^S \subset \mathbb{C}[G]$ be the (right) S-invariants. We call the affine variety

$$\overline{G/S} = \mathrm{Spec}(\mathbb{C}[G]^S)$$

the canonical affine closure of G/U. We have the natural map

$$\overline{G/S} \to X$$

corresponding to the restriction map

$$\mathbb{C}[X] \to \mathbb{C}[G/S] \simeq \mathbb{C}[G]^S.$$

Since S is horospherical, the ring $\mathbb{C}[G]^S$ is simple and graded, and so the affine variety $\overline{G/S}$ is spherical and horospherical.
Although we do not use the following, it clarifies the relation between X and the canonical affine closure $\overline{G/S}$.

Proposition 2.2.1. *Let X be an affine horospherical G-variety containing a dense G-orbit and generic stabilizer $S \subset G$. The semigroup $\check{\Lambda}^+_{\overline{G/S}} \subset \check{\Lambda}_G$ is the intersection of the dominant weights $\check{\Lambda}^+_G \subset \check{\Lambda}_G$ with the group generated by the semigroup $\check{\Lambda}^+_X \subset \check{\Lambda}_G$.*

Proof. Let $P \subset G$ be the normalizer of $S \subset G$. The intersection of $\check{\Lambda}^+_G$ and the group generated by $\check{\Lambda}^+_X$ consists of exactly those weights in $\check{\Lambda}^+_{G,P}$ that annhilate $S/[S,S]$. □

3. Ind-stacks

As usual, let C be a smooth complete complex algebraic curve.

3.1. Labellings. Fix a pair $(\Lambda, \Lambda^{\mathrm{pos}})$ of a lattice Λ and a semigroup $\Lambda^{\mathrm{pos}} \subset \Lambda$. We shall apply the following to the pair $(\Lambda_{M/[M,M]}, \Lambda^{\mathrm{pos}}_{G,P})$.
For $\theta^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}$, we write $\mathfrak{U}(\theta^{\mathrm{pos}})$ for a decomposition

$$\theta^{\mathrm{pos}} = \sum_m n_m \theta^{\mathrm{pos}}_m$$

where $\theta^{\mathrm{pos}}_m \in \Lambda^{\mathrm{pos}} \setminus \{0\}$ are pairwise distinct and n_m are positive integers.

For $\theta^{\text{pos}} \in \Lambda^{\text{pos}}$, and a decomposition $\mathfrak{U}(\theta^{\text{pos}})$, we write $C^{\mathfrak{U}(\theta^{\text{pos}})}$ for the partially symmetrized power $\prod_m C^{(n_m)}$ of the curve C. We write $C_0^{\mathfrak{U}(\theta^{\text{pos}})} \subset C^{\mathfrak{U}(\theta^{\text{pos}})}$ for the complement of the diagonal divisor.
For Θ a pair $(\theta, \mathfrak{U}(\theta^{\text{pos}}))$ consisting of $\theta \in \Lambda$, and $\mathfrak{U}(\theta^{\text{pos}})$ a decomposition of $\theta^{\text{pos}} \in \Lambda^{\text{pos}}$, we write C^{Θ} for the product $C \times C^{\mathfrak{U}(\theta^{\text{pos}})}$. We write $C_0^{\Theta} \subset C^{\Theta}$ for the complement of the diagonal divisor. Although C^{Θ} is independent of θ, it is notationally convenient to denote it as we do.

3.2. $\overline{\textbf{Ind-stack}}$ ASSOCIATED TO PARABOLIC SUBGROUP. Fix a parabolic subgroup $P \subset G$, and let M be its Levi quotient $P/U(P)$. For our application, P will be the normalizer of the generic stabilizer $S \subset G$ of an irreducible affine horospherical G-variety.
Let ${}_\infty\overline{\mathrm{Bun}}_P$ be the ind-stack that classifies data

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_{M/[M,M]} \in \mathrm{Bun}_{M/[M,M]},$$
$$\sigma : \mathcal{P}_{M/[M,M]}|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} \overline{G/[P,P]}|_{C\setminus c})$$

where σ is an $M/[M,M]$-equivariant section which factors

$$\sigma|_{C'} : \mathcal{P}_{M/[M,M]}|_{C'} \to \mathcal{P}_G \overset{G}{\times} G/[P,P]|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overline{G/[P,P]}|_{C'}$$

for some open curve $C' \subset C \setminus c$.

3.2.1. *Stratification.* Let Θ be a pair $(\theta, \mathfrak{U}(\theta^{\text{pos}}))$, with $\theta \in \Lambda_{M/[M,M]}$, and $\theta^{\text{pos}} \in \Lambda^{\text{pos}}_{G,P}$. We recall that we have a locally closed embedding

$$j_\Theta : \mathrm{Bun}_P \times C_0^{\Theta} \to {}_\infty\overline{\mathrm{Bun}}_P$$

defined by

$$j_\Theta(\mathcal{P}_P, (c, \sum_{m,n} \theta_m^{\text{pos}} \cdot c_{m,n})) = (c, \mathcal{P}_P \overset{P}{\times} G, \mathcal{P}_P \overset{P}{\times} [P,P](-\theta \cdot c - \sum_{m,n} \theta_m^{\text{pos}} \cdot c_{m,n}), \sigma)$$

where σ is the natural map

$$\mathcal{P}_P \overset{P}{\times} [P,P](-\theta \cdot c - \sum_{m,n} \theta_m^{\text{pos}} \cdot c_{m,n})|_{C\setminus c} \to \mathcal{P}_P \overset{P}{\times} G \overset{G}{\times} \overline{G/[P,P]}|_{C\setminus c}$$

induced by the inclusion

$$\mathcal{P}_P \overset{P}{\times} \overline{P/[P,P]} \subset \mathcal{P}_P \overset{P}{\times} \overline{G/[P,P]} \simeq \mathcal{P}_P \overset{P}{\times} G \overset{G}{\times} \overline{G/[P,P]}.$$

The following is an ind-version of [BG, Propositions 6.1.2 & 6.1.3], or [BFGM, Proposition 1.5], and we leave the proof to the reader.

PROPOSITION 3.2.2. *Let Θ be a pair $(\theta, \mathfrak{U}(\theta^{\text{pos}}))$, with $\theta \in \Lambda_{M/[M,M]}$, and $\theta^{\text{pos}} \in \Lambda^{\text{pos}}_{G,P}$.*
Every closed point of ${}_\infty\overline{\mathrm{Bun}}_P$ belongs to the image of a unique j_Θ.

For Θ a pair $(\theta, \mathfrak{U}(\theta^{\mathrm{pos}}))$, with $\theta \in \Lambda_{M/[M,M]}$, and $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$, we write ${}_\infty\overline{\mathrm{Bun}}_P^{\Theta} \subset {}_\infty\overline{\mathrm{Bun}}_P$ for the image of j_Θ, and ${}_\infty\overline{\mathrm{Bun}}_P^{\leq\Theta} \subset {}_\infty\overline{\mathrm{Bun}}_P$ for the closure of ${}_\infty\overline{\mathrm{Bun}}_P^{\Theta} \subset {}_\infty\overline{\mathrm{Bun}}_P$.

For Θ a pair $(\theta, \mathfrak{U}(0))$, with $\theta \in \Lambda_{M/[M,M]}$, the substack ${}_\infty\overline{\mathrm{Bun}}_P^{\Theta} \subset {}_\infty\overline{\mathrm{Bun}}_P$ classifies data $(c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}, \sigma)$ for which the map

$$\mathcal{P}_{M/[M,M]}(\theta \cdot c)|_{C\setminus c} \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} \overline{G/[P,P]}|_{C\setminus c}$$

extends to a holomorphic map

$$\mathcal{P}_{M/[M,M]}(\theta \cdot c) \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} \overline{G/[P,P]}$$

which factors

$$\mathcal{P}_{M/[M,M]}(\theta \cdot c) \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} G/[P,P] \to \mathcal{P}_G \overset{G}{\times} \overline{G/[P,P]}.$$

In this case, we write j_θ in place of j_Θ, ${}_\infty\overline{\mathrm{Bun}}_P^{\theta}$ in place of ${}_\infty\overline{\mathrm{Bun}}_P^{\Theta}$, and ${}_\infty\overline{\mathrm{Bun}}_P^{\leq\theta}$ in place of ${}_\infty\overline{\mathrm{Bun}}_P^{\leq\Theta}$. For example, ${}_\infty\overline{\mathrm{Bun}}_P^{\leq 0} \subset {}_\infty\overline{\mathrm{Bun}}_P$ is the closure of the canonical embedding

$$j_0 : \mathrm{Bun}_P \times C \to {}_\infty\overline{\mathrm{Bun}}_P.$$

3.3. $\widetilde{\textbf{Ind-stack}}$ associated to parabolic subgroup. Fix a parabolic subgroup $P \subset G$, and let M be its Levi quotient $P/U(P)$. As usual, for our application, P will be the normalizer of the generic stabilizer $S \subset G$ of an irreducible affine horospherical G-variety.

Let ${}_\infty\widetilde{\mathrm{Bun}}_P$ be the ind-stack that classifies data

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_M \in \mathrm{Bun}_M, \sigma : \mathcal{P}_M|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} \overline{G/U(P)}|_{C\setminus c})$$

where σ is an M-equivariant section which factors

$$\sigma|_{C'} : \mathcal{P}_M|_{C'} \to \mathcal{P}_G \overset{G}{\times} G/U(P)|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overline{G/U(P)}|_{C'}$$

for some open curve $C' \subset C \setminus c$.

3.3.1. *Stratification.* For $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$, we write $\tilde{\mathfrak{U}}(\theta^{\mathrm{pos}})$ for a collection of (not necessarily distinct) elements $\tilde{\theta}_m^{\mathrm{pos}} \in \tilde{\Lambda}_{G,P}^{\mathrm{pos}} \setminus \{0\}$ such that

$$\theta^{\mathrm{pos}} = \sum_m r(\tilde{\theta}_m^{\mathrm{pos}}).$$

We write $r(\tilde{\mathfrak{U}}(\theta^{\mathrm{pos}}))$ for the decomposition such a collection defines.

Let $\tilde{\Theta}$ be a pair $(\tilde{\theta}, \tilde{\mathfrak{U}}(\theta^{\mathrm{pos}}))$ with $\tilde{\theta} \in \Lambda_M^+$, and $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$, and let Θ be the associated pair $(r(\tilde{\theta}), r(\tilde{\mathfrak{U}}(\theta^{\mathrm{pos}})))$. We define the Hecke ind-stack

$$\mathcal{H}_{M,0}^{\tilde{\Theta}} \to C_0^{\Theta}$$

to be that with fiber over $(c, c_{\mathfrak{U}(\theta^{\mathrm{pos}})}) \in C_0^{\Theta}$, where $c_{\mathfrak{U}(\theta^{\mathrm{pos}})} = \sum_m r(\tilde{\theta}_m^{\mathrm{pos}}) \cdot c_m$, the fiber product

$$\mathcal{H}_M^{\tilde{\theta}}|_c \underset{\mathrm{Bun}_M}{\times} \prod_{\mathrm{Bun}_M} \mathcal{H}_M^{\tilde{\theta}_m^{\mathrm{pos}}}|_{c_m}.$$

The following is an ind-version of [BG, Proposition 6.2.5], or [BFGM, Proposition 1.9], and we leave the proof to the reader.

Proposition 3.3.2. *Let $\tilde{\Theta}$ be a pair $(\tilde{\theta}, \tilde{\mathfrak{U}}(\theta^{\mathrm{pos}}))$ with $\tilde{\theta} \in \Lambda_M^+$, and $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$. On the level of reduced ind-stacks, there is a locally closed embedding*

$$j_{\tilde{\Theta}} : \mathrm{Bun}_P \underset{\mathrm{Bun}_M}{\times} \mathcal{H}_{M,0}^{\tilde{\Theta}} \to {}_\infty\widetilde{\mathrm{Bun}}_P.$$

Every closed point of ${}_\infty\widetilde{\mathrm{Bun}}_P$ belongs to the image of a unique $j_{\tilde{\Theta}}$.

For $\tilde{\Theta}$ a pair $(\tilde{\theta}, \tilde{\mathfrak{U}}(\theta^{\mathrm{pos}}))$, with $\tilde{\theta} \in \Lambda_M^+$, and $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$, we write ${}_\infty\widetilde{\mathrm{Bun}}_P^{\tilde{\Theta}} \subset {}_\infty\widetilde{\mathrm{Bun}}_P$ for the image of $j_{\tilde{\Theta}}$, and ${}_\infty\widetilde{\mathrm{Bun}}_P^{\leq\tilde{\Theta}} \subset {}_\infty\widetilde{\mathrm{Bun}}_P$ for the closure of ${}_\infty\widetilde{\mathrm{Bun}}_P^{\tilde{\Theta}} \subset {}_\infty\widetilde{\mathrm{Bun}}_P$.

For $\tilde{\Theta}$ a pair $(\tilde{\theta}, \tilde{\mathfrak{U}}(0))$, with $\tilde{\theta} \in \Lambda_M^+$, we write $j_{\tilde{\theta}}$ in place of $j_{\tilde{\Theta}}$, ${}_\infty\widetilde{\mathrm{Bun}}_P^{\tilde{\theta}}$ in place of ${}_\infty\widetilde{\mathrm{Bun}}_P^{\tilde{\Theta}}$, and ${}_\infty\widetilde{\mathrm{Bun}}_P^{\leq\tilde{\theta}}$ in place of ${}_\infty\widetilde{\mathrm{Bun}}_P^{\leq\tilde{\Theta}}$ For example, ${}_\infty\widetilde{\mathrm{Bun}}_P^{\leq 0}$ is the closure of the canonical embedding

$$j_{\tilde{0}} : \mathrm{Bun}_P \times C \to {}_\infty\widetilde{\mathrm{Bun}}_P.$$

3.4. $\overline{\textbf{Ind-stack}}$ associated to generic stabilizer. Let X be an irreducible affine horospherical G-variety with generic stabilizer $S \subset G$. Recall that the normalizer of S is a parabolic subgroup $P \subset G$ with the same derived group $[P, P] = [S, S]$ and unipotent radical $U(P) = U(S)$. Let M be the Levi quotient $P/U(P)$, and let M_S be the Levi quotient $S/U(S)$.

Let $\overline{Z}_{\mathrm{can}}$ be the ind-stack that classifies data

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_{M_S/[M_S,M_S]} \in \mathrm{Bun}_{M_S/[M_S,M_S]},$$
$$\sigma : \mathcal{P}_{M_S/[M_S,M_S]}|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} \overline{G/[S,S]}|_{C\setminus c})$$

where σ is an $M_S/[M_S, M_S]$-equivariant section which factors

$$\sigma|_{C'} : \mathcal{P}_{M_S/[M_S,M_S]}|_{C'} \to \mathcal{P}_G \overset{G}{\times} G/[S,S]|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overline{G/[S,S]}|_{C'}$$

for some open curve $C' \subset C \setminus c$.

The following is immediate from the definitions.

Proposition 3.4.1. *The diagram*

$$\begin{array}{ccc} \overline{Z}_{\mathrm{can}} & \to & {}_\infty\overline{\mathrm{Bun}}_P \\ \downarrow & & \downarrow \\ \mathrm{Bun}_{M_S/[M_S,M_S]} & \to & \mathrm{Bun}_{M/[M,M]} \end{array}$$

is Cartesian.

3.4.2. *Stratification.* Let Θ be a pair $(\theta, \mathfrak{U}(\theta^{\mathrm{pos}}))$, with $\theta \in \Lambda_{M/[M,M]}$, and $\theta^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P}$.
We write $\overline{Z}^{\Theta}_{\mathrm{can}} \subset \overline{Z}_{\mathrm{can}}$ for the substack which completes the Cartesian diagram

$$\begin{array}{ccc} \overline{Z}^{\Theta}_{\mathrm{can}} & \to & {}_{\infty}\overline{\mathrm{Bun}}^{\Theta}_{P} \\ \downarrow & & \downarrow \\ \mathrm{Bun}_{M_S/[M_S,M_S]} & \to & \mathrm{Bun}_{M/[M,M]}, \end{array}$$

and $\overline{Z}^{\leq\Theta}_{\mathrm{can}} \subset \overline{Z}_{\mathrm{can}}$ for the closure of $\overline{Z}^{\Theta}_{\mathrm{can}} \subset \overline{Z}_{\mathrm{can}}$.
For Θ a pair $(\theta, \mathfrak{U}(0))$, with $\theta \in \Lambda_{M/[M,M]}$, we write $\overline{Z}^{\theta}_{\mathrm{can}}$ in place of $\overline{Z}^{\Theta}_{\mathrm{can}}$, and $\overline{Z}^{\leq\theta}_{\mathrm{can}}$ in place of $\overline{Z}^{\leq\Theta}_{\mathrm{can}}$. For example, $\overline{Z}^{\leq 0}_{\mathrm{can}}$ is the closure of the canonical embedding

$$\mathrm{Bun}_S \times C \subset \overline{Z}_{\mathrm{can}}.$$

3.5. Naive ind-stack associated to X. Let X be an affine horospherical G-variety with dense G-orbit $\overset{\circ}{X} \subset X$ and generic stabilizer $S \subset G$.
Let Z be the ind-stack that classifies data

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \sigma : C \setminus c \to \mathcal{P}_G \overset{G}{\times} X|_{C\setminus c})$$

where σ is a section which factors

$$\sigma|_{C'} : C' \to \mathcal{P}_G \overset{G}{\times} \overset{\circ}{X}|_{C'} \to \mathcal{P}_G \overset{G}{\times} X|_{C'}$$

for some open curve $C' \subset C \setminus c$.
For the canonical affine closure $\overline{G/S}$, we write Z_{can} for the corresponding ind-stack.
We call the ind-stack Z naive, since there is no auxilliary bundle in its definition: it classifies honest sections. Let ${}^{\star}Z$ be the ind-stack that classifies data

$$(c \in C, \mathcal{P}_G \in \mathrm{Bun}_G, \mathcal{P}_{M/M_S} \in \mathrm{Bun}_{M/M_S}, \sigma : \mathcal{P}_{M/M_S}|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} X|_{C\setminus c})$$

where σ is an M/M_S-equivariant section which factors

$$\sigma|_{C'} : \mathcal{P}_{M/M_S}|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overset{\circ}{X}|_{C'} \to \mathcal{P}_G \overset{G}{\times} X|_{C'}$$

for some open curve $C' \subset C\setminus c$. Here as usual, we write M for the Levi quotient $P/U(P)$ of the normalizer $P \subset G$ of the generic stabilizer $S \subset G$, and M_S for the Levi quotient $S/U(S)$.
For the canonical affine closure $\overline{G/S}$, we write ${}^{\star}Z_{\mathrm{can}}$ for the corresponding ind-stack.
The following analogue of Proposition 3.4.1 is immediate from the definitions.

Proposition 3.5.1. *The diagram*

$$\begin{array}{ccc} Z & \to & {}^{\star}Z \\ \downarrow & & \downarrow \\ \mathrm{Bun}_{\langle 1\rangle} & \to & \mathrm{Bun}_{M/M_S} \end{array}$$

is Cartesian.

3.5.2. *Stratification.* We shall content ourselves here with defining the substacks of the naive ind-stack Z which appear in our main theorem. (See [GN] for a different perspective involving a completely local definition.) Recall that we write A for the quotient torus P/S, and Λ_A for its coweight lattice. Similarly, for the identity component $S^0 \subset S$, we write A_0 for the quotient torus P/S^0, and Λ_{A_0} for its coweight lattice. The natural map $A_0 \to A$ provides an inclusion of coweight lattices $\Lambda_{A_0} \to \Lambda_A$. For $\kappa \in \Lambda_A$, we shall define a closed substack $Z^{\leq\kappa} \subset Z$. When $\kappa \in \Lambda_{A_0}$, the closed substack $Z^{\leq\kappa} \subset Z$ appears in our main theorem.

For $\kappa \in \Lambda_A$, let ${}^\star Z^\kappa \subset {}^\star Z$ be the locally closed substack that classifies data $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}, \sigma)$ for which the natural map

$$\mathcal{P}_{M/M_S}(\kappa \cdot c)|_{C\setminus c} \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} X|_{C\setminus c}$$

extends to a holomorphic map

$$\mathcal{P}_{M/M_S}(\kappa \cdot c) \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} X$$

which factors

$$\mathcal{P}_{M/M_S}(\kappa \cdot c) \xrightarrow{\sigma} \mathcal{P}_G \overset{G}{\times} \overset{\circ}{X} \to \mathcal{P}_G \overset{G}{\times} X.$$

We write ${}^\star Z^{\leq\kappa} \subset {}^\star Z$ for the closure of ${}^\star Z^\kappa \subset {}^\star Z$.

For $\kappa \in \Lambda_A$, let $Z^\kappa \subset Z$ be the locally closed substack completing the Cartesian diagram

$$\begin{array}{ccc} Z^\kappa & \to & {}^\star Z^\kappa \\ \downarrow & & \downarrow \\ \mathrm{Bun}_{\langle 1\rangle} & \to & \mathrm{Bun}_{M/M_S} \end{array}.$$

We write $Z^{\leq\kappa} \subset Z$ for the closure of $Z^\kappa \subset Z$.

4. Maps

4.1. The map $\mathfrak{r} : {}_\infty\widetilde{\mathrm{Bun}}_P \to {}_\infty\overline{\mathrm{Bun}}_P$. Let Θ be a pair $(\theta, \mathfrak{U}(\theta^{\mathrm{pos}}))$, with $\theta \in \Lambda_{M/[M,M]}$, and $\theta^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P}$. and $\mathfrak{U}(\theta^{\mathrm{pos}})$ a decomposition $\theta^{\mathrm{pos}} = \sum_m n_m \theta^{\mathrm{pos}}_m$. Let ${}_\infty\widetilde{\mathrm{Bun}}{}^\Theta_P \subset {}_\infty\widetilde{\mathrm{Bun}}_P$ be the inverse image of ${}_\infty\overline{\mathrm{Bun}}{}^\Theta_P \subset {}_\infty\overline{\mathrm{Bun}}_P$ under the natural map

$$\mathfrak{r} : {}_\infty\widetilde{\mathrm{Bun}}_P \to {}_\infty\overline{\mathrm{Bun}}_P.$$

We would like to describe the fibers of the restriction of $\mathfrak{r}$ to the substack ${}_\infty\widetilde{\mathrm{Bun}}{}^\Theta_P \subset {}_\infty\widetilde{\mathrm{Bun}}_P$.

First, we define the Hecke ind-substack

$$\mathcal{H}_M^{\flat(\theta)} \subset \mathcal{H}_M$$

to be the union of the spherical Hecke substacks

$$\mathcal{H}_M^\mu \subset \mathcal{H}_M,$$

for $\mu \in \Lambda_M^+$ such that $r(\mu) = \theta$.

Second, if there exists $\tilde{\mu}^{\mathrm{pos}} \in \tilde{\Lambda}^{\mathrm{pos}}_{G,P}$ such that $r(\tilde{\mu}^{\mathrm{pos}}) = \theta^{\mathrm{pos}}$, we define the Hecke substack

$$\mathcal{H}_M^{\flat(\theta^{\mathrm{pos}})} \subset \mathcal{H}_M$$

to be the union of the spherical Hecke substacks

$$\mathcal{H}_M^{\tilde{\mu}^{\mathrm{pos}}} \subset \mathcal{H}_M,$$

for $\tilde{\mu}^{\mathrm{pos}} \in \tilde{\Lambda}^{\mathrm{pos}}_{G,P}$ such that $r(\tilde{\mu}^{\mathrm{pos}}) = \theta^{\mathrm{pos}}_m$.

Finally, we define the Hecke ind-stack

$$\mathcal{H}_{M,0}^{\flat(\Theta)} \to C_0^{\Theta}$$

to be that with fiber over $(c, c_{\mathfrak{U}(\theta^{\mathrm{pos}})}) \in C_0^{\Theta}$, where $c_{\mathfrak{U}(\theta^{\mathrm{pos}})} = \sum_{m,n} \theta^{\mathrm{pos}}_m \cdot c_{m,n}$, the fiber product

$$\mathcal{H}_M^{\flat(\theta)}|_c \underset{\mathrm{Bun}_M}{\times} \prod_{\mathrm{Bun}_M} \mathcal{H}_M^{\flat(\theta^{\mathrm{pos}}_m)}|_{c_{m,n}}.$$

The following is an ind-version of [BG, Proposition 6.2.5], or [BFGM, Proposition 1.9], and we leave the proof to the reader. It is also immediately implied by Proposition 3.3.2.

Proposition 4.1.1. *Let Θ be a pair $(\theta, \mathfrak{U}(\theta^{\mathrm{pos}}))$, with $\theta \in \Lambda_{M/[M,M]}$, $\theta^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P}$, and $\mathfrak{U}(\theta^{\mathrm{pos}})$ a decomposition $\theta^{\mathrm{pos}} = \sum_m n_m \theta^{\mathrm{pos}}_m$.*

If for all m there exists $\tilde{\mu}^{\mathrm{pos}}_m \in \tilde{\Lambda}^{\mathrm{pos}}_{G,P}$ such that $r(\tilde{\mu}^{\mathrm{pos}}_m) = \theta^{\mathrm{pos}}_m$, then on the level of reduced stacks there is a canonical isomorphism

$$_\infty\widetilde{\mathrm{Bun}}_P^{\Theta} \simeq \mathrm{Bun}_P \underset{\mathrm{Bun}_M}{\times} \mathcal{H}_{M,0}^{\flat(\Theta)}$$

such that the following diagram commutes

$$\begin{array}{ccc} _\infty\widetilde{\mathrm{Bun}}_P^{\Theta} & \simeq & \mathrm{Bun}_P \underset{\mathrm{Bun}_M}{\times} \mathcal{H}_{M,0}^{\flat(\Theta)} \\ \downarrow & & \downarrow \\ _\infty\overline{\mathrm{Bun}}_P^{\Theta} & \simeq & \mathrm{Bun}_P \times C_0^{\Theta} \end{array}$$

where the right hand side is the obvious projection.

If there is an m such that θ^{pos}_m is not equal to $r(\tilde{\mu}^{\mathrm{pos}})$, for any $\tilde{\mu}^{\mathrm{pos}} \in \tilde{\Lambda}^{\mathrm{pos}}_{G,P}$, then $_\infty\widetilde{\mathrm{Bun}}_P^{\Theta}$ is empty.

4.2. The map $\mathfrak{p} : \overline{Z}_{\mathrm{can}} \to Z_{\mathrm{can}}$. Let X be an irreducible affine horospherical G-variety with generic stabilizer $S \subset G$. Recall that the normalizer of a horospherical subgroup $S \subset G$ is a parabolic subgroup $P \subset G$ with the same derived group $[P,P] = [S,S]$ and unipotent radical $U(P) = U(S)$. We write M for the Levi quotient $P/U(P)$, M_S for the Levi quotient $S/U(S)$, and M_S^0 for the identity component of M_S. We write A for the quotient torus P/S, and Λ_A for its coweight lattice. Similarly, for the identity component $S^0 \subset S$, we write A_0 for the quotient torus P/S^0, and Λ_{A_0} for its coweight lattice. The natural map $M/[M,M] \to A_0$ induces a surjection of coweight lattices $\Lambda_{M/[M,M]} \to \Lambda_{A_0}$ which we denote by p. The kernel of p is the coweight lattice $\Lambda_{M_S^0/[M_S,M_S]}$. (Note that the component group of M_S is abelian.)

Associated to the canonical affine closure $\overline{G/S}$, we have a Cartesian diagram of ind-stacks

$$\begin{array}{ccc} \overline{Z}_{\mathrm{can}} & \rightarrow & {}_\infty\overline{\mathrm{Bun}}_P \\ \mathfrak{p}\downarrow & & \downarrow\mathfrak{p} \\ Z_{\mathrm{can}} & \rightarrow & {}^\star Z_{\mathrm{can}} \end{array}$$

We would like to describe some properties of the vertical maps.

PROPOSITION 4.2.1. *The map* $\mathfrak{p} : {}_\infty\overline{\mathrm{Bun}}_P \to {}^\star Z_{\mathrm{can}}$ *is ind-finite.*
For $\theta \in \Lambda_{M/[M,M]}$*, its restriction to* ${}_\infty\overline{\mathrm{Bun}}_P^\theta$ *is an embedding with image* ${}^\star Z_{\mathrm{can}}^{p(\theta)}$*, and its restriction to* ${}_\infty\overline{\mathrm{Bun}}_P^{\leq\theta}$ *is finite with image* ${}^\star Z_{\mathrm{can}}^{\leq p(\theta)}$*.*

Proof. For a point $(c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}, \overline{\sigma}) \in {}_\infty\overline{\mathrm{Bun}}_P$, we write $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}, \sigma) \in {}^\star Z_{\mathrm{can}}$ for its image under $\mathfrak{p}$. Observe that for $\theta \in \Lambda_{M/[M,M]}$, the point $(c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}(\theta \cdot c), \overline{\sigma}) \in {}_\infty\overline{\mathrm{Bun}}_P$ maps to $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}(p(\theta) \cdot c), \sigma) \in {}^\star Z_{\mathrm{can}}$ under $\mathfrak{p}$. Therefore to prove the proposition, it suffices to show that the restriction of $\mathfrak{p}$ to the canonical embedding $\mathrm{Bun}_P \subset {}_\infty\overline{\mathrm{Bun}}_P$ is an embedding with image the canonical embedding $\mathrm{Bun}_P \subset {}^\star Z_{\mathrm{can}}$, and its restriction to ${}_\infty\overline{\mathrm{Bun}}_P^{\leq 0}$ is a finite map with image ${}^\star Z_{\mathrm{can}}^{\leq 0}$. The first assertion is immediate from the definitions. To prove the second, recall that by [BG, Proposition 1.3.6], ${}_\infty\overline{\mathrm{Bun}}_P$ is proper over Bun_G, and so the map $\mathfrak{p}$ is proper since it respects the projection to Bun_G. Therefore it suffices to check that the fibers over closed points of the restriction of $\mathfrak{p}$ to ${}_\infty\overline{\mathrm{Bun}}_P^{\leq 0}$ are finite.
Let Θ be a pair $(0, \mathfrak{U}(\theta^{\mathrm{pos}}))$, with $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$. The stack ${}_\infty\overline{\mathrm{Bun}}_P^\Theta$ classifies data

$$(c, \mathcal{P}_P, c_\Theta, \mathcal{P}_{M/[M,M]})$$

together with an isomorphism

$$\alpha : \mathcal{P}_P \overset{P}{\times} P/[P,P] \simeq \mathcal{P}_{M/[M,M]}(c_\Theta).$$

The fiber of $\mathfrak{p}$ through such a point classifies data

$$(\mathcal{P}_P, c'_{\Theta'}, \mathcal{P}'_{M/[M,M]})$$

together with an isomorphism

$$\alpha' : \mathcal{P}_P \overset{P}{\times} P/[P,P] \simeq \mathcal{P}'_{M/[M,M]}(c'_{\Theta'})$$

such that the labelling $c_\Phi = c_\Theta - c'_{\Theta'}$ takes values in $\Lambda_{M_S^0/[M_S,M_S]}$. Therefore we need only check that for $\theta^{\mathrm{pos}} \in \Lambda_{G,P}^{\mathrm{pos}}$, there are only a finite number of $\phi \in \Lambda_{M_S^0/[M_S,M_S]}$ such that $\theta^{\mathrm{pos}} + \phi \in \Lambda_{G,P}^{\mathrm{pos}}$. By Proposition 2.1.3, the lattice $\Lambda_{M_S^0/[M_S,M_S]}$ intersects the semigroup $\Lambda_{G,P}^{\mathrm{pos}}$ only at 0. Since $\Lambda_{G,P}^{\mathrm{pos}}$ is finitely-generated, this implies that for $\theta^{\mathrm{pos}} \in \Lambda_{M/[M,M]}$, the coset $\theta^{\mathrm{pos}} + \Lambda_{M_S^0/[M_S,M_S]}$ intersects $\Lambda_{G,P}^{\mathrm{pos}}$ in a finite set. □

Corollary 4.2.2. *The map* $\mathfrak{p} : \overline{Z}_{\mathrm{can}} \to Z_{\mathrm{can}}$ *is ind-finite.*
For $\theta \in \Lambda_{M/[M,M]}$*, its restriction to* $\overline{Z}^{\theta}_{\mathrm{can}}$ *is an embedding with image* $Z^{p(\theta)}_{\mathrm{can}}$*, and its restriction to* $\overline{Z}^{\leq\theta}_{\mathrm{can}}$ *is finite with image* $Z^{\leq p(\theta)}_{\mathrm{can}}$*.*

4.3. The map $\mathfrak{s} : Z_{\mathrm{can}} \to Z$. Let X be an affine horospherical variety with dense G-orbit $\mathring{X} \subset X$ and generic stabilizer $S \subset G$.
Associated to the natural map $\overline{G/S} \to X$, we have a Cartesian diagram of ind-stacks

$$\begin{array}{ccc} Z_{\mathrm{can}} & \to & {}^{\star}Z_{\mathrm{can}} \\ \mathfrak{s}\downarrow & & \downarrow\mathfrak{s} \\ Z & \to & {}^{\star}Z. \end{array}$$

We would like to describe some properties of the vertical maps.

Proposition 4.3.1. *The map* $\mathfrak{s} : {}^{\star}Z_{\mathrm{can}} \to {}^{\star}Z$ *is a closed embedding.*
For $\kappa \in \Lambda_A$*, its restriction to* ${}^{\star}Z^{\kappa}_{\mathrm{can}}$ *is an embedding with image* ${}^{\star}Z^{\kappa}$*, and its restriction to* ${}^{\star}Z^{\leq\kappa}_{\mathrm{can}}$ *is a closed embedding with image* ${}^{\star}Z^{\leq\kappa}$*.*

Proof. First note that $\mathfrak{s}$ is injective on scheme-valued points since for $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}\sigma) \in {}^{\star}Z_{\mathrm{can}}$, the map

$$\sigma : \mathcal{P}_{M/M_S}|_{C\setminus c} \to \mathcal{P}_G \overset{G}{\times} \overline{G/S}|_{C\setminus c}$$

factors

$$\sigma|_{C'} : \mathcal{P}_{M/M_S}|_{C'} \to \mathcal{P}_G \overset{G}{\times} G/S|_{C'} \to \mathcal{P}_G \overset{G}{\times} \overline{G/S}|_{C'},$$

for some open curve $C' \subset C \setminus c$, and the map $\overline{G/S} \to X$ restricted to G/S is an embedding.
Now to see $\mathfrak{s}$ is a closed embedding, it suffices to check that $\mathfrak{s}$ satisfies the valuative criterion of properness. Let $D = \operatorname{Spec}\mathbb{C}[[t]]$ be the disk, and $D^{\times} = \operatorname{Spec}\mathbb{C}((t))$ the punctured disk. Let $f : D \to Z$ be a map with a partial lift $F^{\times} : D^{\times} \to Z_{\mathrm{can}}$. Let $\mathcal{P}^f_G$ be the D-family of G-bundles defined by f, and let $\mathcal{P}^f_{M/M_S}$ be the D-family of M/M_S-bundles defined by f. We must check that any partial lift

$$\Sigma^{\times} : \mathcal{P}^f_{M/M_S}|_{(C\setminus c)\times D^{\times}} \to \mathcal{P}^f_G \overset{G}{\times} \overline{G/S}|_{(C\setminus c)\times D^{\times}}$$

of a map

$$\sigma : \mathcal{P}^f_{M/M_S}|_{(C\setminus c)\times D} \to \mathcal{P}^f_G \overset{G}{\times} X|_{(C\setminus c)\times D}$$

which factors

$$\sigma|_{C'\times D} : \mathcal{P}^f_{M/M_S}|_{C'\times D} \to \mathcal{P}^f_G \overset{G}{\times} G/S|_{C'\times D} \to \mathcal{P}^f_G \overset{G}{\times} X|_{C'\times D},$$

for some open curve $C' \subset C \setminus c$, extends to $(C \setminus c) \times D$. Since $\overline{G/S} \to X$ restricted to G/S is an embedding with image G/S, we may lift $\sigma|_{C'\times D}$ to extend $\Sigma^{\times}$ to $C' \times D$. But then $\Sigma^{\times}$ extends completely since $\mathcal{P}^f_{M/M_S}|_{(C\setminus c)\times D}$ is normal and the complement of $\mathcal{P}^f_{M/M_S}|_{C'\times D}$ is of codimension 2.

Finally, for a point $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}, \sigma_{\mathrm{can}}) \in {}^\star Z_{\mathrm{can}}$, we write $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}, \sigma) \in {}^\star Z$ for its image under $\mathfrak{s}$. Observe that for $\kappa \in \Lambda_A$, the point $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}(\kappa \cdot c), \sigma_{\mathrm{can}}) \in {}^\star Z_{\mathrm{can}}$ maps to $(c, \mathcal{P}_G, \mathcal{P}_{M/M_S}(\kappa \cdot c), \sigma) \in {}^\star Z$ under $\mathfrak{s}$. Therefore to complete the proof of the proposition, it suffices to show that the restriction of $\mathfrak{s}$ to the canonical embedding $\mathrm{Bun}_S \times C \subset {}^\star Z_{\mathrm{can}}$ has image the canonical embedding $\mathrm{Bun}_S \times C \subset {}^\star Z$. This is immediate from the definitions. □

COROLLARY 4.3.2. *The map $\mathfrak{s} : Z_{\mathrm{can}} \to Z$ is a closed embedding. For $\kappa \in \Lambda_A$, its restriction to Z^κ_{can} is an embedding with image Z^κ, and its restriction to $Z^{\leq\kappa}_{\mathrm{can}}$ is a closed embedding with image $Z^{\leq\kappa}$.*

5. CONVOLUTION

Let X be an affine horospherical G-variety with dense G-orbit $\mathring{X} \subset X$ and generic stabilizer $S \subset G$.

The following diagram summarizes the ind-stacks and maps under consideration

$$\begin{array}{ccccccc} {}_\infty\widetilde{\mathrm{Bun}}_P & \xrightarrow{\mathfrak{r}} & {}_\infty\overline{\mathrm{Bun}}_P & \xrightarrow{\mathfrak{p}} & {}^\star Z_{\mathrm{can}} & & \\ & & \uparrow \mathfrak{k} & & \uparrow \mathfrak{k} & & \\ & & \overline{Z}_{\mathrm{can}} & \xrightarrow{\mathfrak{p}} & Z_{\mathrm{can}} & \xrightarrow{\mathfrak{s}} & Z. \end{array}$$

Each of the ind-stacks of the diagram projects to $C \times \mathrm{Bun}_G$, and the maps of the diagram commute with the projections.

Let $\mathcal{Z}$ be any one of the ind-stacks from the diagram, and form the diagram

$$\begin{array}{ccccc} \mathcal{Z} & \xleftarrow{h^{\leftarrow}_G} & \mathcal{H}_G \underset{Bun_G \times C}{\times} \mathcal{Z} & \xrightarrow{h^{\rightarrow}_G} & \mathcal{Z} \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{Bun}_G & \xleftarrow{h^{\leftarrow}_G} & \mathcal{H}_G & \xrightarrow{h^{\rightarrow}_G} & \mathrm{Bun}_G \end{array}$$

in which each square is Cartesian.

For $\lambda \in \Lambda^+_G$, we define the convolution functor

$$H^\lambda_G : \mathrm{Sh}(\mathcal{Z}) \to \mathrm{Sh}(\mathcal{Z})$$

on an object $\mathcal{F} \in \mathrm{Sh}(\mathcal{Z})$ to be

$$H^\lambda_G(\mathcal{F}) = h^{\leftarrow}_{G\,!}(\mathcal{A}^\lambda_G \widetilde{\boxtimes} \mathcal{F})^r$$

where $(\mathcal{A}^\lambda_G \widetilde{\boxtimes} \mathcal{F})^r$ is the twisted product defined with respect to $h^{\rightarrow}_G$, and $\mathcal{A}^\lambda_G$ is the simple spherical sheaf on the fibers of $h^{\rightarrow}_G$ corresponding to λ. (See Section 1.4 for more on the twisted product and spherical sheaf.)

5.1. CONVOLUTION ON ${}_\infty\widetilde{\mathrm{Bun}}_P$.

Recall that for a reductive group H, and $\lambda \in \Lambda^+_H$, we write $V^\lambda_{\check{H}}$ for the irreducible representation of the dual group $\check{H}$ of highest weight λ.

We shall deduce our results from the following.

Theorem 5.1.1. [BG, Theorem 4.1.5]. *For $\lambda \in \Lambda^+_G$, there is a canonical isomorphism*

$$H^\lambda_G(\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P}) \simeq \sum_{\mu\in\Lambda^+_M} \mathrm{IC}^{\leq\mu}_{\infty\widetilde{\mathrm{Bun}}_P} \otimes \mathrm{Hom}_{\check{M}}(V^\mu_{\check{M}}, V^\lambda_{\check{G}}).$$

5.2. Convolution on $_\infty\overline{\mathrm{Bun}}_P$. Recall that $r : \Lambda_M \to \Lambda_{M/[M,M]}$ denotes the natural projection, $2\check{\rho}_M$ the sum of the positive roots of M, and $\langle 2\check{\rho}_M, \mu\rangle$ the natural pairing, for $\mu \in \Lambda_M$.

Theorem 5.2.1. *For $\lambda \in \Lambda^+_G$, there is an isomorphism*

$$H^\lambda_G(\mathrm{IC}^{\leq 0}_{\infty\overline{\mathrm{Bun}}_P}) \simeq \sum_{\theta\in\Lambda_{M/[M,M]}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathrm{IC}^{\leq\theta}_{\infty\overline{\mathrm{Bun}}_P} \otimes \mathrm{Hom}_{\check{T}}(V^\mu_{\check{T}}, V^\lambda_{\check{G}})[\langle 2\check{\rho}_M, \mu\rangle].$$

Proof. Step 1. For the projection

$$\mathfrak{r} : {_\infty\widetilde{\mathrm{Bun}}_P} \to {_\infty\overline{\mathrm{Bun}}_P},$$

we clearly have

$$H^\lambda_G(\mathfrak{r}_! \,\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P}) \simeq \mathfrak{r}_! H^\lambda_G(\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P}). \tag{1}$$

Let us first analyze the left hand side of equation 1. We may write the push-forward $\mathfrak{r}_! \,\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P}$ in the form

$$\mathfrak{r}_! \,\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P} \simeq \mathrm{IC}^{\leq 0}_{\infty\overline{\mathrm{Bun}}_P} \oplus \mathcal{J}^{\leq 0}$$

where $\mathcal{J}^{\leq 0} \in \mathrm{Sh}({_\infty\overline{\mathrm{Bun}}_P})$ is isomorphic to a direct sum of shifts of sheaves of the form

$$\mathrm{IC}^{\leq\Theta}_{\infty\overline{\mathrm{Bun}}_P}, \text{ for pairs } \Theta = (0, \mathfrak{U}(\theta^{\mathrm{pos}})), \text{ with } \theta^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P} \setminus \{0\}.$$

The asserted form of $\mathcal{J}^{\leq 0}$ follows from the Decomposition Theorem, the fact that the restrictions of $\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P}$ to the strata of $_\infty\widetilde{\mathrm{Bun}}_P$ are constant [BFGM, Theorem 1.12], and the structure of the map $\mathfrak{r}$ described in Proposition 4.1.1. For any $\eta^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P} \setminus \{0\}$, and decomposition $\mathfrak{U}(\eta^{\mathrm{pos}})$, we have the finite map

$$\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})} : C^{\mathfrak{U}(\eta^{\mathrm{pos}})} \times {_\infty\overline{\mathrm{Bun}}_P} \to {_\infty\overline{\mathrm{Bun}}_P}$$

defined by

$$\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})}(\sum_{m,n} \eta^{\mathrm{pos}}_m \cdot c_{m,n}, (c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}, \sigma))$$
$$= (c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}(-\sum_{m,n} \eta^{\mathrm{pos}}_m \cdot c_{m,n}), \sigma).$$

Note that for $\eta \in \Lambda_{M/[M,M]}$, and Θ the pair $(\eta, \mathfrak{U}(\eta^{\mathrm{pos}}))$, the restriction of $\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})}$ provides an isomorphism

$$\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})} : (C^{\mathfrak{U}(\eta^{\mathrm{pos}})} \times {_\infty\overline{\mathrm{Bun}}^\eta_P})_0 \xrightarrow{\sim} {_\infty\overline{\mathrm{Bun}}^\Theta_P}$$

where the domain completes the Cartesian square

$$\begin{array}{ccc}(C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\times{}_\infty\overline{\mathrm{Bun}}_P^\eta)_0 & \to & C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\times{}_\infty\overline{\mathrm{Bun}}_P^\eta \\ \downarrow & & \downarrow \\ (C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\times C)_0 & \to & C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\times C\end{array}$$

where as usual

$$(C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\times C)_0\subset C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\times C$$

denotes the complement to the diagonal divisor.
We define the strict full triangulated subcategory of irrelevant sheaves

$$\mathrm{IrrelSh}({}_\infty\overline{\mathrm{Bun}}_P)\subset\mathrm{Sh}({}_\infty\overline{\mathrm{Bun}}_P)$$

to be that generated by sheaves of the form

$$\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})!}(\mathrm{IC}_C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\boxtimes\mathcal{F})$$

where η^{pos} runs through $\Lambda^{\mathrm{pos}}_{G,P}\setminus\{0\}$, $\mathfrak{U}(\eta^{\mathrm{pos}})$ runs through decompositions of η^{pos}, $\mathrm{IC}_C^{\mathfrak{U}(\eta^{\mathrm{pos}})}$ denotes the intersection cohomology sheaf of $C^{\mathfrak{U}(\eta^{\mathrm{pos}})}$, and $\mathcal{F}$ runs through objects of $\mathrm{Sh}({}_\infty\overline{\mathrm{Bun}}_P)$.

LEMMA 5.2.2. *The sheaf* $\mathcal{I}^{\le 0}$ *is irrelevant.*

Proof. Let Θ be a pair $(\theta,\mathfrak{U}(\theta^{\mathrm{pos}}))$, with $\theta\in\Lambda_{M/[M,M]}$, and $\theta^{\mathrm{pos}}\in\Lambda^{\mathrm{pos}}_{G,P}\setminus\{0\}$. Then we may realize the sheaf $\mathrm{IC}^{\le\Theta}_{{}_\infty\overline{\mathrm{Bun}}_P}$ as the pushforward

$$\mathrm{IC}^{\le\Theta}_{{}_\infty\overline{\mathrm{Bun}}_P}\simeq\tau_{\mathfrak{U}(\theta^{\mathrm{pos}})!}(\mathrm{IC}_C^{\mathfrak{U}(\theta^{\mathrm{pos}})}\boxtimes\mathrm{IC}^\theta_{{}_\infty\overline{\mathrm{Bun}}_P})$$

To see this, we use the isomorphism

$$\tau_{\mathfrak{U}(\theta^{\mathrm{pos}})}:(C^{\mathfrak{U}(\theta^{\mathrm{pos}})}\times{}_\infty\overline{\mathrm{Bun}}^\theta_P)_0\xrightarrow{\sim}{}_\infty\overline{\mathrm{Bun}}^\Theta_P,$$

and the fact that $\tau_{\mathfrak{U}(\theta^{\mathrm{pos}})}$ is finite. □

LEMMA 5.2.3. *If* $\mathcal{E}$ *is an irrelevant sheaf, then* $H^\lambda_G(\mathcal{E})$ *is an irrelevant sheaf.*

Proof. Clearly we have a canonical isomorphism

$$H^\lambda_G(\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})!}(\mathrm{IC}_C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\boxtimes\mathcal{F}))\simeq\tau_{\mathfrak{U}(\eta^{\mathrm{pos}})!}(\mathrm{IC}_C^{\mathfrak{U}(\eta^{\mathrm{pos}})}\boxtimes H^\lambda_G(\mathcal{F})).$$

□

By the preceding lemmas, we may write the left hand side of equation 1 in the form

$$H^\lambda_G(\mathfrak{r}_!\,\mathrm{IC}^{\le 0}_{{}_\infty\widetilde{\mathrm{Bun}}_P})\simeq H^\lambda_G(\mathrm{IC}^{\le 0}_{{}_\infty\overline{\mathrm{Bun}}_P})\oplus H^\lambda_G(\mathcal{I}^{\le 0})\tag{2}$$

where $H^\lambda_G(\mathcal{I}^{\le 0})$ is an irrelevant sheaf.
Let us next analyze the right hand side of equation 1. By Theorem 5.1.1, we have

$$\mathfrak{r}_!H^\lambda_G(\mathrm{IC}^{\le 0}_{{}_\infty\widetilde{\mathrm{Bun}}_P})\simeq\sum_{\mu\in\Lambda^+_M}\mathfrak{r}_!\,\mathrm{IC}^{\le\mu}_{{}_\infty\widetilde{\mathrm{Bun}}_P}\otimes\mathrm{Hom}_{\check{M}}(V^\mu_{\check{M}},V^\lambda_{\check{G}}).$$

Lemma 5.2.4. *For $\mu \in \Lambda_M^+$, we have*

$$\mathfrak{r}_! \operatorname{IC}^{\leq \mu}_{\infty\widetilde{\operatorname{Bun}}_P} \simeq \sum_{\nu \in \Lambda_M} (\operatorname{IC}^{\leq r(\mu)}_{\infty\overline{\operatorname{Bun}}_P} \oplus \mathcal{J}^{\leq \mu}) \otimes \operatorname{Hom}_{\check{T}}(V^{\nu}_{\check{T}}, V^{\mu}_{\check{M}})[\langle 2\check{\rho}_M, \nu\rangle].$$

where $\mathcal{J}^{\leq\mu}$ is isomorphic to a direct sum of shifts of sheaves of the form

$$\operatorname{IC}^{\leq \Theta}_{\infty\overline{\operatorname{Bun}}_P}, \text{ for pairs } \Theta = (\theta, \mathfrak{U}(\theta^{\mathrm{pos}})).$$

Proof. We may form the diagram

$$\begin{array}{ccccc}
{}_{\infty}\widetilde{\operatorname{Bun}}_P & \overset{h^{\leftarrow}_M}{\leftarrow} & \mathcal{H}_M \underset{\operatorname{Bun}_M \times C}{\times} {}_{\infty}\widetilde{\operatorname{Bun}}_P & \overset{h^{\rightarrow}_M}{\rightarrow} & {}_{\infty}\widetilde{\operatorname{Bun}}_P \\
\downarrow & & \downarrow & & \downarrow \\
\operatorname{Bun}_G & \overset{h^{\leftarrow}_M}{\leftarrow} & \mathcal{H}_G & \overset{h^{\rightarrow}_M}{\rightarrow} & \operatorname{Bun}_G
\end{array}$$

in which each square is Cartesian. We define the convolution functor

$$H^{\mu}_M : \operatorname{Sh}({}_{\infty}\widetilde{\operatorname{Bun}}_P) \to \operatorname{Sh}({}_{\infty}\widetilde{\operatorname{Bun}}_P)$$

on an object $\mathcal{F} \in \operatorname{Sh}({}_{\infty}\widetilde{\operatorname{Bun}}_P)$ to be

$$H^{\mu}_M(\mathcal{F}) = h^{\leftarrow}_{M!}(\mathcal{A}^{\mu}_M \widetilde{\boxtimes} \mathcal{F})^r$$

where $(\mathcal{A}^{\mu}_M \widetilde{\boxtimes} \mathcal{F})^r$ is the twisted product defined with respect to $h^{\rightarrow}_M$, and $\mathcal{A}^{\mu}_M$ is the simple spherical sheaf on the fibers of $h^{\rightarrow}_M$ corresponding to μ. Theorem 4.1.3 of [BG] provides a canonical isomorphism

$$H^{\mu}_M(\operatorname{IC}^{\leq 0}_{\infty\widetilde{\operatorname{Bun}}_P}) \simeq \operatorname{IC}^{\leq\mu}_{\infty\widetilde{\operatorname{Bun}}_P}.$$

We also have a commutative diagram

$$\begin{array}{ccc}
{}_{\infty}\widetilde{\operatorname{Bun}}_P & \overset{h^{\leftarrow}_M}{\leftarrow} & \mathcal{H}_M \underset{\operatorname{Bun}_M \times C}{\times} {}_{\infty}\widetilde{\operatorname{Bun}}_P \\
\mathfrak{r}\downarrow & & \downarrow \mathfrak{r}' \\
{}_{\infty}\overline{\operatorname{Bun}}_P & \overset{h^{\leftarrow}_{M/[M,M]}}{\leftarrow} & \mathcal{H}_{M/[M,M]} \underset{\operatorname{Bun}_{M/[M,M]} \times C}{\times} {}_{\infty}\overline{\operatorname{Bun}}_P
\end{array}$$

where the modification map $h^{\leftarrow}_{M/[M,M]}$ is given by

$$h^{\leftarrow}_{M/[M,M]}(\theta, (c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}, \sigma)) = (c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}(-\theta\cdot c), \sigma).$$

We conclude that there is an isomorphism

$$\mathfrak{r}_! \operatorname{IC}^{\leq\mu}_{\infty\widetilde{\operatorname{Bun}}_P} \simeq h^{\leftarrow}_{M/[M,M]!}\mathfrak{r}'_!(\mathcal{A}^{\mu}_M \widetilde{\boxtimes} \operatorname{IC}^{\leq 0}_{\infty\widetilde{\operatorname{Bun}}_P})^r.$$

Now the map $\mathfrak{r}'$ factors into the projection of the left hand factor

$$\mathcal{H}_M \underset{\operatorname{Bun}_M \times C}{\times} {}_{\infty}\widetilde{\operatorname{Bun}}_P \to \mathcal{H}_{M/[M,M]} \underset{\operatorname{Bun}_{M/[M,M]} \times C}{\times} {}_{\infty}\widetilde{\operatorname{Bun}}_P$$

followed by the projection of the right hand factor

$$\mathcal{H}_{M/[M,M]} \underset{\operatorname{Bun}_{M/[M,M]} \times C}{\times} {}_{\infty}\widetilde{\operatorname{Bun}}_P \xrightarrow{\mathfrak{r}} \mathcal{H}_{M/[M,M]} \underset{\operatorname{Bun}_{M/[M,M]} \times C}{\times} {}_{\infty}\overline{\operatorname{Bun}}_P.$$

Thus we have an isomorphism

$$\mathfrak{r}'_!(\mathcal{A}^{\mu}_{M}\widetilde{\boxtimes}\,\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P})^r \simeq \sum_{\nu\in\Lambda_M}(\mathrm{IC}^{\leq 0}_{\infty\overline{\mathrm{Bun}}_P}\oplus\mathcal{I}^{\leq 0})\otimes\mathrm{Hom}_{\check{T}}(V^{\nu}_{\check{T}},V^{\mu}_{\check{M}})[\langle 2\check{\rho}_M,\nu\rangle]$$

where as before

$$\mathfrak{r}_!\,\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P}\simeq\mathrm{IC}^{\leq 0}_{\infty\overline{\mathrm{Bun}}_P}\oplus\mathcal{I}^{\leq 0}$$

where $\mathcal{I}^{\leq 0}$ is isomorphic to a direct sum of shifts of sheaves of the form

$$\mathrm{IC}^{\leq\Theta}_{\infty\overline{\mathrm{Bun}}_P},\ \text{for pairs } \Theta=(0,\mathfrak{U}(\theta^{\mathrm{pos}})),\text{ with } \theta^{\mathrm{pos}}\in\Lambda^{\mathrm{pos}}_{G,P}\setminus\{0\}.$$

Finally, applying the modification $h^{\leftarrow}_{M/[M,M]!}$ with twist $r(\mu)$ to the above isomorphism, we obtain an isomorphism

$$\mathfrak{r}_!\,\mathrm{IC}^{\leq\mu}_{\infty\widetilde{\mathrm{Bun}}_P}\simeq\sum_{\nu\in\Lambda_M}(\mathrm{IC}^{\leq r(\mu)}_{\infty\overline{\mathrm{Bun}}_P}\oplus\mathcal{I}^{\leq\mu})\otimes\mathrm{Hom}_{\check{T}}(V^{\nu}_{\check{T}},V^{\mu}_{\check{M}})[\langle 2\check{\rho}_M,\nu\rangle].$$

Here we write $\mathcal{I}^{\leq\mu}$ for the result of applying the modification $h^{\leftarrow}_{M/[M,M]!}$ with twist $r(\mu)$ to $\mathcal{I}^{\leq 0}$. Clearly the modification $h^{\leftarrow}_{M/[M,M]!}$ takes strata to strata so we conclude that $\mathcal{I}^{\leq\mu}$ is isomorphic to a direct sum of shifts of sheaves of the form

$$\mathrm{IC}^{\leq\Theta}_{\infty\overline{\mathrm{Bun}}_P},\ \text{for pairs } \Theta=(\theta,\mathfrak{U}(\theta^{\mathrm{pos}})).$$

□

Note that the proof actually shows that $\mathcal{I}^{\leq\mu}$ is isomorphic to a direct sum of shifts of sheaves of the form

$$\mathrm{IC}^{\leq\Theta}_{\infty\overline{\mathrm{Bun}}_P},\ \text{for pairs } \Theta=(0,\mathfrak{U}(\theta^{\mathrm{pos}})),\text{ with } \theta^{\mathrm{pos}}\in\Lambda^{\mathrm{pos}}_{G,P}\setminus\{0\},$$

and so in particular is irrelevant, but we shall have no need for this.

Combining the formulas given by Theorem 5.1.1 and the preceding lemma, we may write the right hand side of equation 1 in the form

(3)

$$\mathfrak{r}_! H^{\lambda}_{G}(\mathrm{IC}^{\leq 0}_{\infty\widetilde{\mathrm{Bun}}_P})\simeq\sum_{\theta\in\Lambda_{M/[M,M]}}\ \sum_{\mu\in\Lambda_M,r(\mu)=\theta}\mathrm{IC}^{\leq\theta}_{\infty\overline{\mathrm{Bun}}_P}\otimes\mathrm{Hom}_{\check{T}}(V^{\mu}_{\check{T}},V^{\lambda}_{\check{G}})[\langle 2\check{\rho}_M,\mu\rangle]\oplus\mathcal{J}$$

where $\mathcal{J}$ is isomorphic to a direct sum of shifts of sheaves of the form

$$\mathrm{IC}^{\leq\Theta}_{\infty\overline{\mathrm{Bun}}_P},\ \text{for pairs } \Theta=(\theta,\mathfrak{U}(\theta^{\mathrm{pos}})).$$

Finally, comparing the left hand side (equation 2) and the right hand side (equation 3), and noting that $\mathrm{IC}^{\leq\theta}_{\infty\overline{\mathrm{Bun}}_P}$ is not irrelevant, we conclude that

$$H^{\lambda}_{G}(\mathrm{IC}^{\leq 0}_{\infty\overline{\mathrm{Bun}}_P})\simeq\sum_{\theta\in\Lambda_{M/[M,M]}}\ \sum_{\mu\in\Lambda_M,r(\mu)=\theta}\mathrm{IC}^{\leq\theta}_{\infty\overline{\mathrm{Bun}}_P}\otimes\mathrm{Hom}_{\check{T}}(V^{\mu}_{\check{T}},V^{\lambda}_{\check{G}})[\langle 2\check{\rho}_M,\mu\rangle]\oplus\mathcal{M}$$

where $\mathcal{M}$ is is isomorphic to a direct sum of shifts of sheaves of the form

$$\mathrm{IC}^{\leq\Theta}_{\infty\overline{\mathrm{Bun}}_P},\ \text{for pairs } \Theta=(\theta,\mathfrak{U}(\theta^{\mathrm{pos}})).$$

Step 2. Now we shall show that $\mathcal{M}$ is in fact zero. To do this, we shall show that its restriction to each stratum of ${}_\infty\overline{\mathrm{Bun}}_P$ is zero.
Let Φ be a pair $(\phi, \mathfrak{U}(\phi^{\mathrm{pos}}))$, with $\phi \in \Lambda_{M/[M,M]}$, and $\phi^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P}$. Let $H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})_\Phi$ be the restriction of $H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})$ to the stratum ${}_\infty\overline{\mathrm{Bun}}^\Phi_P$. For $\theta \in \Lambda_{M/[M,M]}$, let $\mathcal{A}^\theta_\Phi$ be the restriction of $\mathrm{IC}^{\leq\theta}_{{}_\infty\overline{\mathrm{Bun}}_P}$ to the stratum ${}_\infty\overline{\mathrm{Bun}}^\Phi_P$, and let $\mathcal{M}_\Phi$ be the restriction of $\mathcal{M}$. Note that by step 1, [BFGM, Theorem 7.3] and Lemma 5.2.5 below, all of the restrictions are locally constant.
We shall calculate $H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})_\Phi$ in two different ways and compare the results.
On the one hand, by Step 1, we have
(4)
$$H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})_\Phi \simeq \sum_{\theta\in\Lambda_{M/[M,M]}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathcal{A}^\theta_\Phi \otimes \mathrm{Hom}_{\check{T}}(V^\mu_{\check{T}}, V^\lambda_{\check{G}})[\langle 2\check{\rho}_M, \mu\rangle] \oplus \mathcal{M}_\Phi$$

On the other hand, let us return to the definition of the convolution, and consider the diagram

$$\begin{array}{ccccc}
{}_\infty\overline{\mathrm{Bun}}_P & \overset{h^{\leftarrow}_G}{\leftarrow} & \mathcal{H}_G \underset{\mathrm{Bun}_G\times C}{\times} {}_\infty\overline{\mathrm{Bun}}^{\leq 0}_P & \overset{h^{\rightarrow}_G}{\rightarrow} & {}_\infty\overline{\mathrm{Bun}}^{\leq 0}_P \\
\downarrow & & \downarrow & & \downarrow \\
\mathrm{Bun}_G & \overset{h^{\leftarrow}_G}{\leftarrow} & \mathcal{H}_G & \overset{h^{\rightarrow}_G}{\rightarrow} & \mathrm{Bun}_G
\end{array}$$

Recall that by definition

$$H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P}) = h^{\leftarrow}_{G\,!}(\mathcal{A}^\lambda_G \widetilde{\boxtimes}\, \mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})^r$$

where $(\mathcal{A}^\lambda_G \widetilde{\boxtimes}\, \mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})^r$ is the twisted product defined with respect to $h^{\rightarrow}_G$, and $\mathcal{A}^\lambda_G$ is the simple spherical sheaf on the fibers of $h^{\rightarrow}_G$ corresponding to λ.
To calculate $H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})_\Phi$, consider the inverse image ${h^{\leftarrow}_G}^{-1}({}_\infty\overline{\mathrm{Bun}}^\Phi_P)$. Projecting along $h^{\rightarrow}_G$, we may decompose the inverse image into a union of locally closed substacks

$${h^{\leftarrow}_G}^{-1}({}_\infty\overline{\mathrm{Bun}}^\Phi_P) \simeq \bigsqcup_{\xi\in R^{\mathrm{pos}}_{G,P}} \mathcal{S}^\lambda_{P,\phi-\xi} \underset{\mathrm{Bun}_P}{\times} {}_\infty\overline{\mathrm{Bun}}^{(\xi,\mathfrak{U}(\phi^{\mathrm{pos}}))}_P.$$

Projecting each piece back along $h^{\leftarrow}_G$, we arrive at a spectral sequence for $H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P})_\Phi$ with E_2 term

$$\sum_{\xi\in R^{\mathrm{pos}}_{G,P}} \sum_{\mu\in\Lambda_M, r(\mu)=\phi-\xi} \mathcal{A}^0_{(\xi,\mathfrak{U}(\phi^{\mathrm{pos}}))} \otimes \mathrm{Hom}_{\check{T}}(V^\mu_{\check{T}}, V^\lambda_{\check{G}})[\langle 2\check{\rho}_M, \mu\rangle]$$

In fact, the spectral sequence degenerates here for reasons of parity, but we shall not need this. What we do need is the following cyclicity.

Lemma 5.2.5. *Let Ψ be a pair $(\psi, \mathfrak{U}(\psi^{\mathrm{pos}}))$, with $\psi \in \Lambda_{M/[M,M]}$, and $\psi^{\mathrm{pos}} \in \Lambda^{\mathrm{pos}}_{G,P}$. Let $\theta \in \Lambda_{M/[M,M]}$. Then $\mathcal{A}^0_{(\psi,\mathfrak{U}(\psi^{\mathrm{pos}}))} \simeq \mathcal{A}^\theta_{(\psi+\theta,\mathfrak{U}(\psi^{\mathrm{pos}}))}$.*

Proof. The modification

$$(c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}, \sigma) \mapsto (c, \mathcal{P}_G, \mathcal{P}_{M/[M,M]}(\theta \cdot c), \sigma).$$

defines an isomorphism ${}_\infty\overline{\mathrm{Bun}}_P \xrightarrow{\sim} {}_\infty\overline{\mathrm{Bun}}_P$ which restricts to an isomorphism

$${}_\infty\overline{\mathrm{Bun}}_P^{(\psi,\mathfrak{U}(\psi^{\mathrm{pos}}))} \xrightarrow{\sim} {}_\infty\overline{\mathrm{Bun}}_P^{(\psi+\theta,\mathfrak{U}(\psi^{\mathrm{pos}}))}.$$

□

We apply the lemma with $\psi = \xi$, $\psi^{\mathrm{pos}} = \phi^{\mathrm{pos}}$, and make the substitution $\theta = \phi - \xi$, to write the E_2 term

$$\sum_{\phi-\theta\in R^{\mathrm{pos}}_{G,P}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathcal{A}^{\theta}_{(\phi,\mathfrak{U}(\phi^{\mathrm{pos}}))} \otimes \mathrm{Hom}_{\check{T}}(V^{\mu}_{\check{T}}, V^{\lambda}_{\check{G}})[\langle 2\check{\rho}_M, \mu\rangle] \tag{5}$$

Comparing our two calculations (equations 4 and 5), we conclude by a dimension count that $\mathcal{M}_\Phi$ must be zero. □

5.3. CONVOLUTION ON $\overline{Z}_{\mathrm{can}}$.

THEOREM 5.3.1. *For $\lambda \in \Lambda^+_G$, there is an isomorphism*

$$H^\lambda_G(\mathrm{IC}^{\leq 0}_{\overline{Z}_{\mathrm{can}}}) \simeq \sum_{\theta\in\Lambda_{M/[M,M]}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathrm{IC}^{\leq\theta}_{\overline{Z}_{\mathrm{can}}} \otimes \mathrm{Hom}_{\check{T}}(V^{\mu}_{\check{T}}, V^{\lambda}_{\check{G}})[\langle 2\check{\rho}_M, \mu\rangle].$$

Proof. By Proposition 3.4.1, for $\theta \in \Lambda_{M/[M,M]}$, we have

$$\mathfrak{k}^* \, \mathrm{IC}^{\leq\theta}_{{}_\infty\overline{\mathrm{Bun}}_P} \simeq \mathrm{IC}^{\leq\theta}_{\overline{Z}_{\mathrm{can}}},$$

Clearly the pullback $\mathfrak{k}^*$ commutes with convolution

$$H^\lambda_G(\mathfrak{k}^* \, \mathrm{IC}^{\leq\theta}_{{}_\infty\overline{\mathrm{Bun}}_P}) \simeq \mathfrak{k}^* H^\lambda_G(\mathrm{IC}^{\leq\theta}_{{}_\infty\overline{\mathrm{Bun}}_P}).$$

Thus by Theorem 5.2.1, we conclude

$$\begin{aligned}
&H^\lambda_G(\mathrm{IC}^{\leq 0}_{\overline{Z}_{\mathrm{can}}}) \\
&\simeq H^\lambda_G(\mathfrak{k}^* \, \mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P}) \\
&\simeq \mathfrak{k}^* H^\lambda_G(\mathrm{IC}^{\leq 0}_{{}_\infty\overline{\mathrm{Bun}}_P}) \\
&\simeq \sum_{\theta\in\Lambda_{M/[M,M]}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathfrak{k}^* \, \mathrm{IC}^{\leq\theta}_{{}_\infty\overline{\mathrm{Bun}}_P} \otimes \mathrm{Hom}_{\check{T}}(V^{\mu}_{\check{T}}, V^{\lambda}_{\check{G}})[\langle 2\check{\rho}_M, \mu\rangle] \\
&\simeq \sum_{\theta\in\Lambda_{M/[M,M]}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathrm{IC}^{\leq\theta}_{\overline{Z}_{\mathrm{can}}} \otimes \mathrm{Hom}_{\check{T}}(V^{\mu}_{\check{T}}, V^{\lambda}_{\check{T}})[\langle 2\check{\rho}_M, \mu\rangle].
\end{aligned}$$

□

5.4. Convolution on Z. Recall the map of coweight lattices

$$q:\Lambda_M \xrightarrow{r} \Lambda_{M/[M,M]} \xrightarrow{p} \Lambda_{A_0}.$$

Theorem 5.4.1. *For $\lambda \in \Lambda_G^+$, there is an isomorphism*

$$H_G^\lambda(\mathrm{IC}_{\overline{Z}}^{\leq 0}) \simeq \sum_{\kappa\in\Lambda_{A_0}} \sum_{\mu\in\Lambda_T, q(\mu)=\kappa} \mathrm{IC}_{\overline{Z}}^{\leq\kappa} \otimes \mathrm{Hom}_{\check{T}}(V_{\check{T}}^\mu, V_{\check{G}}^\lambda)[\langle 2\check{\rho}_M,\mu\rangle].$$

Proof. By Corollary 4.2.2, for $\theta\in\Lambda_{M/[M,M]}$, we have

$$\mathfrak{p}_! \, \mathrm{IC}_{\overline{Z}_{\mathrm{can}}}^{\leq\theta} \simeq \mathrm{IC}_{\overline{Z}_{\mathrm{can}}}^{\leq p(\theta)},$$

By Corollary 4.3.2, for $\kappa\in\Lambda_{A_0}$, we have

$$\mathfrak{s}_! \, \mathrm{IC}_{\overline{Z}_{\mathrm{can}}}^{\leq\kappa} \simeq \mathrm{IC}_{\overline{Z}}^{\leq\kappa}.$$

Clearly the pushforwards $\mathfrak{p}_!$ and $\mathfrak{s}_!$ commute with convolution

$$H_G^\lambda(\mathfrak{s}_!\mathfrak{p}_! \, \mathrm{IC}_{\overline{Z}_{\mathrm{can}}^{\leq 0}}) \simeq \mathfrak{s}_!\mathfrak{p}_! H_G^\lambda(\mathrm{IC}_{\overline{Z}_{\mathrm{can}}^{\leq 0}}).$$

Thus by Theorem 5.3.1, we conclude

$$\begin{aligned}
&H_G^\lambda(\mathrm{IC}_{\overline{Z}}^{\leq 0})\\
&\simeq H_G^\lambda(\mathfrak{s}_!\mathfrak{p}_! \, \mathrm{IC}_{\overline{Z}_{\mathrm{can}}}^{\leq 0})\\
&\simeq \mathfrak{s}_!\mathfrak{p}_! H_G^\lambda(\mathrm{IC}_{\overline{Z}_{\mathrm{can}}}^{\leq 0})\\
&\simeq \sum_{\theta\in\Lambda_{M/[M,M]}} \sum_{\mu\in\Lambda_M, r(\mu)=\theta} \mathfrak{s}_!\mathfrak{p}_! \, \mathrm{IC}_{\overline{Z}_{\mathrm{can}}}^{\leq\theta} \otimes \mathrm{Hom}_{\check{T}}(V_{\check{T}}^\mu, V_{\check{G}}^\lambda)[\langle 2\check{\rho}_M,\mu\rangle]\\
&\simeq \sum_{\kappa\in\Lambda_{A_0}} \sum_{\mu\in\Lambda_T, q(\mu)=\kappa} \mathrm{IC}_{\overline{Z}}^{\leq\kappa} \otimes \mathrm{Hom}_{\check{T}}(V_{\check{T}}^\mu, V_{\check{G}}^\lambda)[\langle 2\check{\rho}_M,\mu\rangle].
\end{aligned}$$

□

6. Complements

For our application [GN], we need a slight modification of our main result. As usual, let X be an affine horospherical G-variety with dense G-orbit $\mathring{X}\subset X$ and generic stabilizer $S\subset G$. Let S^0 be the identity component of S, and let $\pi_0(S)$ be the component group S/S^0.

For a scheme $\mathcal{S}$, we write $C_{\mathcal{S}}$ for the product $\mathcal{S}\times C$. For an $\mathcal{S}$-point $(c,\mathcal{P}_G,\sigma)$ of the ind-stack Z, the section σ defines a reduction of the G-bundle $\mathcal{P}_G$ to an S-bundle $\mathcal{P}'_S$ over an open subscheme $C'_{\mathcal{S}}\subset C_{\mathcal{S}}$ which is the complement $C_{\mathcal{S}}\setminus\mathcal{D}$ of a subscheme $\mathcal{D}\subset C_{\mathcal{S}}$ which is finite and flat over $\mathcal{S}$. By induction, the S-bundle $\mathcal{P}'_S$ defines a $\pi_0(S)$-bundle over $C'_{\mathcal{S}}$. We call this the generic $\pi_0(S)$-bundle associated to the point $(c,\mathcal{P}_G,\sigma)$.

We define $'Z\subset Z$ to be the ind-substack whose $\mathcal{S}$-points $(c,\mathcal{P}_G,\sigma)$ have the property that for every geometric point $s\in\mathcal{S}$, the restriction of the associated generic $\pi_0(S)$-bundle to $\{s\}\times C\subset C_{\mathcal{S}}$ is trivial. It is not difficult (see [GN]) to show that $'Z$ is closed in Z. Observe that we have a short exact sequence

$$0\to\Lambda_{A_0}\to\Lambda_A\to S/S^0\to 0.$$

Thus for $\kappa \in \Lambda_{A_0}$, it makes sense to consider the locally closed substack $'Z^\kappa \subset {}'Z$ and its closure $'Z^{\leq\kappa} \subset {}'Z$. Observe as well that from the fibration $S \to G \to G/S$, we have an exact sequence

$$\pi_1(G) \to \pi_1(\mathring{X}) \to \pi_0(S).$$

Thus for $\lambda \in \Lambda_G^+$, we have the convolution functor

$$H_G^\lambda : \mathrm{Sh}('Z) \to \mathrm{Sh}('Z).$$

The same arguments show that our main result holds equally well in this context.

Theorem 6.0.2. *For $\lambda \in \Lambda_G^+$, there is an isomorphism*

$$H_G^\lambda(\mathrm{IC}_{'Z}^{\leq 0}) \simeq \sum_{\kappa\in\Lambda_{A_0}} \sum_{\mu\in\Lambda_T, q(\mu)=\kappa} \mathrm{IC}_{'Z}^{\leq\kappa} \otimes \mathrm{Hom}_{\check{T}}(V_{\check{T}}^\mu, V_{\check{G}}^\lambda)[\langle 2\check{\rho}_M, \mu\rangle].$$

References

[BFGM] A. Braverman, M. Finkleberg, D. Gaitsgory, and I. Mirković. Intersection cohomology of Drinfeld's compactifications, *Selecta Math. (N.S.)*, vol. 8, 2002, 381–418.

[BG] A. Braverman and D. Gaitsgory, Geometric Eisenstein series. *Invent. Math.*, 150 (2), 2002, 287–384.

[GN] D. Gaitsgory and D. Nadler, Spherical varieties and Langlands duality, To appear in *Moscow Math. J., volume dedicated to Pierre Deligne.*

[Kn] F. Knop, Weylgruppe und Momentabbildung, *Invent. Math.*, vol. 99 (1), 1990, 1–23.

[Pop] V. L. Popov, Contractions of actions of reductive algebraic groups. *Mat. Sb. (N.S.)*, 130(172)(3), 1986, 310–334, 431.

Dennis Gaitsgory
Department of Mathematics
Harvard University
Cambridge
MA 02138
gaitsgde@math.harvard.edu

David Nadler
Department of Mathematics
Northwestern University
Evanston
IL 60208
nadler@math.northwestern.edu

Projective Homogeneous Varieties Birational to Quadrics

Mark L. MacDonald

Received: December 10, 2008
Revised: February 2, 2009

Communicated by Alexander S. Merkurev

Abstract. We will consider an explicit birational map between a quadric and the projective variety $X(J)$ of traceless rank one elements in a simple reduced Jordan algebra J. $X(J)$ is a homogeneous G-variety for the automorphism group $G = \mathrm{Aut}(J)$. We will show that the birational map is a blow up followed by a blow down. This will allow us to use the blow up formula for motives together with Vishik's work on the motives of quadrics to give a motivic decomposition of $X(J)$.

2000 Mathematics Subject Classification: Primary 11E04; Secondary 14E05, 14L30, 14C15
Keywords and Phrases: Motivic decompositions, Sarkisov links, Jordan algebras

Recently Totaro has solved the birational classification problem for a large class of quadrics [To08]. In particular, let ϕ be an r-Pfister form over a field k of characteristic not 2, and $b = \langle b_1, \cdots b_n\rangle$ be a non-degenerate quadratic form with $n \geq 2$.

Proposition 0.1. *[To08, Thm. 6.3] The birational class of the quadric defined by*

$$q = \phi \otimes \langle b_1, \cdots , b_{n-1}\rangle \perp \langle b_n\rangle$$

only depends on the isometry classes of ϕ and $\phi \otimes b$, and not on the choice of diagonalization of b.

The Sarkisov program [Co94] predicts that any birational map between quadrics (in fact between any two Mori fibre spaces) factors as a chain of composites of "elementary links". In 2.16 we will explicitly factor many of Totaro's birational maps into chains of elementary links, and also prove the following theorem.

Theorem 0.2. *For $r = 0, 1, 2$ and $n \geq 3$, or $r = 3$ and $n = 3$, for each of the birational equivalences from Prop. 0.1, there is a birational map which factors into two elementary links, each of which is the blow up of a reduced subscheme followed by a blow down. Furthermore, if $r \neq 1$ or ϕ is not hyperbolic, then the intermediate Mori fibre space of this factorization will be the projective homogeneous variety $X(J)$ of traceless rank one elements in a Jordan algebra J.*

The birational map from a quadric to $X(J)$ will be the codimension 1 restriction of a birational map between projective space and the projective variety V_J of rank one elements of J, first written down by Jacobson [Ja85, 4.26].

0.3 Motivic decompositions. Let G a semisimple linear algebraic group of inner type, and X a projective homogeneous G-variety such that G splits over the function field of X, which is to say, X is generically split (see [PSZ08, 3.6] for a convenient table). Then [PSZ08] gives a direct sum decomposition of the Chow motive $\mathcal{M}(X; \mathbb{Z}/p\mathbb{Z})$ of X. They show that it is the direct sum of some Tate twists of a single indecomposable motive $\mathcal{R}_p(G)$, which generalizes the Rost motive. This work unified much of what was previously known about motivic decompositions of anisotropic projective homogeneous varieties.
In the non-generically split cases less is known. Quadrics are in general not generically split, but much is known by the work of Vishik and others, especially in low dimensions [Vi04].

Theorem 0.4. *(See Thm. 3.6) The motive of the projective quadric defined by the quadratic forms in Prop. 0.1 may be decomposed into the sum, up to Tate twists, of Rost motives and higher forms of Rost motives.*

In the present paper we will use this knowledge of motives of quadrics to produce motivic decompositions for the non-generically split projective homogeneous G-varieties $X(J)$ which appear in Thm. 0.2. The algebraic groups G are of Lie type ${}^2A_{n-1}$, C_n and F_4, and are automorphism groups of simple reduced Jordan algebras of degree ≥ 3. These varieties $X(J)$ come in four different types which we label $r = 0, 1, 2$ or 3, corresponding to the 2^r dimensional composition algebra of the simple Jordan algebra J (see Thm. 2.4 for a description of $X(J)$ as G/P for a parabolic subgroup P).

Theorem 0.5. *(See Thm. 3.12) The motive of $X(J)$ is the direct sum of a higher form of a Rost motive, F_n^r, together with several Tate twisted copies of the Rost motive R^r.*

The $r = 1$ case of this theorem provides an alternate proof of Krashen's motivic equivalence [Kr07, Thm. 3.3]. On the other hand, the $r = 1$ case of this theorem is shown in [SZ08, Thm. (C)] by using Krashen's result (See Remark 3.14).

0.6 Notational conventions. We will fix a base field k of characteristic 0 (unless stated otherwise), and an algebraically closed (equivalently, a separably closed) field extension $\bar{k}$ of k. We only use the characteristic 0 assumption to

show the varieties $X(J)$ and Z_1 are homogeneous. We will assume a *scheme* over k is a separated scheme of finite type over k, and a *variety* will be an irreducible reduced scheme.
For a scheme X over k, $\overline{X} = X \times_k \bar{k}$.
G denotes an algebraic group over k.
a_i are coefficients of the r-Pfister form ϕ over k.
b_i are coefficients of the n-dimensional quadratic form b over k.
q denotes a quadratic form over k, and Q is the associated projective quadric.
$i_W(q)$ is the Witt index of the quadratic form q.
C is a composition algebra (not to be confused with the Lie type C_n), and c_i are elements of C.
J is a Jordan algebra, x is an element of J, and u is an idempotent in J.
$X(J)$, $Q(J,u)$, Z_1 and Z_2 are complete schemes over k defined in Section 2.
F_n^r and R^r are motives defined in Section 3.1 (not to be confused with the Lie type F_4).
$\mathcal{M}(X)$ denotes the motive of a smooth complete scheme X, and $M\{i\}$ denotes the i^{th} Tate twist of the motive M.
The paper is organized as follows. In Section 1 we will recall the terminology and classification of reduced simple Jordan algebras. In Section 2 we describe the variety $X(J)$ and show it is homogeneous. Also we will define the birational map v_2 from a quadric to $X(J)$ and show that it is a Sarkisov link by analyzing its scheme of base points. In Section 3 we deduce motivic decompositions for a class of quadrics, as well as for the indeterminacy locus of v_2 introduced in Section 2. Finally we put these decompositions together to give a motivic decomposition of $X(J)$.

1 Jordan algebras

A *Jordan algebra* over k is a commutative, unital (not necessarily associative) k-algebra J whose elements obey the identity

$$x^2(xy) = x(x^2y) \text{ for all } x, y \in J.$$

A *simple* Jordan algebra is one with no proper ideals. An *idempotent* in J is an element $u^2 = u \neq 0 \in J$. Two idempotents are *orthogonal* if they multiply to zero, and an idempotent is *primitive* if it is not the sum of two orthogonal idempotents in J. For any field extension l/k, we can *extend scalars* to l by taking $J_l = J \otimes_k l$, for example $\bar{J} = J \otimes \bar{k}$. A Jordan algebra has *degree n* if the identity in $\bar{J}$ decomposes into n pairwise orthogonal primitive idempotents over $\bar{k}$. A degree n Jordan algebra is *reduced* if the identity decomposes into n orthogonal primitive idempotents over k.
The classification of reduced simple Jordan algebras of degree ≥ 3 is closely related to the classification of composition algebras. A *composition algebra* over k is a unital k-algebra C together with a non-degenerate quadratic form ϕ on C (called the *norm form*) such that for any $c_1, c_2 \in C$ we have that

$\phi(c_1c_2) = \phi(c_1)\phi(c_2)$. Two composition algebras are isomorphic as k-algebra iff their norm forms are isometric. Every norm form is an r-fold *Pfister form*, which is to say

$$\phi = \langle\langle a_1, \cdots, a_r\rangle\rangle := \langle 1, -a_1\rangle \otimes \cdots \otimes \langle 1, -a_r\rangle.$$

Furthermore, r must be $0, 1, 2$ or 3, and for any such r-fold Pfister form ϕ, there is a composition algebra with ϕ as its norm form and a canonical conjugation map $^- : C \to C$.

Let C be a composition algebra with norm form $\phi = \langle\langle a_1, \cdots, a_r\rangle\rangle$, and let $b = \langle b_1, \cdots, b_n\rangle$ be a non-degenerate quadratic form. Then we can define a reduced Jordan algebra in the following way. Let $\Gamma = \mathrm{diag}(b_1, \cdots, b_n)$, and let $\sigma_b(x) := \Gamma^{-1}\bar{x}^t\Gamma$ define a map from $M_n(C)$ to $M_n(C)$. Then σ_b is an involution (i.e. an anti-homomorphism such that $\sigma_b^2 = \sigma_b$), so we can define $\mathrm{Sym}(M_n(C), \sigma_b)$ to be the commutative algebra of symmetric elements (i.e. elements x such that $\sigma_b(x) = x$). The product structure is defined by $x \circ y = \frac{1}{2}(xy + yx)$, using the multiplication in C. When C is associative (i.e. $r = 0, 1$ or 2) we know $\mathrm{Sym}(M_n(C), \sigma_b)$ is Jordan. For $r = 3$, it is only Jordan when $n \leq 3$, so in what follows we will always impose this condition in the $r = 3$ case.

The Jordan algebra isomorphism class of $\mathrm{Sym}(M_n(C), \sigma_b)$ only depends on the isomorphism classes of b and C, and not on the diagonalization we have chosen for b. The following theorem states that in degrees ≥ 3 these make up all of the reduced Jordan algebras up to isomorphism.

Theorem 1.1. *(Coordinatization [Mc04, 17],[Ja68, p.137]) Let J be a reduced simple Jordan algebra of degree $n \geq 3$. Then there exists a composition algebra C and an n-dimensional quadratic form b such that $J \cong \mathrm{Sym}(M_n(C), \sigma_b)$.*

2 The Sarkisov link

We will define a birational map from a projective quadric to a projective homogeneous variety, $X(J)$, and show it is an elementary link in terms of Sarkisov (see 2.17).

Let $r = 0, 1, 2, 3$ and $n \geq 3$, and if $r = 3$ then $n = 3$. Throughout we will fix a composition algebra C of dimension 2^r over k, and elements $b_i \in k^*$ such that $b = \langle b_1, \cdots, b_n\rangle$ is a non-degenerate quadratic form. Let $J = \mathrm{Sym}(M_n(C), \sigma_b)$ (see Section 1). Then J is a central simple reduced Jordan algebra. Jacobson defined the closed subset $V_J \subset \mathbb{P}J$ of *rank 1* elements of J (he used the terminology *reduced elements*) and showed it is a variety defined over k [Ja85, §4].

2.1 The Veronese map. The following rational map is a generalization of the $r = 0$ case where it is the degree 2 Veronese morphism [Ch06, 3] [Za93, Last page].

$$v_2 : \mathbb{P}(C^n) \dashrightarrow \mathbb{P}J$$
$$[c_1, \cdots, c_n] \mapsto [b_i c_i \bar{c_j}].$$

If the composition algebra is associative (so $r \neq 3$), then the set-theoretic image of v_2 (where it is defined) is precisely V_J. If $r = 3$, then the set-theoretic image of v_2 isn't closed, but its closure is V_J [Ch06, Prop. 4.2]. Note that this map specifies a choice of n orthogonal primitive idempotents, $v_2([0, \cdots, 1, \cdots, 0])$, so it depends on more than just the isomorphism class of J.
Let us restrict the map v_2 to the projective space defined by $c_n \in k1$, and abuse notation by sometimes considering v_2 as a rational map from $\mathbb{P}(C^{n-1} \times k) \dashrightarrow V_J$. This map is an isomorphism on the open subset $U = (c_n \neq 0) \subset \mathbb{P}(C^{n-1} \times k)$ [Ja85, Thm. 4.26], and hence birational. The projective homogeneous variety we will be interested in is $X(J) \subset V_J$ the hyperplane of traceless matrices, which has dimension $2^r(n-1) - 1$.

2.2 The quadric $Q(J, u)$. Define the quadric $Q(J, u) \subset \mathbb{P}(C^{n-1} \times k)$ by

$$\phi \otimes \langle b_1, \cdots b_{n-1} \rangle \perp \langle b_n \rangle = (\sum_{i=1}^{n-1} b_i c_i \bar{c_i}) + b_n c_n^2 = 0.$$

Here ϕ is the norm form of C. The right hand side is simply the trace in V_J, so the restriction of the birational map v_2 to $Q(J, u)$ has image in $X(J)$. We will often further abuse notation and consider v_2 to be the birational map from $Q(J, u)$ to $X(J)$.
Although the definition of $Q(J, u)$ depends on the diagonalization of b, the isomorphism class of $Q(J, u)$ depends only on the isomorphism class of J together with a choice of primitive idempotent u, which we will usually take to be $u = \mathrm{diag}(0, \cdots, 0, 1) \in J$, as we have done above.

Remark 2.3. Since the birational class of $Q(J, u)$ is independent of $u \in J$, we have another proof of Prop. 0.1 when $r \leq 3$, and if $r = 3$ then $n = 3$. For more on this, see 2.16.

For connected algebraic groups G over $\bar{k}$, projective homogeneous G-varieties G/P are classified by conjugacy classes of parabolic subgroups P in G. Furthermore, the conjugacy classes of parabolics are classified by specifying subsets θ of the set Δ of nodes of the Dynkin diagram of G, as in [Ti65, 1.6]. In fact we will use the complement to his notation, so that $\theta = \Delta$ corresponds to a Borel subgroup $P_\Delta = B$, and $\theta = \emptyset$ corresponds to $P_\emptyset = G$. We use the Bourbaki root numberings. G^0 denotes the connected component of the identity in G.

Theorem 2.4. *V_J is the union of two* $\mathrm{Aut}(J)$*-orbits: $X(J)$ and $V_J - X(J)$. Furthermore, we have:*
(r=0): $\overline{X(J)} \cong G/P_\theta$, for $G = \mathrm{Aut}(\bar{J}) \cong \mathrm{SO}(n)$, if $n \neq 4$ then $\theta = \{1\}$, and if $n = 4$ then the Dynkin diagram is two disjoint nodes, where θ is both nodes.

In all cases, these varieties are quadrics.
(r=1): $\overline{X(J)} \cong G^0/P_\theta$, *for* $G = \mathrm{Aut}(\bar{J}) \cong \mathbb{Z}/2 \ltimes \mathrm{PGL}(n)$ *and* $\theta = \{1, n-1\}$, *this is the variety of flags of dimension 1 and codimension 1 linear subspaces in a vector space.*
(r=2): $\overline{X(J)} \cong G/P_\theta$, *for* $G = \mathrm{Aut}(\bar{J}) \cong \mathrm{PSp}(2n)$ *and* $\theta = \{2\}$, *this is the second symplectic Grassmannian.*
(r=3): $\overline{X(J)} \cong G/P_\theta$, *for* $G = \mathrm{Aut}(\bar{J}) \cong F_4$ *and* $\theta = \{4\}$, *this may be viewed as a hyperplane section of the Cayley plane.*

Proof. $\mathrm{Aut}(J)$ acts on V_J, since the rank is preserved by automorphisms. So it is sufficient to prove this theorem for $k = \bar{k}$. Every element of $V_J - X(J)$ is $[u]$ for some rank one idempotent u [Ch06, Prop. 3.8], and $\mathrm{Aut}(J)$ is transitive on rank one idempotents by Jacobson's coordinatization theorem, since the field is algebraically closed [Mc04, 17].
Clearly $X(J)$ is preserved by $\mathrm{Aut}(J)$, since the trace is preserved by automorphisms. All that remains is to show that $\mathrm{Aut}(J)$ is transitive on $X(J)$, which we will do in cases. Consider the $2^{r-1}n(n-1)+n$ dimensional $\mathrm{Aut}(J)$ representation $J = k \oplus J_0$, where J_0 is the subrepresentation of traceless elements in J. In all cases we will show that J_0 is an irreducible $\mathrm{Aut}(J)$ representation, find the highest weight, and show that there is a closed orbit in $\mathbb{P}(J_0)$ which is contained in $X(J)$ and is of the same dimension. Therefore, by uniqueness of the closed orbit, which follows from the irreducibility of J_0, $X(J)$ is the closed orbit.

Case $r = 0$: For simplicity, we will modify the definition of J. Instead of taking $n \times n$ matrices such that $x^t = x$, we will take matrices such that $M^{-1}x^tM = x$ where

$$M = \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix} \text{ for } n = 2m, \text{ and } M = \begin{bmatrix} 0 & I_m & 0 \\ I_m & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ for } n = 2m+1.$$

This change is justified by recalling that any two orthogonal involutions in the same matrix algebra over an algebraically closed field are isomorphic. Now the Lie algebra of derivations $\mathrm{Der}(J) \cong \mathfrak{so}(n)$ is in the more standard form, and we can choose elements of the Cartan subalgebra $\mathfrak{h}$ as diagonal matrices $H_i = E_{i,i} - E_{m+i,m+i}$ as in [FH91, 18]. Following the conventions of [FH91], we have a dual basis $L_i(H_j) = \delta_{ij}$ of $\mathfrak{h}^*$, and we wish to find the highest weight of the representation J_0.
For $n = 2m$, the roots of $\mathfrak{so}(2m)$ are $\pm L_i \pm L_j$ for $1 \le i \ne j \le m$. One can check that the non-zero weights of J_0 are $\pm L_i \pm L_j$ for all i, j. In particular, the element $E_{1,m+1}$ is a weight vector in J_0 for the weight $2L_1$, and the irreducible representation with highest weight $2L_1$ is of the same dimension as J_0. Therefore J_0 is the irreducible representation with highest weight $2L_1$, and since $\mathrm{Aut}(J)$ is simple, there is a unique closed orbit in $\mathbb{P}(J_0)$, and it is the orbit of $E_{1,m+1}$. To determine the dimension of the orbit, we ask which root spaces $\mathfrak{g}_{-\alpha_i}$ in the Lie algebra for the negative simple roots $-\alpha_i$, kill the

weight space of $2L_1$. For $n = 4$, neither root space, for $-\alpha_1 = -L_1 - L_2$ nor $\alpha_2 = -L_1 + L_2$, kills this weight space. For any $n \geq 6$ even, all of the negative simple root spaces kill the weight space $2L_1$ except for the one for $-L_1 + L_2$. In either case the dimension of the parabolic fixing $E_{1,m+1}$ is $2m^2 - 3m + 2$, so the dimension of the orbit is $n-2$. This is the dimension of the closed invariant subset $X(J)$, which must contain a closed orbit. Since there is only one closed orbit, $X(J)$ must be the entire orbit.
A similar analysis may be carried out in the $n = 2m + 1$ case, where again $E_{1,m+1}$ is a weight vector for the highest weight $2L_1$.

Case $r = 1$*:* We have the action of the connected component $\mathrm{Aut}(J)^0 = PGL(n)$ on $J \cong M_n(k)$, acting by conjugation. The induced action of the Lie algebra of derivations $\mathrm{Der}(J) \cong \mathfrak{sl}(n)$ on J_0 is just the adjoint action on $\mathfrak{sl}(n)$. With the standard diagonal Cartan subalgebra, and choice of positive roots dual to $H_i = E_{i,i} - E_{i+1,i+1}$, the highest weight is in the representation J_0 is $2L_1 + L_2 + \cdots + L_{n-1}$ with multiplicity 1. A dimension count shows this representation is irreducible, and the dimension of the parabolic fixing a highest weight vector is $n^2 - 2n + 2$. So the dimension of the unique closed orbit is $2n - 3$, which is the dimension of $X(J)$. Therefore $X(J)$ is the closed orbit.

Case $r = 2$*:* As in the $r = 0$ case, we will change our symplectic involution $\sigma(x) = \bar{x}^t$ to $\sigma_M(x) = M^{-1}x^tM$ for

$$M = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}.$$

Then the Lie algebra of derivations $\mathrm{Der}(J) \cong \mathfrak{sp}(2n)$ is in the standard form, by choosing a Cartan subalgebra of diagonal matrices, with $H_i = E_{i,i} - E_{n+i,n+i}$ and dual basis $L_i \in \mathfrak{h}^*$. The roots of $\mathfrak{sp}(2n)$ are $\pm L_i \pm L_j$ for all i, j, and the non-zero weights of J_0 are $\pm L_i \pm L_j$ for $i \neq j$. In particular, the highest weight is $L_1 + L_2$ in the standard weight ordering of [FH91, p.257]. Comparing dimensions shows that J_0 is irreducible, and the parabolic fixing a highest weight vector is of dimension $2n^2 - 3n + 5$. So the unique closed orbit in $\mathbb{P}(J_0)$ is of dimension $4n - 5$, which is the same as the dimension of $X(J)$. Therefore $X(J)$ is the unique closed orbit.

Case $r = 3$*:* First notice that J_0 is a 26-dimensional non-trivial representation of $F_4 = \mathrm{Aut}(J)$. It is well-known that such a representation is unique, and has a 15-dimensional unique closed orbit in $\mathbb{P}(J_0)$. Since $X(J)$ is a 15-dimensional closed invariant subset, it must be equal to the closed orbit. $\square$

Remark 2.5. Over the complex numbers the varieties with exactly two G-orbits for some semisimple algebraic group G, one of which is of codimension one, have been classified by [Ah86]. The varieties V_J account for most of these.

2.6 Blowing up the base loci

Any birational map of projective varieties over a field can be expressed as a blow up followed a blow down of closed subschemes (Prop. 2.7). In this section

we will show that these closed subschemes, for our birational map from $Q(J,u)$ to $X(J)$, are (usually) smooth varieties, and hence see that the map is an elementary link in terms of Sarkisov.
Given a rational map between projective varieties $f : Y \dashrightarrow X$, we can define the scheme of base points of f as a closed subscheme of Y [Ha77, II. Example 7.17.3].

Proposition 2.7. *Let $f : Y \dashrightarrow X$ be a birational map of projective varieties over a field k with $g : X \dashrightarrow Y$ the inverse birational map. Let Z_Y and Z_X be the schemes of base points of f and g respectively. Then the blow up $\tilde{Y}$ of Y along Z_Y is isomorphic to the blow up $\tilde{X}$ of X along Z_X.*

Proof. Let $U \subset Y$ be the open subset on which f is an isomorphism. Then the graph Γ_f of $f|_U$ is a subset of $U \times f(U) \subset Y \times X$. The closure of Γ_f in $Y \times X$, given the structure of a closed reduced subscheme, is the blow up $\tilde{Y}$ [EH00, Prop. IV.22][1].
Similarly, $\tilde{X}$ is the closure of $\Gamma_g \subset U \times f(U)$. Since the inverse of f on U is g, we have that $\tilde{X}$ and $\tilde{Y}$ are both closures in $Y \times X$ of the same subset of $U \times f(U)$. So they have the same structure as reduced schemes, and hence $\tilde{X} \cong \tilde{Y}$. □

2.8 Indeterminacy locus of v_2. Let Z_1 be the closed *reduced* subscheme associated to the scheme of base points in $Q(J,u)$ of the birational map v_2. We will show that Z_1 is isomorphic to the scheme of base points. We denote by $\mathrm{Aut}(J,u)$ the subgroup of automorphisms of J that fix the primitive idempotent u.

Theorem 2.9. *Z_1 is homogeneous under an action of* $\mathrm{Aut}(J,u)$*.*

Proof. To describe the action we will use the vector space isomorphism $C^{n-1} \cong J_{\frac{1}{2}}(u) = \{x \in J | x \cdot u = \frac{1}{2}x\}$. Here, as above, we take $u = \mathrm{diag}(0, \cdots, 0, 1) = E_{n,n}$. This isomorphism is given by sending an element $c \in C^{n-1}$ to the matrix element in $J_{\frac{1}{2}}(u) \subset M_n(C)$ with n^{th} row equal to $[c, 0]$.
So we have an $\mathrm{Aut}(J,u)$ action on $\mathbb{P}(C^{n-1})$. By considering the defining equations, one see that Z_1 is isomorphic to the reduced subscheme of $\mathbb{P}(J_{\frac{1}{2}}(u))$ defined by the matrix equation $x^2 = 0$. So it is clear that the underlying closed subset is stable under $\mathrm{Aut}(J,u)$.
Finally, to show the action is transitive, it is enough to show it after extending scalars to an algebraically closed field $\bar{k}$. We will use similar arguments as in the proof of Thm. 2.4.

Case $r = 2$: Using the notation from the proof of Thm. 2.4, the roots of the Lie algebra of $\mathrm{Aut}(J,u)$ are $\pm L_i \pm L_j$ for $i, j \leq n-1$ together with $\pm 2L_n$. One can check that the non-zero weights of the representation $J_{\frac{1}{2}}(u)$ are $\pm L_i \pm L_n$ for $i \leq n-1$. A dimension count reveals that $J_{\frac{1}{2}}(u)$ is therefore

[1] They assume Y is affine, but we can drop this assumption since the blow up is determined locally.

an irreducible representation with highest weight $L_1 + L_n$. The only negative simple roots that don't kill a highest weight vector are $L_2 - L_1$ and $-2L_n$, so the dimension of the parabolic subgroup that fixes a point in the unique closed orbit in $\mathbb{P}(J_{\frac{1}{2}}(u))$ is $2n^2 - 5n + 6$. So the dimension of this orbit is $2n - 2$.

To see this is the same as the dimension of Z_1, consider the affine cone $\tilde{Z}_1$ over Z_1 inside $J_{\frac{1}{2}}(u)$. Then consider the Jacobian matrix of the equations given by $\{x_i\bar{x}_j = 0\}$ with respect to the $4(n-1)$ variables: 4 variables for each coordinate $x_i \in C$. The rank of this matrix at any point in the affine cone over Z_1 is $\leq \dim(J_{\frac{1}{2}}(u)) - \dim(\tilde{Z}_1)$, where equality holds if the ideal spanned by the polynomials $\{x_i\bar{x}_j\}$ is radical. By choosing a convenient point, we see that the dimension of Z_1 is at most $2n-2$, which is the dimension of the closed orbit. So if Z_1 contained another $\mathrm{Aut}(J,u)$-orbit, then it would contain another closed orbit. But the closed orbit is unique, and therefore Z_1 is the closed orbit.

Case $r = 3$: It is well known that the $\mathrm{Aut}(J,u) \cong \mathrm{Spin}(9)$ representation given by $J_{\frac{1}{2}}(u)$ for $u = E_{3,3}$ is the 16-dimensional spin representation. The unique closed orbit in $\mathbb{P}(J_{\frac{1}{2}}(u))$ is therefore the 10-dimensional spinor variety. Using a similar argument to the $r = 2$ case, we can show the dimension of Z_1 is at most 10, so by the uniqueness of the closed orbit we can conclude that Z_1 is the closed orbit.

Case $r = 1$: This case is slightly different from the other two because $\mathrm{Aut}(J,u) \cong \mathbb{Z}/2 \ltimes GL(n-1)$ is a disconnected group, and the connected component has *two* closed orbits in $\mathbb{P}(J_{\frac{1}{2}}(u))$. The argument is similar to the $r = 2$ case, except we find that the $\mathfrak{sl}(n-1)$-representation $J_{\frac{1}{2}}(u)$ is the direct sum of the standard representation V with its dual V^*. So the two closed orbits in $\mathbb{P}(J_{\frac{1}{2}}(u))$ are the orbits of weight vectors for the weights $L_1 - L_n$ and $L_n - L_1$, which are the respective closed orbits in $\mathbb{P}V$ and $\mathbb{P}V^*$. Each $\mathfrak{sl}(n-1)$-orbit has dimension $n-2$. Furthermore, the $\mathbb{Z}/2$ part of $\mathrm{Aut}(J,u)$ swaps these two representations, since it acts on matrices as the transpose. So there is a unique closed $\mathrm{Aut}(J,u)$-orbit, and it is of dimension $n-2$.

As in the $r = 2$ case, by considering the rank of the Jacobian at a closed point in $\tilde{Z}_1$, we see that the dimension of Z_1 is at most $n-2$. Since Z_1 is $\mathrm{Aut}(J,u)$-stable, we can conclude that it is the closed orbit. □

Corollary 2.10. *The reduced scheme Z_1 is isomorphic to the scheme of base points of v_2 in $Q(J,u)$.*

Proof. The $r = 0$ case is trivial, since v_2 is a morphism and hence Z_1 is empty. It is sufficient to assume k is algebraically closed.

The other cases follow from the proof of Thm. 2.9, as follows. We can choose a convenient closed point in the scheme of base points, and show that the rank of the Jacobian of the defining polynomials given by $\{v_2(x) = 0\}$ is equal to the codimension. This implies the scheme is smooth at that point (and therefore at all points), so in particular, it is reduced. □

Corollary 2.11. *Over $\bar{k}$, the smooth subscheme Z_1 is isomorphic to the following.*
$(r=0) : \emptyset$
$(r=1) : \mathbb{P}^{n-2} \sqcup \mathbb{P}^{n-2}$
$(r=2) : \mathbb{P}^1 \times \mathbb{P}^{2n-3}$
$(r=3) :$ *The 10-dimensional spinor variety*

Proof. This follows from our representation theoretic understanding of Z_1 from the proof of Thm. 2.9.
There are much more explicit ways of understanding the $r \neq 3$ cases. For example, in the $r=2$ case, if $c=[c_1, \cdots, c_{n-1}] \in \mathbb{P}(M_2(\bar{k})^{n-1})$ is in $\overline{Z_1}$, then the c_i's are rank 1 matrices that have a common non-zero vector in their kernels. This can be used to get an explicit isomorphism with $\mathbb{P}^1 \times \mathbb{P}^{2n-3}$. $\square$

Remark 2.12. These varieties are written in [Za93, Final pages], where it is implicitly suggested that they are the base locus of the rational map v_2.

Remark 2.13. It is shown in [Kr07] that $Z_1 \cong \mathrm{Spec}(k(\sqrt{a_1})) \times_k \mathbb{P}^{n-2}$, where $\langle\langle a_1 \rangle\rangle$ is the norm form associated to C. So the above corollary shows that Z_1 is irreducible over k except for the single case when $r=1$ and C is split.

2.14 Indeterminacy locus of v_2^{-1}. Let Z_2 be the scheme of base points of the inverse birational map $v_2^{-1} : X(J_n) \dashrightarrow Q(J,u)$. We have that $v_2^{-1}([x_{ij}]) = [x_{n,1}, \cdots, x_{n,n}]$, where this is defined.
We will use the notation $J_{n-1} = \mathrm{Sym}(M_{n-1}(C), \sigma_{\langle b_1, \cdots, b_{n-1}\rangle})$, and sometimes $J_n = J$ for emphasis. The isomorphism class of J_{n-1} depends on the choice of primitive idempotent $u = E_{n,n} \in J$, but is otherwise independent of the diagonalization of $\langle b_1, \cdots, b_{n-1}\rangle$.

Lemma 2.15. *The scheme of base points Z_2 is isomorphic to the smooth subvariety $X(J_{n-1})$.*

Proof. The indeterminacy locus of v_2^{-1} is simply the closed subset of matrices in $X(J_n)$ whose bottom row (and therefore right-most column) is zero. In other words, Z_2 is defined by linear polynomials. The ideal of these polynomials is radical, and therefore the scheme Z_2 is reduced. For $n \geq 4$, one sees that Z_2 is isomorphic to $X(J_{n-1})$. For $n=3$, by considering the matrix equation $x^2=0$, we see that the base locus of Z_2 is the quadric defined by $\phi \otimes \langle b_1 \rangle \perp \langle b_2 \rangle = 0$. We will define $X(J_2)$ to be this quadric. $\square$

2.16 The chain between two quadrics

The Sarkisov program [Co94] predicts that any birational map between two Mori fibre spaces X and Y factors into a chain of elementary links between intermediate Mori fibre spaces. An example of such a link (of type II [Co94, 3.4.2]) would be $X \leftarrow W \rightarrow V$ where both morphisms are blow ups of smooth subvarieties, and X and V are projective homogeneous varieties with Picard number 1 (and hence Mori fibre spaces).

Theorem 2.17. *For $r \neq 1$ or C non-split, the birational map v_2 from $Q(J,u)$ to $X(J)$ is an elementary link of type II.*

Proof. We have that Z_1 is irreducible (see Remark 2.13). The blow up of an irreducible smooth subscheme increases the Picard number by 1, and a blow down decreases it by 1. So in this situation, by Lemma 2.15 and Lemma 2.10 we see that $X(J)$ has Picard number 1. So by Prop. 2.7 we have that v_2 is a blow up of a smooth subvariety followed by a blow down to a smooth subvariety, and therefore it is an elementary link of type II. □

Let $b' = \langle b'_1, \cdots, b'_n \rangle$, and $q' = \phi \otimes \langle b'_1, \cdots, b'_{n-1} \rangle \perp \langle b'_n \rangle$. Then Totaro's Prop. 0.1 states that if $\phi \otimes b \cong \phi \otimes b'$, then the quadrics defined by q and q' are birational. By defining the Jordan algebra J' using ϕ and b', we have a birational map v'_2 from $Q(J',u')$ to $X(J')$.

Proof of Thm. 0.2. If $\phi \otimes b \cong \phi \otimes b'$, then the Jordan algebras $J \cong J'$ are isomorphic as algebras ([KMRT98, Prop. 4.2, p. 43], [Ja68, Ch. V.7, p. 210]), and therefore the varieties $X(J) \cong X(J')$ are also isomorphic. So, as noted in Remark 2.3, $Q(J,u)$ is birational to $Q(J',u')$, and moreover by Thm. 2.17 this map is the composition of two elementary links, with intermediate variety $X(J)$. Notice that if C is a split composition algebra (equivalently, ϕ is hyperbolic) then $Q(J,u)$ and $Q(J',u')$ are already isomorphic. □

2.18 Transposition maps. Now we will explicitly factor the birational maps of Roussey ([Ro05]) and Totaro ([To08]), which in general have more than two elementary links. The most basic case they consider, though, is that of *transposition.* This corresponds to finding a birational map between quadrics q and q', where $b'_i = b_i$ for $1 \leq i \leq n-2$, and $b'_{n-1} = b_n$, $b'_n = b_{n-1}$. So b and b' differ by transposing the last two entries. Totaro proves Prop. 0.1 by finding a suitable chain of such transposition maps.

Proposition 2.19. *For $r = 0, 1, 2$ and $n \geq 3$, and if $r = 3$ then $n = 3$, Totaro's transposition map factors as the composite of two elementary links.*

Proof. Let q and q' be as above, and let $J = \mathrm{Sym}(M_n(C_\phi), \sigma_b)$. Then the quadric $(q = 0) = Q(J,u)$ is defined using the idempotent $u = \mathrm{diag}(0, \cdots 0, 1) \in J$ (see 2.2). General rational points on this quadric are elements in $\mathbb{P}(C^{n-1} \times k)$ such that $v_2([c_1, \cdots, c_n]) \in \mathbb{P}J$ has trace zero. Here $c_i \in C$ for $i \neq n$, and $c_n \in k$. The inverse birational map v_2^{-1} simply takes the n^{th} row of the matrix in J.
Then the quadric for $(q' = 0) = Q(J,u')$ can be defined using the idempotent $u' = \mathrm{diag}(0, \cdots, 1, 0) \in J$. General rational points on this quadric are elements in $\mathbb{P}(C^{n-2} \times k \times C)$ such that $v'_2([c'_1, \cdots, c'_n]) \in \mathbb{P}J$ has trace zero, where we use the same Jordan algebra J. Here $c'_i \in C$ for $i \neq n-1$, and $c'_{n-1} \in k$. The inverse birational map $(v'_2)^{-1}$ takes the $n-1^{th}$ row of the matrix in J.
So the composition $(v'_2)^{-1} \circ v_2$ defines a birational map from $Q(J,u)$ to $Q(J,u')$. From Thm. 2.17 this is the composite of two elementary links. So it remains to show this composite is the same as Totaro's transposition map.

To see this, consider the map $(v_2')^{-1} \circ v_2$ over $\bar{k}$, and observe where it sends a general point from $Q(J,u)$. Recall that v_2 sends $[c_1, \cdots, c_n]$ to the matrix $[b_i c_i \overline{c_j}] \in X(J)$, and then taking the $n-1^{th}$ row of this matrix gives us

$$[b_{n-1}c_{n-1}\overline{c_1}, \cdots, b_{n-1}c_{n-1}\overline{c_{n-1}}, b_{n-1}c_{n-1}\overline{c_n}] \in Q(J,u') \subset \mathbb{P}(C^{n-2} \times \bar{k} \times C).$$

After using the isomorphism $\mathbb{P}(C^{n-2} \times k \times C) \cong \mathbb{P}(C^{n-1} \times k)$ to swap the last two coordinates, we can now recognize that this is exactly a map from [To08, Lemma 5.1], where the "multiplication" of elements in C, is $x * y := x\bar{y}$. $\square$

Remark 2.20. We may also view this chain of birational maps as a "weak factorization" in the sense of [AKMW02]. They prove that any birational map between smooth projective varieties can be factored into a sequence of blow ups and blow downs of smooth subvarieties. But a chain of Sarkisov links (of type II) is stronger, because then each blow up is immediately followed by a blow down, and the intermediate varieties are Mori fibre spaces.

3 Motives

For a smooth complete scheme X defined over k, we will denote the Chow motive of X with coefficients in a ring Λ by $\mathcal{M}(X;\Lambda)$, following [EKM08] (see also [Vi04], [Ma68]). We will briefly recall the definition of the category of graded Chow motives with coefficients in Λ.

Let us define the category $\mathcal{C}(k,\Lambda)$. The objects will be pairs (X,i) for X a smooth complete scheme over k, and $i \in \mathbb{Z}$, and the morphisms will be *correspondences*:

$$\mathrm{Hom}_{\mathcal{C}(k,\Lambda)}((X,i),(Y,j)) = \coprod_m \mathrm{CH}_{\dim(X_m)+i-j}(X_m \times_k Y, \Lambda).$$

Here $\{X_m\}$ is the set of irreducible components of X. If $f : X \to Y$ is a morphism of k-schemes, then the graph of f is an element of $\mathrm{Hom}_{\mathcal{C}(k,\Lambda)}((X,0),(Y,0))$. There is a natural composition on correspondences that generalizes the composition of morphisms of schemes.

We denote the *additive completion* of this pre-additive category by $CR(k,\Lambda)$. Its objects are finite direct sums of objects in $\mathcal{C}(k,\Lambda)$, and the morphisms are matrices of morphisms in $\mathcal{C}(k,\Lambda)$. Then $CR(k,\Lambda)$ is the category of *graded correspondences* over k with coefficients in Λ.

Finally, we let $CM(k,\Lambda)$ be the *idempotent completion* of $CR(k,\Lambda)$. Here the objects are pairs (A,e), where A is an object in $CR(k,\Lambda)$ and $e \in \mathrm{Hom}_{CR(k,\Lambda)}(A,A)$ such that $e \circ e = e$. Then the morphisms are

$$\mathrm{Hom}_{CM(k,\Lambda)}((A,e),(B,f)) = f \circ \mathrm{Hom}_{CR(k,\Lambda)}(A,B) \circ e.$$

This is the category of graded Chow motives over k with coefficients in Λ. For any smooth complete scheme X over k, we denote $\mathcal{M}(X) = ((X,0), id_X)$ its

Chow motive, and $\mathcal{M}(X)\{i\} = ((X, i), id_X)$ its i^{th} Tate twist. Any object in $CM(k, \Lambda)$ is the direct summand of a finite sum of motives $\mathcal{M}(X)\{i\}$.
In this section we will describe direct sum motivic decompositions of $Q(J, u)$, Z_1 and finally $X(J)$. A non-degenerate quadratic form q of dimension ≥ 2 defines a smooth projective quadric Q, and we will sometimes write $\mathcal{M}(q) = \mathcal{M}(Q)$.

3.1 Motives of neighbours of multiples of Pfister quadrics

In this section until 3.8 we can assume our base field k is of any characteristic other than 2, and $r \geq 1$ may be arbitrarily large. Given an r-fold Pfister form ϕ and an n-dimensional non-degenerate quadratic form $b = \langle b_1, \cdots, b_n \rangle$ over k we will describe the motivic decomposition of the projective quadric Q defined by

$$q = \phi \otimes \langle b_1, \cdots, b_{n-1} \rangle \perp \langle b_n \rangle.$$

This quadric is dependent on the choice of diagonalization of b. The following is Vishik's motivic decomposition of the quadric defined by $\phi \otimes b$.

Theorem 3.2. *([Vi04, 6.1])*
For $n \geq 1$, there exists a motive F_n^r such that

$$\mathcal{M}(\phi \otimes b) = \bigoplus_{i=0}^{2^r-1} F_n^r\{i\} \oplus \begin{cases} \emptyset & \text{if } n \text{ is even} \\ \mathcal{M}(\phi)\{2^{r-1}(n-1)\} & \text{if } n \text{ is odd.} \end{cases}$$

Vishik uses the notation $F_\phi(\mathcal{M}(b))$ for F_n^r, and calls it a *higher form of* $\mathcal{M}(b)$. It only depends on the isometry classes of ϕ and b.
If ϕ is anisotropic, Rost defined an indecomposable motive R^r such that $\mathcal{M}(\phi)$ is the direct sum of Tate twists of R^r. This is called the Rost motive of ϕ. If ϕ is split, then this motive is no longer indecomposable, but we will still call $R^r = \mathbb{Z} \oplus \mathbb{Z}\{2^{r-1} - 1\}$ the Rost motive. In fact, F_2^r is just the Rost motive of $\phi \otimes b$ (which is similar to a Pfister form). Also note that $F_1^r = 0$.
In particular, for $n \geq 1$, by counting Tate motives one sees that

$$F_n^r|_{\bar{k}} = \bigoplus_{i=0}^{\lfloor \frac{n}{2} \rfloor - 1} (\mathbb{Z}\{2^r i\} \oplus \mathbb{Z}\{2^r(n-1) - 2^r i - 1\}).$$

So the summand has $2\lfloor \frac{n}{2} \rfloor$ Tate motives, which is the same number that $\mathcal{M}(b)|_{\bar{k}}$ has.
A summand M is said to *start at* d if $d = \min\{i | \mathbb{Z}\{i\}$ is a summand of $M_{\bar{k}}\}$. Similarly, a summand M *ends at* d if $d = \max\{i | \mathbb{Z}\{i\}$ is a summand of $M_{\bar{k}}\}$. We will use the following theorem of Vishik. Here $i_W(q)$ denotes the Witt index of the quadratic form q. This is the number of hyperbolic plane summands in q.

Theorem 3.3. *([Vi04, 4.15]) Let P, Q be smooth projective quadrics over k, and $d \geq 0$. Assume that for every field extension E/k, we have that*

$$i_W(p|_E) > d \Leftrightarrow i_W(q|_E) > m.$$

Then there is an indecomposable summand in $\mathcal{M}(P)$ starting at d, and it is isomorphic to a (Tate twisted) indecomposable summand in $\mathcal{M}(Q)$ starting at m.

With this theorem, it becomes straight forward to prove the following motivic decomposition (Thm. 3.6), by translating it into some elementary facts about multiples of Pfister forms. First we will state two lemmas for convenience.

Lemma 3.4. *Let ϕ be an r-fold Pfister form ($r \geq 1$) and let b be an n-dimensional non-degenerate quadratic form ($n \geq 2$). For any $0 \leq d \leq \lfloor \frac{n}{2} \rfloor - 1$, we have $i_W(\phi \otimes b) > 2^r d$ implies $i_W(\phi \otimes b) > 2^r(d+1) - 1$.*

Proof. This follows from the fact that if ϕ is anisotropic then 2^r divides $i_W(\phi \otimes b)$ [Vi04, Lemma 6.2] or [WS77, Thm. 2(c)]. □

Lemma 3.5. *If Q is a smooth projective quadric of dimension N, then for any $0 \leq d \leq N$, an indecomposable summand of $\mathcal{M}(Q)$ starting at d is isomorphic (up to Tate twist) to an indecomposable summand of $\mathcal{M}(Q)$ ending at $N-d$. The same is true for indecomposable summands of F_n^r for any $r \geq 1$ and $n \geq 1$.*

Proof. This is proved in [Vi04, Thm. 4.19] for anisotropic Q, but it is also true for isotropic Q by using [Vi04, Prop. 2.1] to reduce to the anisotropic case. The statement for the motive F_n^r follows easily from its construction. □

Theorem 3.6. *Let ϕ be an r-fold Pfister form ($r \geq 1$), and for non-zero b_i and $n \geq 2$ we let $q = \phi \otimes \langle b_1, \cdots, b_{n-1}\rangle \perp \langle b_n \rangle$ over k of characteristic not 2. Then we have the following motivic decomposition.*

$$\mathcal{M}(q) = F_n^r \oplus \bigoplus_{i=1}^{2^r-1} F_{n-1}^r\{i\} \oplus \begin{cases} \emptyset & \text{if } n \text{ is odd} \\ \bigoplus_{j=1}^{2^{r-1}-1} R^r\{2^{r-1}(n-1)-j\} & \text{if } n \text{ is even.} \end{cases}$$

Proof. We will split the proof into steps, including one step for each of the three summands. We will use the notation $b' = \langle b_1, \cdots, b_{n-1}\rangle$ and $b = b' \perp \langle b_n \rangle$. Note that we can assume that ϕ is anisotropic, because when it is isotropic both sides split into Tate motives, and we get the isomorphism by checking that on the right hand side there is exactly one copy of $\mathbb{Z}\{i\}$ for each $0 \leq i < 2^r(n-1)$.

Step 1: The first summand. To show that F_n^r is isomorphic to a summand of $\mathcal{M}(q)$, we need to show that given an indecomposable summand in F_n^r starting at d, then there is an isomorphic indecomposable summand in $\mathcal{M}(q)$ starting at d. In fact, by Lemma 3.5 it is enough to only consider indecomposable summands starting in the 'first half', which is to say starting at $i < 2^{r-1}(n-1)$. Since the only Tate motives in the first half of $F_n^r|_{\bar{k}}$ are $\mathbb{Z}\{2^r d\}$ for some $0 \leq d \leq \lfloor \frac{n}{2} \rfloor - 1$, by Thm. 3.3 it is enough to show that for each such d and E/k field extension we have $i_W(\phi \otimes b|_E) > 2^r d$ iff $i_W(\phi \otimes b' \perp \langle b_n \rangle|_E) > 2^r d$.
The "if" part is clear. So assume $i_W(\phi \otimes b|_E) > 2^r d$. Then by Lemma 3.4 we know $i_W(\phi \otimes b|_E) \geq 2^r(d+1)$. So the $2^r(d+1)$-dimensional totally isotropic

subspace must intersect the 2^r-1-codimensional subform $\phi\otimes b' \perp \langle b_n\rangle \subset \phi\otimes b$ in dimension at least 2^rd+1. In other words, $i_W(\phi\otimes b' \perp \langle b_n\rangle|_E) > 2^rd$.

Step 2: The second summand. Fix a $1 \le i \le 2^r-1$. As argued in Step 1, we want to show that if $0 \le d \le \lfloor \frac{n}{2} \rfloor - 1$, and if there is an indecomposable summand of F^r_{n-1} starting at 2^rd, then there is an isomorphic indecomposable summand of $\mathcal{M}(q)$ starting at 2^rd+i. By Thm. 3.3 it is enough to show that for any E/k we have $i_W(\phi\otimes b'|_E) > 2^rd$ iff $i_W(\phi\otimes b' \perp \langle b_n\rangle|_E) > 2^rd+i$.

$$\begin{aligned} i_W(\phi\otimes b'|_E) > 2^rd &\Rightarrow i_W(\phi\otimes b' \perp \langle b_n\rangle|_E) > 2^r(d+1)-1 && \text{Lemma 3.4}\\ &\Rightarrow i_W(\phi\otimes b' \perp \langle b_n\rangle|_E) > 2^rd+i && \\ &\Rightarrow i_W(\phi\otimes b') > 2^rd && \text{See below}\end{aligned}$$

The last implication follows since the $\ge 2^rd+2$ dimensional totally isotropic subspace must intersect the codimension 1 subform in dimension at least 2^rd+1. So, by Lemma 3.5, we have shown that $F^r_{n-1}\{i\}$ is isomorphic to a summand of $\mathcal{M}(q)$ for $1 \le i \le 2^r-1$.

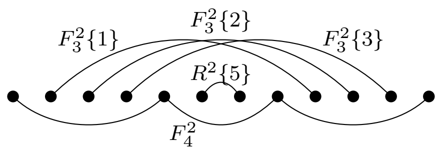
Step 3: The third summand. Assume n is even. Since the summand is empty for $r=1$, we can assume $r \ge 2$. Fix an $2^{r-1}(n-2) < i < 2^{r-1}(n-1)$.

$$\begin{aligned} i_W(\phi) > 0 &\Rightarrow i_W(\phi) = 2^{r-1} && \text{Property of Pfister forms}\\ &\Rightarrow i_W(\phi\otimes b' \perp \langle b_n\rangle) > i && \\ &\Rightarrow i_W(\phi) > 0 && \text{See below}\end{aligned}$$

For the last implication, we have that the hyperbolic part of $\phi\otimes b' \perp \langle b_n\rangle$ is of dimension $\ge 2^r(n-2)+4$. So the anisotropic part is of dimension $\le 2^r-2$. So by the Arason-Pfister hauptsatz, $\phi\otimes b'$ is hyperbolic. Now if ϕ were anisotropic, then $2\dim(\phi)$ would divide $\dim(\phi\otimes b')$ [WS77, Thm. 2(c)]. But this says $2^{r+1}|2^r(n-1)$, which is impossible for n even. Therefore ϕ is isotropic.
To finish Step 3, we use Thm. 3.3 to get the isomorphism of motivic summands.

Step 4: Counting Tate motives. To finish the proof, one needs to show that the summands we have described in these three steps are all possible summands. This can easily be checked by counting the Tate motives over $\bar{k}$. For a visualization of this, see Example 3.7 below.
We have implicitly used [Vi04, Cor. 4.4] here. Note also that for the $n=2$ case the second summand is zero. □

Example 3.7. As an illustration of the counting argument in Step 4 above, consider $r=2$ and $n=4$. Then Thm. 3.6 says that $\mathcal{M}(\langle\langle a_1,a_2\rangle\rangle\otimes\langle b_1,b_2,b_3\rangle \perp \langle b_4\rangle)$ has 5 motivic (possibly decomposable) summands in this decomposition. We can visualize this decomposition, as in [Vi04], with a node for each of the 12 Tate motives over $\bar{k}$, and a line between the nodes if they are in the same summand. Then the motive of the 11-dimensional quadric, $\mathcal{M}(q)$, is as follows, with each summand labelled:



Notice that these summands might be decomposable, for example if the Pfister form $\langle\langle a_1, a_2\rangle\rangle$ is split. So this differs slightly from Vishik's diagrams, since he used solid lines to denote indecomposable summands, and dotted lines for possibly decomposable ones.

3.8 The motive of the base locus Z_1

Now we will use our understanding of Z_1 from Thm. 2.9 and its proof, to decompose its motive into the direct sum of Tate twisted Rost motives.

Proposition 3.9. *(1) For $r = 1$, we have that $\mathcal{M}(Z_1, \mathbb{Z}/2) \cong \oplus_{i=0}^{n-1} R^1\{i\}$*
(2) For $r = 2$, we have that $\mathcal{M}(Z_1, \mathbb{Z}/2) \cong \oplus_{i=0}^{2n-3} R^2\{i\}$.
(3) For $r = 3$, we have that $\mathcal{M}(Z_1, \mathbb{Z}/2) \cong \oplus_{i=0}^{7} R^3\{i\}$.

Proof. For $r = 1$, it is shown in [Kr07] that $Z_1 \cong \mathbb{P}^{n-2} \times_k \mathrm{Spec}(k\sqrt{a_1})$. We know that $\mathcal{M}(\mathrm{Spec}(k[\sqrt{a_1}])) \cong R^1$, so the result follows because the motive of projective space splits into Tate motives.
We have seen that in all cases Z_1 is a smooth scheme that is homogeneous for $\mathrm{Aut}(J, u)$. Moreover, for $r = 2$ or 3, we know that Z_1 is a *generically split* variety in the sense of [PSZ08]. So by their theorem [PSZ08, 5.17] we have that $\mathcal{M}(Z_1, \mathbb{Z}/2)$ is isomorphic to a direct sum of Tate twisted copies of an indecomposable motive $\mathcal{R}_2(\mathrm{Aut}(J, u))$.
Now let V be the projective quadric defined by the r-Pfister form ϕ, the norm form of the composition algebra C. It is a homogeneous $\mathrm{SO}(\phi)$ variety. Since C splits over the function field $k(V)$, by Jacobson's coordinatization theorem J must also split over $k(V)$, and therefore so does the group $\mathrm{Aut}(J, u)$. Furthermore, over $k(Z_1)$, we have a rational point in Z_1. Then for any non-zero coordinate $c_i \in C$ of such a point, there exists $0 \neq y \in C$ such that $c_i y = 0$ in C. But then $\phi(c_i)y = (\bar{c}_i c_i)y = \bar{c}_i(c_i y) = 0$, and so C has an isotropic vector, and is therefore split. Therefore $\mathrm{SO}(\phi)$ splits over $k(Z_1)$.
Now we may apply [PSZ08, Prop. 5.18(iii)] to conclude that $\mathcal{R}_2(\mathrm{Aut}(J, u)) \cong \mathcal{R}_2(\mathrm{SO}(\phi))$. Finally, observe that $\mathcal{R}_2(\mathrm{SO}(\phi))$ is isomorphic to the Rost motive of ϕ ([PSZ08, Last example in 7]), which is the motive R^r. The proposition can be deduced now by counting the Betti numbers of Z_1 (see [Kö91]). $\square$

3.10 Motivic decomposition of $X(J)$

We are ready to decompose the motive $\mathcal{M}(X(J))$ for any reduced simple Jordan algebra J. Recall that $X(J)$ is a homogeneous space for $\mathrm{Aut}(J)$ (Lemma 2.4).

Proposition 3.11. *Let $r = 0, 1, 2$ or 3 and $n \geq 3$, and if $r = 3$ then $n = 3$. We have the following isomorphism of motives with coefficients in $\mathbb{Z}$.*

$$\mathcal{M}(Q(J_n, u)) \oplus \bigoplus_{i=1}^{d_1-1} \mathcal{M}(Z_1)\{i\} \cong \mathcal{M}(X(J_n)) \oplus \bigoplus_{i=1}^{d_2-1} \mathcal{M}(X(J_{n-1}))\{i\}.$$

Here d_i are the respective codimensions of the subschemes Z_i. In particular, for $r \neq 0$, $d_1 = 2^{r-1}n - 2$ and $d_2 = 2^r$.

Proof. If n is the degree of J_n, we have by Section 2.6 that the blow up of $X(J_n)$ along the smooth subvariety $X(J_{n-1})$ is isomorphic to the blow up of $Q(J_n, u)$ along the smooth subscheme Z_1. So by applying the blow up formula for motives [Ma68, p.463], we get the above isomorphism. □

Theorem 3.12. *Let $r = 0, 1, 2$ or 3, and $n \geq 3$ (and if $r = 3$ then $n = 3$). And let $J = \mathrm{Sym}(M_n(C), \sigma_b)$ where C is a 2^r-dimensional composition algebra over k, and $b = \langle b_1, \cdots, b_n \rangle$ is a non-degenerate quadratic form over k. Then*
$(r = 0)$:

$$\mathcal{M}(X(J)) \cong F_n^0 = \mathcal{M}(b),$$

$(r = 1)$:

$$\mathcal{M}(X(J), \mathbb{Z}/2) \cong F_n^1 \oplus \bigoplus_{j=0}^{\lfloor \frac{n-3}{2} \rfloor} \left(\bigoplus_{i=1}^{2\lfloor \frac{n}{2} \rfloor} R^1\{i + 2j\} \right),$$

$(r = 2)$:

$$\mathcal{M}(X(J), \mathbb{Z}/2) \cong F_n^2 \oplus \bigoplus_{j=0}^{\lfloor \frac{n-2}{2} \rfloor} \left(\bigoplus_{i=1}^{4\lfloor \frac{n-1}{2} \rfloor + 1} R^2\{i + 4j\} \right),$$

$(r = 3)$:

$$\mathcal{M}(X(J), \mathbb{Z}/2) \cong F_3^3 \oplus \bigoplus_{i=1}^{11} R^3\{i\}.$$

Proof. The motive of $Q(J, u)$ may be decomposed in terms of the motives F_n^r, F_{n-1}^r and R^r (Thm. 3.6). The motive of Z_1 with $\mathbb{Z}/2$ coefficients may be decomposed in terms of R^r (Prop. 3.9). The subvariety $X(J_2)$ is isomorphic to the quadric defined by $\phi \otimes \langle b_1 \rangle \perp \langle b_2 \rangle$ (see proof of Lemma 2.15), so we have already decomposed its motive in terms of F_2^r and R^r (Thm. 3.6).
So the last ingredient we need is the cancellation theorem. It gives conditions for when it is true that an isomorphism of motives $A \oplus B \cong A \oplus C$ implies an isomorphism of motives $B \cong C$. This does not hold in general; there are counter-examples when $\Lambda = \mathbb{Z}$ [CPSZ06, Remark 2.8]. But if we take Λ to be

any field, then the stronger Krull-Schmidt theorem holds, which says that any motivic decomposition into indecomposables is unique [CM06, Thm. 34][2].
When we put these pieces into the isomorphism from Prop. 3.11, we may proceed by induction on n. One sees that we can cancel the F^r_{n-1} terms in the decomposition, leaving us with the motive $\mathcal{M}(X(J))$ on the right hand side, F^r_n on the left hand side, and several Tate twisted copies of R^r on both sides. To finish the proof one just needs to count the number of copies of R^r remaining after the cancellation theorem, and verify that the given expressions are correct. We leave this induction argument to the reader. □

Remark 3.13. When ϕ is isotropic, the above motives split. When ϕ is anisotropic, R^r is indecomposable, but the motive F^r_n could still be decomposable, depending on the quadratic form b.

Remark 3.14. The $r = 1$ case of the above theorem may be used to prove Krashen's motivic equivalence [Kr07, Thm. 3.3]. To see this, notice that a $1-$Pfister form ϕ defines a quadratic étale extension l/k, and any hermitian form h over l/k is defined by a quadratic form b over k. So in Krashen's notation, $V(h) = X(J)$. Furthermore, his $V(q_h)$ is the projective quadric defined by $\phi \otimes b$, and his $\mathbb{P}_L(N)$ is isomorphic to the base locus Z_1. So in the notation of this paper, his motivic equivalence is

$$\mathcal{M}(\phi \otimes b) \oplus \bigoplus_{i=1}^{n-2} \mathcal{M}(Z_1)\{i\} \cong \mathcal{M}(X(J)) \oplus \mathcal{M}(X(J))\{1\}.$$

Since we have motivic decompositions of all of these summands in terms of F^1_n and R^1 (see Thm. 3.2, Prop. 3.9 and Thm. 3.12), it is easy to verify his motivic equivalence, at least for $\mathbb{Z}/2$ coefficients.
On the other hand, the $r = 1$ case of Thm. 3.12 follows from Krashen's motivic equivalence, together with the $r = 1$ cases of Thm. 3.2 and Prop. 3.9; this is pointed out in [SZ08, Thm. (C)].

Acknowledgements. I would like to thank Burt Totaro for many useful discussions, and for originally noticing the $r = 1$, $n = 3$ case of Prop. 2.19. I would also like to thank Patrick Brosnan, Nikita Semenov and the referee for their comments.

References

[AKMW02] D. Abramovich, K. Karu, K. Matsuki and J. Wlodarczyk. Torification and factorization of birational maps. *J. Amer. Math. Soc.* 15 (2002), 531-572.

[Ah86] D. Ahiezer. Algebraic groups transitive in the complement of a homogeneous hypersurface. *Trans. Moscow Math. Soc.* 48 (1986), 83-103.

[2] Although this theorem is only stated for Λ a discrete valuation ring, the same proof works for any field.

[Al47] A.A. Albert. A structure theory for Jordan algebras. *Ann. of Math. (2)* 48 (1947), 546-567.

[Bo91] A. Borel. *Linear algebraic groups*. Graduate Texts in Mathematics (Springer, 1991).

[CPSZ06] B. Calmès, V. Petrov, N. Semenov and K. Zainoulline. Chow motives of twisted flag varieties. *Compositio Math.* 142 (2006), 1063-1080.

[Ch06] P.-E. Chaput. Geometry over composition algebras: projective geometry. *J. Algebra* 298 (2006), 340-362.

[CM06] V. Chernousov and A. Merkurjev. Motivic decomposition of projective homogeneous varieties and the Krull-Schmidt theorem. *Transform. Groups* 11 (2006), 371-386.

[Co94] A. Corti. Factoring birational maps of threefolds after Sarkisov. *J. Algebraic Geom.* 4 (1994), 223-254.

[De01] O. Debarre. *Higher-dimensional algebraic geometry*. Universitext (Springer, 2001).

[EKM08] R. Elman, N. Karpenko and A. Merkurjev. *The algebraic and geometric theory of quadratic forms*. Colloquium Publications (American Mathematical Society, 2008).

[EH00] D. Eisenbud and J. Harris. *The geometry of schemes*. Graduate Texts in Mathematics (Springer, 2000).

[FH91] W. Fulton and J. Harris. *Representation theory: A first course*. Graduate Texts in Mathematics (Springer, 1991).

[Ha77] R. Hartshorne. *Algebraic geometry*. Graduate Texts in Mathematics (Springer, 1977).

[Ja68] N. Jacobson. *Structure and representations of Jordan algebras*. Amer. Math. Soc. Colloq. Publ. vol. XXXIX (American Mathematical Society, 1968).

[Ja85] N. Jacobson. Some projective varieties defined by Jordan algebras. *J. Algebra* 97 (1985), 565-598.

[KMRT98] M.-A. Knus, A. Merkurjev, M. Rost and J.-P. Tignol. *The book of involutions*. Colloquium Publications (American Mathematical Society, 1998).

[Kö91] B. Köck. Chow motif and higher Chow theory of G/P. *Manuscripta Math.* 70 (1991), no. 4, 363-372.

[Kr07] D. Krashen. Motives of unitary and orthogonal homogeneous varieties. *J. Algebra* 318 (2007), 135-139.

[Mac08] M.L. MacDonald. Cohomological invariants of odd degree Jordan algebras. *Math. Proc. Cambridge Philos. Soc.* 145 (2008), no.2, 295-303.

[Ma68] Ju. I. Manin. Correspondences, motifs and monoidal transformations. *Math. USSR-Sb.* 6 (1968), 439-470.

[Mc04] K. McCrimmon. *A taste of Jordan algebras*. Universitext (Springer, 2004).

[PSZ08] V. Petrov, N. Semenov and K. Zainoulline. J-invariant of linear algebraic groups. *Ann. Sci. Ecole Norm. Sup.* (4) 41 (2008), no.6, 1023-1053.

[Ro05] S. Roussey. Isotropie, corps de fonctions et équivalences birationelles des formes quadratiques. Ph.D. thesis, Université de Franche-Comté (2005).

[SZ08] N. Semenov and K. Zainoulline. Essential dimension of Hermitian spaces. Preprint, arxiv.org/abs/0805.1161

[Se02] J.-P. Serre. *Galois Cohomology*. Graduate Texts in Mathematics (Springer, 2002).

[SV00] T.A. Springer and F.D. Veldkamp. *Octonions, Jordan algebras and exceptional groups*. Springer Monographs in Mathematics (Springer, 2000).

[Ti65] J. Tits. Classification of algebraic semisimple groups. *Algebraic Groups and Discontinuous Subgroups*. Proc. Sympos. Pure Math. IX (Amer. Math. Soc., Providence, R.I., 1966), 32-62.

[To08] B. Totaro. Birational geometry of quadrics. *Bull. Soc. Math. France*, to appear.

[Vi04] A. Vishik. Motives of quadrics with applications to the theory of quadratic forms. *Geometric Methods in the Algebraic Theory of Quadratic Forms. Summer School, Lens 2000* (ed. J.-P. Tignol), Lect. Notes in Math., 1835 (Springer, Berlin, 2004), 25-101.

[WS77] A.R. Wadsworth and D.B. Shapiro. On multiples of round and Pfister forms. *Math. Z.* 157 (1977), 53-62.

[Za93] F.L. Zak. *Tangents and secants of algebraic varieties*. Translations of Math. Monographs, 127 (Amer. Math. Soc., Providence, R.I., 1993).

M.L. MacDonald
Department of Mathematics
The University of British Columbia
Room 121, 1984 Mathematics Road
Vancouver, B.C.
Canada V6T 1Z2
MLM@math.ubc.ca

On the Nonexistence of Certain Morphisms from Grassmannian to Grassmannian in Characteristic 0

Ajay C. Ramadoss

Received: March 17, 2007
Revised: April 8, 2009

Communicated by Takeshi Saito

Abstract. This paper proves some properties of the big Chern classes of a vector bundle on a smooth scheme over a field of characteristic 0. These properties together with the explicit computation of the big Chern classes of universal quotient bundles of Grassmannians are used to prove the main Theorems (Theorems 1,2 and 3) of this paper.

The nonexistence certain morphisms between Grassmannians over a field of characteristic 0 follows directly from these theorems. One of our theorems, for instance, states that the higher Adams operations applied to the class of a universal quotient bundle of a Grassmannian that is not a line bundle yield elements in the K-ring of the Grassmannian that are not representable as classes of genuine vector bundles. This is not true for Grassmannians over a field of characteristic p.

2000 Mathematics Subject Classification: 14M15, 14F99, 19A99
Keywords and Phrases: Chern character, big Chern classes, Grassmannian, universal quotient bundle, Adams operations.

1 Introduction

1.1 Motivation

Problems regarding the constraints that morphisms between homogeneous spaces must satisfy have been studied by Kapil Paranjape and V. Srinivas [7], [8]. In [7], they characterize self maps of finite degree between homogeneous spaces and prove that finite surjective morphisms from Grassmannian to Grassmannian are actually isomorphisms. In [8], they prove that if S is a smooth quadric hypersurface in $\mathbb{P}^{n+1}$, where $n = 2k+1$, and if $2^k | d$, then there exist continuous maps $f : \mathbb{P}^n \to S$ so that $f^*(\mathcal{O}_S(1)) = \mathcal{O}_{\mathbb{P}^n}(d)$. Let $G(r,n)$ denote the Grassmannian of r-dimensional quotient spaces of an n-dimensional vector space over a field of characteristic 0. In the same spirit, given an integer $p \geq 2$, one can ask questions like whether there exists a map from a Grassmannian $G(r,n)$ to another Grassmannian $G(r,M)$ so that $f^*[Q_{G(r,M)}] = \psi^p[Q_{G(r,n)}]$ where $[V]$ denotes the class of a vector bundle V in K-theory and $Q_{G(r,n)}$ and $Q_{G(r,M)}$ denote the universal quotient bundles of $G(r,n)$ and $G(r,M)$ respectively. Another question in the same spirit would be whether there exist morphisms $f : G(r,n) \to G(r-1,M)$ so that $f^*(\mathrm{ch}_l(Q)) = \mathrm{ch}_l(Q)$. The answers to the first question is in the negative for all $r \geq 2, n \geq 2r+1$ and the answer to the second question is in the negative for infinitely many r, with n assumed to be large enough. It may be noted that in these questions, our attention is not restricted solely to dominant/finite morphisms unlike in the results in [7] and [8]. Indeed, the results proven here are not obtainable by the methods of [7] and [8] as far I can see.

1.2 Statements of the results

The following theorems contain the answers obtained for the above questions. These theorems are proven in this paper. Before we proceed, we state that all varieties in this paper are smooth projective varieties over a field of characteristic 0. For any smooth projective variety X, let $K(X)$ denote the K-ring of X. For any vector bundle V on X, let $[V]$ denote the class of V in $K(X) \otimes \mathbb{Q}$.

Theorem 1. *Let Q denote the universal quotient bundle of a Grassmannian $G(r,n)$. Suppose that $r \geq 2$ and that $n \geq 2r+1$. Then, for all $p \geq 2$, the element $\psi^p[Q]$ of $K(G(r,n)) \otimes \mathbb{Q}$ is not equal to $[V]$ for any genuine vector bundle V on $G(r,n)$.*

Corollary 1. *If $f : G(r,n) \to G(r,\infty)$ is a morphism of schemes with $r \geq 2$ and $n \geq 2r+1$, then $f^*[Q_{G(r,\infty)}] \neq \psi^p[Q_{G(r,n)}]$ for any $p \geq 2$.*

Let X be a smooth variety, and let $F_r\mathrm{CH}^l(X) \otimes \mathbb{Q}$ denote the subspace of $\mathrm{CH}^l(X) \otimes \mathbb{Q}$ spanned by $\{\mathrm{ch}_l(V) | V \text{ a vector bundle of rank } \leq r\}$. Then, this filtration is nontrivial as a theory. Let $Q_{G(r,n)}$ denote the universal quotient bundle of $G(r,n)$, and let ch denotes the Chern character map, with ch_l denoting the degree l component of ch.

Theorem 2. *Given any natural number $l \geq 2$, there exist infinitely many natural numbers $r > 0$, and a constant C depending on l so that whenever*

$n > Cr^2 + r$,

$$\mathrm{ch}_l(Q_{G(r,n)}) \in F_r\, CH^l(G(r,n)) \otimes \mathbb{Q} \setminus F_{r-1}\, CH^l(G(r,n)) \otimes \mathbb{Q}.$$

Corollary 2. *Given any natural number $l \geq 2$, there exist infinitely many natural numbers $r > 0$, and a constant C depending on l so that whenever $n > Cr^2 + r$, and $f : G(r,n) \to G(r-1,\infty)$ is a morphism of varieties, then*

$$f^*(\mathrm{ch}_l(Q_{G(r-1,\infty)})) \neq \mathrm{ch}_l(Q_{G(r,n)}).$$

Corollary 3. *There exist infinitely many r so that if $f : G(r,n) \to G(r-1,\infty)$ is any morphism of schemes with $n > 7r^2 + r + 2$, then*

$$f^*\,\mathrm{ch}_2(Q_{G(r-1,\infty)}) = \kappa \mathrm{ch}_1(Q_{G(r,n)})^2$$

for some constant $\kappa \in \mathbb{K}$ that possibly depends on r.

Theorem 3. *If $f : G(3,6) \to G(2,\infty)$ is a morphism, then*

$$f^*(\mathrm{ch}_2(Q_{G(2,\infty)})) = \kappa \mathrm{ch}_1(Q_{G(3,6)})^2$$

for some constant $\kappa \in \mathbb{K}$.

1.3 An outline of the set up of the proofs

All these results are proven using certain facts about certain characteristic classes. These characteristic classes were discovered by M. Kapranov [6] (and independently by M.V. Nori [1]) as far as I know. In this paper, I shall show that these objects are characteristic classes that commute with Adams operations (Lemma 9 and Lemma 13 of Section 4.2 in this paper). These characteristic classes are defined as follows.

Let X be a smooth projective variety and let V be a vector bundle on X. Consider the Atiyah class

$$\theta_V \in \mathrm{H}^1(X, \mathrm{End}(V) \otimes \Omega)$$

of V. Denote the k -fold cup product of θ_V with itself by θ_V^k. Applying the composition map $\mathrm{End}(V)^{\otimes k} \to \mathrm{End}(V)$, followed by the trace map $tr : \mathrm{End}(V) \to \mathcal{O}_X$ to θ_V^k, we obtain the characteristic class

$$\mathrm{t}_k(V) \in \mathrm{H}^k(X, \Omega^{\otimes k}).$$

Note that the projection $\Omega^{\otimes k} \to \wedge^k \Omega$ when applied to $\mathrm{t}_k(V)$ gives us $k!\,\mathrm{ch}_k(V)$ where $\mathrm{ch}_k(V)$ denotes the degree k part of the Chern character of V. The classes t_k are referred to in the paper by Kapranov [1] as the *big Chern classes*. These classes and their properties are discussed in greater detail in Section 4 of this paper. The big Chern classes together give a ring homomorphism

$\oplus\, \mathrm{t}_k : K(X) \otimes \mathbb{Q} \to \oplus\, \mathrm{H}^k(X, \Omega^{\otimes k})$ where the right hand side is equipped with a commutative product that shall be described in the Section 2. The commutative ring $\oplus\, \mathrm{H}^k(X, \Omega^{\otimes k})$ shall henceforth be denoted by $\mathrm{R}(X)$. Both this product and the usual cup product in addition to some other (λ-ring) structure on this ring are preserved under pullbacks. Moreover, the two products are distinct and the Adams gradation on $\mathrm{R}(X)$ is distinct from the obvious one (unlike in the case of the usual cohomology ring). These facts place serious restrictions on what pullback maps $f^* : \mathrm{R}(X) \to \mathrm{R}(Y)$ corresponding to morphisms $f : Y \to X$ look like. An important subring of the ring $\mathrm{R}(X)$ will be calculated explicitly for the Grassmannian $G(r, n)$ at the end of Section 3.

Notation: Throughout this paper, $\mathbb{K}$ shall be used to denote the base field. We assume throughout this paper that the characteristic of $\mathbb{K}$ is zero.

1.4 Brief outlines of the proofs

1.4.1 Outline for Theorems 2 and 3

The basic idea behind the proofs of Theorem 2, Corollary 2 and Theorem 3 is the same.

If $\sigma \in S_k$ is a permutation of $\{1, \ldots, k\}$, and if $\mathcal{F}$ is a vector bundle on X, then σ gives us a homomorphism $\sigma : \mathcal{F}^{\otimes k} \to \mathcal{F}^{\otimes k}$ of $\mathcal{O}_X$ modules. If $f_1, \ldots, f_k$ are sections of $\mathcal{F}$ over an affine open subscheme $\mathrm{Spec}(U)$ of X, then

$$\sigma(f_1 \otimes \cdots \otimes f_k) = f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(k)}.$$

This gives us a right action of S_k on $\mathcal{F}^{\otimes k}$. If $\mathcal{F} = \Omega$, the cotangent bundle of X, then $\sigma : \Omega^{\otimes k} \to \Omega^{\otimes k}$ induces a map $\sigma_* : \mathrm{H}^k(X, \Omega^{\otimes k}) \to \mathrm{H}^k(X, \Omega^{\otimes k})$. Extending this action of S_k on $\mathrm{H}^k(X, \Omega^{\otimes k})$ gives us an endomorphism β_* of $\mathrm{H}^k(X, \Omega^{\otimes k})$ corresponding to each element β of the group ring $\mathbb{K}S_k$ of S_k.

To prove Corollary 2, it suffices to show that for l fixed, there exist infinitely many r such that there is some natural number k with the property that there exists an element β of $\mathbb{K}S_k$ such that

$$\beta_*\, \mathrm{t}_k(\alpha_l(Q_{G(r,n)})) \neq 0$$

and

$$\beta_*\, \mathrm{t}_k(\alpha_l(Q_{G(r-1,\infty)})) = 0.$$

Here $\alpha_l(V) = \mathrm{ch}^{-1} \mathrm{ch}_l(V)$ for any vector bundle V. This is enough because t_k,α_l and β_* commute with pullbacks. If Corollary 2 were to be violated with the above situation being true, we would have something that is 0 [in this case, $\beta_*\, \mathrm{t}_k(\alpha_l(Q_{G(r-1,\infty)}))$] pulling back to something that is nonzero [in this case, $\beta_*\, \mathrm{t}_k(\alpha_l(Q_{G(r,n)}))$]. This gives us a contradiction. A little more work is required to prove Theorem 2.

1.4.2 Outline for Theorem 1

The proof of Theorem 1 is in the same spirit, though much more complicated. We will define a functor of type (k, l) (or a functor of "Adams weight l") to be a map (not necessarily a ring homomorphism/abelian group homomorphism) from $K(X) \otimes \mathbb{Q} \to \mathrm{R}_k(X)$ which takes an element $x \in K(X) \otimes \mathbb{Q}$ to a linear combination of expressions of the form

$$\beta_* \left(\mathrm{t}_{\lambda_1}(\alpha_{l_1}(x)) \cup \cdots \cup \mathrm{t}_{\lambda_s}(\alpha_{l_s}(x))\right)$$

where $\beta \in \mathbb{K}S_k$. If v_l is a functor of type (k, l) then v_l commutes with pullbacks and

$$v_l(\psi^p x) = p^l v_l(x).$$

Corollary 1 will be proven by showing that there is a linear dependence relation

$$\sum_l a_l v_l(Q_{G(r,n)}) = 0$$

for all $n \geq 2r + 1$, with $v_l(Q_{G(r,n)}) \neq 0$, where v_l's are functors of type $(2r, l)$. We will pick a linear dependence relation of this type of shortest length. If Corollary 1 is false, we will obtain yet another linear dependence relation $\sum_l p^l a_l v_l(Q_{G(r,n)}) = 0$, contradicting the fact that the chosen linear dependence relation is of shortest length. A little more work will give us Theorem 1.

Detailed proofs are given in Sections 6 and 7, but the previous sections are required to understand the set up for the proofs. An important ingredient required to flesh-out the proof outlined above is the explicit calculation of $\mathrm{t}_k(Q_{G(r,n)})$. This is done in section 5.

1.5 Remarks about possible future extensions

It can be easily shown that any linear dependence relation between functors of type (k, l) applied to the universal quotient bundle of $G(r, n)$

$$\sum_l a_l v_l(Q_{G(r,n)}) = 0$$

that holds for all n large enough will apply to a vector bundle of rank r on a smooth projective variety X. Thus, if we are able to prove that we have a linear dependence relation

$$\sum_l a_l v_l(Q_{G(r,n)}) = 0$$

for all n large enough with $v_l(V) \neq 0$ then we will be able to apply the same argument to show that in K-Theory, higher Adams operations applied to $[V]$ give us elements not expressible as the class of any genuine vector bundle.

One can try doing this for other homogenous vector bundles in the Grassmannian, and in general, other vector bundles on a G/P space arising out of P-representations, where G is a linear reductive group and P is a parabolic subgroup. This could lead to further progress towards finding the P representations that give rise to vector bundles satisfying Theorem 1. More intricate combinatorics than was used here in this paper may be required for further progress along these lines.

At first sight, it may look that theorem 2 needs to be strengthened. Indeed, on going through the proof, one feels strongly that the filtration F_r of $\mathrm{CH}^l() \otimes \mathbb{Q}$, which theorem 2 says is nontrivial as a theory, is in fact, strictly increasing as a theory. More specifically, I feel that given any $l \geq 2$ fixed, and $r \geq 2$, there exists some Grassmannian $G = G(r, n)$ so that $\mathrm{ch}_l(Q) \in F_r\mathrm{CH}^l(G) \otimes \mathbb{Q} \setminus F_{r-1}\mathrm{CH}^l(G) \otimes \mathbb{Q}$.

One approach to this question is entirely combinatorial (along the lines of the proof to theorems 2 and 3). Let V_λ denote the irreducible representation of S_k corresponding to the partition λ of k. Let $|\lambda|$ denote the number of rows in the Young diagram of λ. The combinatorial approach to this question is to try to show that for some k and a *particular* $\beta \in \mathbb{K}S_k$ depending on l and k only, the subspace spanned by the conjugates of β_{r-1} is of strictly smaller dimension than that spanned by conjugates of β_r. Here, β_i is the image of β under the projection $\mathbb{K}S_k \to \oplus_{|\lambda| \leq i} End(V_\lambda)$. Approaching this question along these lines would indeed involve algebraic combinatorics extensively.

Acknowledgements. This work would not have been possible without the many useful discussions I had with my advisor, Prof. Madhav Nori. It is difficult to convey my gratitude to him. I also received a lot of encouragement (as well as many useful suggestions) from Prof. Spencer Bloch, to whom I am deeply grateful. I am also very grateful to Prof. Shrawan Kumar for pointing out a theorem of Bott [4] used in this work and to Prof. Victor Ginzburg for making me aware of the paper by M. Kapranov [6] where the characteristic classes used are introduced. I thank my friend and colleague Apoorva Khare for helping me LaTeX this work and Dr. Victor Protsak, Prof. Kaan Akin and Prof. Mohan Ramachandran for useful suggestions.

2 The λ-ring $\mathrm{R}(X)$

We recall that a (p, q)-shuffle is a permutation σ of $\{1, 2, \ldots, p+q\}$ such that $\sigma(1) < \cdots < \sigma(p)$ and $\sigma(p+1) < \cdots < \sigma(p+q)$. We denote the set of all

(p,q)-shuffles by $\mathrm{Sh}_{p,q}$ throughout the rest of this work. Also, for the rest of this work, the sign of a permutation σ shall be denoted by $\mathrm{sgn}(\sigma)$.

If $\sigma \in S_k$ is a permutation of $\{1,\ldots,k\}$, and $\mathcal{F}$ a vector bundle on X, then σ gives us a homomorphism $\sigma : \mathcal{F}^{\otimes k} \to \mathcal{F}^{\otimes k}$ of $\mathcal{O}_X$ modules. If $f_1,\ldots,f_k$ are sections of $\mathcal{F}$ over an affine open subscheme $\mathrm{Spec}(U)$ of X, then $\sigma(f_1 \otimes \cdots \otimes f_k) = f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(k)}$.

This gives us a right action of S_k on $\mathcal{F}^{\otimes k}$. If $\mathcal{F} = \Omega$, the cotangent bundle of X, then $\sigma : \Omega^{\otimes k} \to \Omega^{\otimes k}$ induces a map $\sigma_* : \mathrm{H}^k(X, \Omega^{\otimes k}) \to \mathrm{H}^k(X, \Omega^{\otimes k})$.

If $f : Y \to X$ is a morphism of varieties, we have a natural pullback map $\tilde{f}^* : \mathrm{H}^k(X, {\Omega_X}^{\otimes k}) \to \mathrm{H}^k(Y, f^*{\Omega_X}^{\otimes k})$. This can be composed by the map ${\iota^{\otimes k}}_* : \mathrm{H}^k(Y, f^*{\Omega_X}^{\otimes k}) \to \mathrm{H}^k(Y, {\Omega_Y}^{\otimes k})$ to define the pullback $f^* : \mathrm{H}^k(X, {\Omega_X}^{\otimes k}) \to \mathrm{H}^k(Y, {\Omega_Y}^{\otimes k})$, where $\iota : f^*\Omega_X \to \Omega_Y$. We note that

$$f^* \circ \sigma_* = \sigma_* \circ f^*.$$

If $\alpha \in \mathrm{H}^l(X, {\Omega_X}^{\otimes l})$ and $\beta \in \mathrm{H}^m(X, {\Omega_X}^{\otimes m})$, define

$$\alpha \odot \beta := \sum_{\sigma \in \mathrm{Sh}_{l,m}} \mathrm{sgn}(\sigma) \sigma_*^{-1}(\alpha \cup \beta).$$

$\odot$ gives us a product on $\oplus \mathrm{H}^k(X, \Omega^{\otimes k})$. Moreover,

Proposition 1. *If α and β are as in the previous paragraph, then $\alpha \odot \beta = \beta \odot \alpha$. In other words, $\odot$ equips $\mathrm{R}(X)$ with the structure of a commutative ring.*

Proof. If γ is the permutation of $\{1,\ldots,k+l\}$ where $\gamma(i) = l+i$ for $1 \le i \le k$ and $\gamma(i) = i-k$ for $k = 1 \le i \le l+k$, then $\mathrm{sgn}(\gamma) = (-1)^{kl}$. Also, $\sigma \to \sigma \circ \gamma$ gives us a bijection between $\mathrm{Sh}_{l,k}$ and $\mathrm{Sh}_{k,l}$.
Thus

$$\alpha \odot \beta = \sum_{\sigma \in \mathrm{Sh}_{k,l}} \mathrm{sgn}(\sigma)\sigma_*^{-1}(\alpha \cup \beta) = \sum_{\tau \in \mathrm{Sh}_{l,k}} \mathrm{sgn}(\gamma)\,\mathrm{sgn}(\tau)(\tau \circ \gamma)_*^{-1}(\alpha \cup \beta)$$

$$= \sum_{\tau \in \mathrm{Sh}_{l,k}} \mathrm{sgn}(\tau)(\gamma^{-1} \circ \tau^{-1})_*\,\mathrm{sgn}(\gamma)(\alpha \cup \beta) = \sum_{\tau \in \mathrm{Sh}_{l,k}} \mathrm{sgn}(\tau)\tau_*^{-1}(\mathrm{sgn}(\gamma)\gamma_*^{-1}(\alpha \cup \beta))$$

$$= \sum_{\tau \in \mathrm{Sh}_{l,k}} \mathrm{sgn}(\tau)\tau_*^{-1}(\beta \cup \alpha) = \beta \odot \alpha.$$

(Note that $(\gamma^{-1} \circ \tau^{-1})_* = \tau_*^{-1} \circ \gamma_*^{-1}$ since the action of S_{k+l} on $\Omega^{\otimes k+l}$ is a right action). $\square$

We recall from Fulton and Lang [9] that a special λ-ring A is a commutative ring together with operations $\psi^p : A \to A$ indexed by the natural numbers so that

a) ψ^p is a ring homomorphism for all p.
b) $\psi^p \circ \psi^q = \psi^{pq}$.
c) $\psi^1 = \mathrm{id}$.

Here, we show that $\mathrm{R}(X)$ has a special λ-ring structure (i.e, has Adams operations). This is done in Lemma 2. It will be clear from their definition that the Adams operations commute with pullbacks. *The graded tensor co-algebra $T^*\Omega$ of the cotangent bundle Ω_X is a sheaf of graded-commutative Hopf-algebras on X. The product on $T^*\Omega$ and the Adams operations on $T^*\Omega$ therefore induce corresponding operations on the cohomology ring of $T^*\Omega$. Proposition 1 in fact, proves that the ring $\mathrm{R}(X)$ is a subring of the cohomology ring of $T^*\Omega$. It turns out that the Adams operations on the cohomology of $T^*\Omega$ restrict to Adams operations on $\mathrm{R}(X)$ as well.* The rest of this section is devoted to explaining the details of the outline we have just highlighted. We begin with a digression on Hopf-algebras.

2.1 Adams operations on commutative Hopf-algebras

We recall that a Hopf-algebra over a field $\mathbb{K}$ of characteristic 0 is a vector space H together with maps $\mu : H \otimes H \to H$ (multiplication), $\Delta : H \to H \otimes H$ (comultiplication), $u : \mathbb{K} \to H$ (unit) and $c : H \to \mathbb{K}$ (counit) such that the six properties listed below are satisfied.
1. Multiplication is associative and comultiplication is coassociative.
2. Multiplication is a coalgebra homomorphism and comultiplication is an algebra homomorphism.
3. $\mu \circ (u \otimes \mathrm{id}) = \mu \circ (\mathrm{id} \otimes u) = \mathrm{id} : H \to H$.
4. $(\mathrm{id} \otimes c) \circ \Delta = (c \otimes \mathrm{id}) \circ \Delta = \mathrm{id} : H \to H$.
5. u is a coalgebra map and c is an algebra map.
6. $c \circ u = \mathrm{id} : \mathbb{K} \to \mathbb{K}$.

One can define a Hopf algebra in the category of $\mathcal{O}_X$ modules in the same spirit. It is an $\mathcal{O}_X$ module $\mathcal{H}$ together with maps of $\mathcal{O}_X$ modules $\mu : \mathcal{H} \otimes \mathcal{H} \to \mathcal{H}$ (multiplication), $\Delta : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$ (comultiplication), $u : \mathcal{O}_X \to \mathcal{H}$ (unit) and $c : \mathcal{H} \to \mathcal{O}_X$ (counit) such that
1. Multiplication is associative and comultiplication is coassociative.
2. Multiplication is a coalgebra homomorphism and comultiplication is an algebra homomorphism.
3. $\mu \circ (u \otimes \mathrm{id}) = \mu \circ (\mathrm{id} \otimes u) = \mathrm{id} : \mathcal{H} \to \mathcal{H}$.
4. $(\mathrm{id} \otimes c) \circ \Delta = (c \otimes \mathrm{id}) \circ \Delta = \mathrm{id} : \mathcal{H} \to \mathcal{H}$.
5. u is a coalgebra map and c is an algebra map.
6. $c \circ u = \mathrm{id} : \mathcal{O}_X \to \mathcal{O}_X$.

The Hopf algebra $\mathcal{H}$ is said to be (graded) commutative if $\mu \circ \tau = \mu$ where τ is the (signed) swap map from $\mathcal{H} \otimes \mathcal{H}$ to itself. In the graded case $\tau(a \otimes b) = (-1)^{|a||b|} b \otimes a$, where a and b are homogenous sections of $\mathcal{H}$ over an affine open subset of X. $|a|$ and $|b|$ denote the degrees of a and b respectively.

The following four facts are completely analogous to statements in section 4.5.1 of Loday [2]. The checks Loday [2] asks us to do to make these observations for the case of a commutative Hopf algebra over a field also go through in our case, that of a graded commutative Hopf algebra in the category of $\mathcal{O}_X$ modules. These checks are left to the reader as they are fairly simple.

Fact 1. If $\mathcal{H}$ is a (graded) commutative Hopf algebra in the category of $\mathcal{O}_X$ modules, we can define the convolution of two maps $f, g \in \mathrm{End}_{\mathcal{O}_X}(\mathcal{H})$ by

$$f * g = \mu \circ (f \otimes g) \circ \Delta.$$

The convolution product $*$ is an associative product on $\mathrm{End}_{\mathcal{O}_X}(\mathcal{H})$.

Fact 2. If f is an algebra morphism, then if g and h are any $\mathcal{O}_X$ linear maps,

$$f \circ (g * h) = (f \circ g) * (f \circ h).$$

Fact 3. If $\mathcal{H}$ is (graded) commutative and f and g are algebra morphisms, then $f * g$ is an algebra morphism.

Fact 4. It follows from Fact 3 that

$$\psi^k := \mathrm{id} * \cdots * \mathrm{id} \in \mathrm{End}_{\mathcal{O}_X}(\mathcal{H})$$

is an algebra morphism for all natural numbers k. It also follows from Fact 2 that

$$\psi^p \circ \psi^q = \psi^{pq}$$

for all natural numbers p, q.

Further, the following proposition, which is an extension of Proposition 4.5.3 of Loday [2] to graded commutative Hopf algebras in the category of $\mathcal{O}_X$ modules, holds as well. Since the proof of Proposition 4.5.3 of [2] given by Loday [2] goes through in this case with trivial modifications, we omit the proof of the following proposition.

Proposition 2. *If $\mathcal{H} = \oplus_{n \geq 0} \mathcal{H}_n$ is a (graded) commutative Hopf algebra in the category of $\mathcal{O}_X$ modules, then*
a) ψ^p maps $\mathcal{H}_n$ to itself for all p and n.
b) There exist elements $e_n^{(i)}$ of $\mathrm{End}_{\mathcal{O}_X}(\mathcal{H}_n)$ such that

$$\psi^k = \sum_{i=1}^{n} k^i e_n^{(i)}.$$

Further,

$$e_n^{(i)} \circ e_n^{(j)} = \delta_{ij} e_n^{(i)}$$

where δ_{ij} is the Kronecker delta.

An immediate consequence (when $k = 1$) of this proposition is that

$$\mathrm{id} = e_n^{(1)} + \cdots + e_n^{(n)}.$$

The Hopf algebra that is relevant to us is the (graded) tensor co-algebra of a vector bundle $\mathcal{F}$. Here,

$$T^*(\mathcal{F})_n = \mathcal{F}^{\otimes n}$$

$$\Delta(f_1 \otimes \cdots \otimes f_n) = \sum_{0 \le i \le n} f_1 \otimes \cdots \otimes f_i \bigotimes f_{i+1} \otimes \cdots \otimes f_n \in T^*(\mathcal{F}) \otimes T^*(\mathcal{F})$$

(cut coproduct) and

$$\mu(f_1 \otimes \cdots \otimes f_p \bigotimes f_{p+1} \otimes \cdots \otimes f_{p+q}) \\ = \sum_{\sigma \in \mathrm{Sh}_{p,q}} \mathrm{sgn}(\sigma) f_{\sigma^{-1}(1)} \otimes \cdots \otimes f_{\sigma^{-1}(p+q)}$$

where f_i is a section of $\mathcal{F}$ over an affine open subscheme U of X for each i.

We note that in this case,

$$\psi^2(f_1 \otimes \cdots \otimes f_n) = \sum_{p+q=n} \sum_{\sigma \in \mathrm{Sh}_{p,q}} \mathrm{sgn}(\sigma) f_{\sigma^{-1}(1)} \otimes \cdots \otimes f_{\sigma^{-1}(n)}.$$

In this particular case, we also want to find out about the idempotents $e_n^{(i)} \in \mathrm{End}_{\mathcal{O}_X}(\mathcal{F})^{\otimes n}$. The following extension of Proposition 4.5.6 from Loday [2] is what we want. Again, since the proof given in [2] extends with trivial modifications to our case. We therefore, leave the proof of the following proposition to the reader.

Lemma 1.

$$e_n^{(i)} = \sum_{j=1}^{n} a_n^{i,j} l_n^j$$

where

$$\sum_{i=1}^{n} a_n^{i,j} X^i = \binom{X - j + n}{n}$$

and

$$l_n^j = \sum_{\sigma \in S_{n,j}} (\mathrm{sgn}\,\sigma) {\sigma_*}^{-1}.$$

Here, $S_{n,j} = \{\sigma \in S_n | card\{i | \sigma(i) > \sigma(i+1)\} = j - 1\}$.

For example, $e_n^{(n)} = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) {\sigma_*}^{-1}$.

2.2 Description of λ-ring structure on $\mathrm{R}(X)$

Consider the tensor co-algebra $T^*\Omega$. Consider the Adams operations ψ^k on $T^*\Omega$ as described in the previous subsection. Note that $\psi^k|_{\Omega^{\otimes n}}$ induces a map $\psi^k_* : \mathrm{R}_n(X) \to \mathrm{R}_n(X)$. Thus the Adams operation ψ^k induces a map $\psi^k_* : \mathrm{R}(X) \to \mathrm{R}(X)$ that is $\mathbb{K}$-linear. That $\psi^p \circ \psi^q = \psi^{pq}$ implies that $\psi^p_* \circ \psi^q_* = \psi^{pq}_*$. Define the k-th Adams operation on $\oplus \mathrm{H}^n(X, \Omega^{\otimes n})$ to be ψ^k_*. That the Adams operations so defined are ring endomorphisms of $\mathrm{R}(X)$ follows from the fact that the product in $\mathrm{R}(X)$ is induced by the product in $T^*\Omega$. We have therefore, proven the following Lemma.

Lemma 2. $\mathrm{R}(X)$ *is a special λ-ring with Adams operations ψ^p given by ψ^p_*.*

Remark. The Adams operations on $\mathrm{R}(X)$ are thus seen to be defined combinatorially.

3 The ring $\mathrm{R}(G(r,n))^{Gl(n)}$

In this section we explicitly compute an important part of $\mathrm{R}(G(r,n))$, where $G(r,n)$ is the Grassmannian of r dimensional quotients of an n-dimensional vector space. $G(r,n)$ is a homogenous space $Gl(n)/P$ where P is the appropriate parabolic subgroup of $Gl(n)$. Let N denote the unipotent normal subgroup of P.

All the vector bundles that arise during the course of stating and proving the main theorems are $Gl(n)$- equivariant. Thus, the big Chern classes of these vector bundles lie in the part of $\mathrm{R}(G(r,n))$ fixed by $Gl(n)$. If V is an n dimensional vector space, let S be the subspace of V preserved by P and Q the corresponding quotient. The cotangent bundle Ω of the $G(r,n)$ is the vector bundle arising out of the P-representation $Q^* \otimes S$ on which N acts trivially.

Convention. When we refer to Ω in the category of P-representations, we shall refer to the P representation giving rise to the cotangent bundle of $G(r,n)$.

We are now in a position to make the following four observations. Together with the step by step justifications that follow them, these observations describe the method we will use to compute $\mathrm{R}(G(r,n))^{Gl(n)}$ while rigorously justifying our computations at the same time. Observation 1 that follows is a serious statement. We devote the appendix of this paper to sketch its proof. Observations 2 and 3 are first stated "proposition style" and then followed up with proofs. Observation 4 is a sequence of four computations that is crucial to the explicit description of $\mathrm{R}(G(r,n))^{Gl(n)}$ that we provide.

Observation 1. *Let $\mathcal{S}V$ denote the vector bundle on G/P arising out of a P-representation V. Then,* $\mathrm{H}^k(G/P, \mathcal{S}V)^G$ *is isomorphic to* $\mathrm{H}^k(P, V)$.

Here, $\mathrm{H}^k(P, V)$ is in the category of P-modules. This statement follows from a theorem of Bott [4]. Though the base field is the field of complex numbers in [4], an extension of this result to an arbitrary base field of characteristic 0 can be shown using the method of flat descent [11] (Theorem 6 in the appendix to this paper). We sketch a proof of this fact in the appendix to this paper.

Observation 2. *In the case of a Grassmannian,*

$$\mathrm{H}^k(G/P, \mathcal{S}V)^G \cong \mathrm{H}^k(N, V)^{P/N}.$$

Proof. We have the Lyndon-Hochschild-Serre spectral sequence

$$E_2{}^{pq} = \mathrm{H}^p(P/N; \mathrm{H}^q(N; A)) \implies \mathrm{H}^{p+q}(P; A)$$

where A is any P-representation. In the case of a Grassmannian, P/N is isomorphic to $Gl(Q) \times Gl(S)$. The category of P/N -representations is semisimple, and all but the bottom row of the spectral sequence vanish. Thus in the case of a Grassmannian,

$$\mathrm{H}^k(G/P, \mathcal{S}V)^G \cong \mathrm{H}^k(N, V)^{P/N}.$$

□

Observation 3. *From now on $G = Gl(n)$ and P is a parabolic subgroup such that G/P is the Grassmannian $G(r, n)$. Let $\mathcal{N}$ denote the category of N -representations. For any P-representations V and W on which N acts trivially,*

$$\mathrm{Ext}^k_{\mathcal{N}}(W, V) \cong \mathrm{Hom}_{\mathbb{K}}(W \otimes \wedge^k \Omega, V).$$

Proof. We prove the above assertion as follows.

Step 1: Note that N is a Lie group, and in our case (that of a Grassmannian) the exponential map gives a bijection between the Lie-algebra η associated to N and N itself. The category of (finite dimensional) η representations is thus equivalent to a full subcategory of $\mathcal{N}$ in which all our N representations lie. Note that characteristic 0 is needed to formally define the exponential map and its inverse. Also, the category of η-representations is equivalent to the category of $U(\eta)$-representations, where $U(\eta)$ is the universal enveloping algebra of η. Since η is abelian, (in the case of the Grassmannian) $U(\eta) = \mathrm{Sym}^* \eta$. In what follows, we shall work in the category of $\mathrm{Sym}^* \eta$-modules.

Step 2: Consider the Ad action of P on η. The resulting P representation is the P-representation $Q^* \otimes S$ on which N acts trivially. Since co-tangent bundle

Ω of $G(r,n)$ arises out of this P-representation, we abuse notation and denote this P-representation by Ω. For the rest of this section as well as in Sections 5.2 and 5.3, Ω shall denote this P-representation from which the cotangent bundle of $G(r,n)$ arises. As vector spaces, $\eta \simeq \Omega$. As algebras,

$$U(\eta) \simeq \mathrm{Sym}^*(\Omega).$$

Step 3: Note that $\mathrm{Sym}^*(\Omega)$ acts trivially on W. In other words, $y.w = 0$ for any $w \in W$ and any y in the ideal of $\mathrm{Sym}^*(\Omega)$ generated by Ω. Therefore, a projective $\mathrm{Sym}^*(\Omega)$-module resolution of W can be obtained by taking the Koszul complex

$$\ldots \to W \otimes \wedge^k \Omega \otimes \mathrm{Sym}^* \Omega \to W \otimes \wedge^{k-1} \Omega \otimes \mathrm{Sym}^* \Omega \to \ldots \to W \otimes \mathrm{Sym}^* \Omega \to W \to 0.$$

It follows that if V is any other $\mathrm{Sym}^* \Omega$-module, then $\mathrm{Ext}^k(W,V)$ is just the k-th cohomology of the complex

$$0 \to \mathrm{Hom}(W \otimes \mathrm{Sym}^* \Omega, V) \to \ldots \to \ldots \mathrm{Hom}(W \otimes \wedge^k \Omega \otimes \mathrm{Sym}^* \Omega, V) \to \ldots.$$

If V is also a trivial $\mathrm{Sym}^* \Omega$-module, then we see that

$$\mathrm{Hom}(W \otimes \wedge^k \Omega \otimes \mathrm{Sym}^* \Omega, V) = \mathrm{Hom}_{\mathbb{K}}(W \otimes \wedge^k \Omega, V)$$

and the Koszul differential in the previous complex is 0. Thus,

$$\mathrm{Ext}^k_{\mathcal{N}}(W,V) \cong \mathrm{Hom}_{\mathbb{K}}(W \otimes \wedge^k \Omega, V).$$

$\square$

Observation 4. $\mathrm{R}(G(r,n))^{Gl(n)}$ *is isomorphic to a quotient of the group ring* $\mathbb{K}S_k$ *as a* $\mathbb{K}$*-vector space. For the rest of this paper we identify* $\mathrm{R}(G(r,n))^{Gl(n)}$ *with this quotient via a particular isomorphism. An explicit step by step construction of this isomorphism is provided in paragraphs* a).-d). *below.*

a). It follows from Observation 3, Observation 2 and the fact that $P/N \cong Gl(Q) \times Gl(S)$ that

$$\mathrm{H}^k(G(r,n), \Omega^{\otimes k})^{Gl(n)} \cong \mathrm{Hom}_{\mathbb{K}}(\wedge^k \Omega, \Omega^{\otimes k})^{Gl(Q) \times Gl(S)}.$$

We recall from Weyl [10] that if V is any vector space, the map

$$\varphi_V : \mathbb{K}S_k \to \mathrm{End}_{\mathbb{K}}(V^{\otimes k})^{Gl(V)}$$

$$v_1 \otimes \cdots \otimes v_n \mapsto v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}$$

is a surjection. It follows from this that

$$\varphi_{Q^*} \otimes \varphi_S : \mathbb{K}S_k \otimes \mathbb{K}S_k \to (\mathrm{End}_{\mathbb{K}}(Q^{*\otimes k}) \otimes \mathrm{End}_{\mathbb{K}}(S^{\otimes k}))^{Gl(Q) \times Gl(S)}$$

is surjective.

b). Let $i : \wedge^k\Omega \to \Omega^{\otimes k}$ denote the standard inclusion. Let $p : \Omega^{\otimes k} \to \Omega^{\otimes k}$ denote standard projection onto the image of i. Note that $p = \frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega)\omega$. If $\alpha \in (\operatorname{End}_{\mathbb{K}}(Q^{*\otimes k}) \otimes \operatorname{End}_{\mathbb{K}}(S^{\otimes k}))^{Gl(Q)\times Gl(S)}$, then $\alpha \circ i = 0$ iff $\alpha \circ p = 0$. Recall that $\Omega \cong Q^* \otimes S$. Therefore, every element in $\operatorname{Hom}_{\mathbb{K}}(\wedge^k\Omega, \Omega^{\otimes k})^{Gl(Q)\times Gl(S)}$ is the image of a linear combination of elements of the form

$$(\tau\otimes\sigma)\circ\frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega)(\omega\otimes\omega).$$

Also, since we are using the right action of $S_k \times S_k$ on $Q^{*\otimes k}\otimes S^{\otimes k}$,

$$\begin{aligned}(\tau\otimes\sigma)\circ\frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega)(\omega\otimes\omega) &= \frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega)(\omega\otimes\omega)(\tau\otimes\sigma)\\ &= \frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega\sigma)\operatorname{sgn}(\sigma^{-1})(\omega\sigma\otimes\omega\sigma)(\sigma^{-1}\tau\otimes\mathrm{id})\\ &= \frac{1}{k!}\operatorname{sgn}(\sigma)\sum_{\omega\in S_k}\operatorname{sgn}(\omega)(\omega\otimes\omega)(\sigma^{-1}\tau\otimes\mathrm{id}).\end{aligned}$$

c). Identify $\operatorname{End}_{\mathbb{K}}(\Omega^{\otimes k})$ with $(\operatorname{End}_{\mathbb{K}}(Q^{*\otimes k})\otimes\operatorname{End}_{\mathbb{K}}(S^{\otimes k}))$ and think of $S_k\times S_k$ as acting on this with the left copy of S_k permuting the Q^* and the right copy permuting the S. Then, the map p is identified with $\frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega)(\omega\otimes\omega)$. It follows from the above computation that if $\sigma,\tau\in S_k$ then

$$(\sigma\otimes\tau)\circ p = \operatorname{sgn}(\sigma)(\sigma^{-1}\tau\otimes\mathrm{id})\circ p.$$

Therefore, every element in $\operatorname{Hom}_{\mathbb{K}}(\wedge^k\Omega, \Omega^{\otimes k})^{Gl(Q)\times Gl(S)}$ is the image of a linear combination of elements of the form

$$(\sigma^{-1}\tau\otimes\mathrm{id})\circ p.$$

It follows that as a $\mathbb{K}$-vector space, $\operatorname{Hom}_{\mathbb{K}}(\wedge^k\Omega, \Omega^{\otimes k})^{Gl(Q)\times Gl(S)}$ can be identified with a quotient of the group ring $\mathbb{K}S_k$. We shall shortly determine this quotient precisely – but not before making a final computation.

d). Identify Ω with $Q^*\otimes S$. With this identification, if $\sigma\in S_k$, the right action of σ on $\Omega^{\otimes k}$ corresponds to the right action of $\sigma\otimes\sigma$ on $Q^{*\otimes k}\otimes S^{\otimes k}$. Also, if $\beta\in\mathbb{K}S_k$, then

$$\begin{aligned}&\frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega)(\omega\otimes\omega)(\beta\otimes\mathrm{id})(\sigma\otimes\sigma)\\ &\qquad\qquad= \frac{1}{k!}\sum_{\omega\in S_k}\operatorname{sgn}(\omega\sigma)\operatorname{sgn}(\sigma)(\omega\sigma\otimes\omega\sigma)(\sigma^{-1}\beta\sigma\otimes\mathrm{id})\end{aligned}$$

$$= \frac{1}{k!}\,\mathrm{sgn}(\sigma)\sum_{\omega\in S_k}\mathrm{sgn}(\omega)(\omega\otimes\omega)(\sigma^{-1}\beta\sigma\otimes\mathrm{id}).$$

The main result of this section. Henceforth $\mathrm{B}(G(r,n))$ shall denote $\mathrm{R}(G(r,n))^{Gl(n)}$. Observations 1-4 above enable us to conclude that $\mathrm{B}(G(r,n))$ is isomorphic to a quotient of $\mathbb{K}S_k$ as a $\mathbb{K}$-vector space.

We need to specify which quotient of $\mathbb{K}S_k$ gives $\mathrm{B}(G(r,n))$. Recall that the irreducible representations of S_k over $\mathbb{C}$ can be realized over $\mathbb{Q}$ and hence over any field of characteristic 0. We also recall that the irreducible representations of S_k are indexed by partitions λ of k. They are self-dual, and $V_\lambda\otimes Alt = V_{\bar\lambda}$, where $\bar\lambda$ is the partition conjugate to λ. Note that $\mathbb{K}S_k$ is isomorphic to $\oplus_\lambda \mathrm{End}(V_\lambda)$.

Notation. Let $|\lambda|$ denote the rank (number of summands) of the partition λ. Let P_r denote the projection from $\mathbb{K}S_k$ to $\oplus_{|\lambda|\le r}\mathrm{End}(V_\lambda)$ for $1\le r\le k$, and let $P_{r,n}$ denote the projection from $\mathbb{K}S_k$ to $\oplus_{|\lambda|\le r,|\bar\lambda|\le n-r}\mathrm{End}(V_\lambda)$. If n is large enough, $P_{r,n}=P_r$.

The main result in this section is the following.

Lemma 3. *1. As a vector space,*

$$\mathrm{B}(G(r,n))\cong\oplus_k P_{r,n}(\mathbb{K}S_k).$$

2. If $\sigma\in S_k$ then

$$\sigma_* P_{r,n}(\beta)=P_{r,n}(\mathrm{sgn}(\sigma)\sigma^{-1}\beta\sigma)\ \forall\ \beta\in\mathbb{K}S_k.$$

3. If $\alpha\in S_k$ and $\beta\in S_l$ then

$$P_{r,n}(\alpha)\cup P_{r,n}(\beta)=P_{r,n}(\alpha\times\beta)$$

where $\alpha\times\beta$ is thought of as an element of S_{k+l} in the obvious fashion.

The second part of this lemma follows from the paragraph d). of Observation 4 in this subsection. The following sequence of lemmas proves the remaining parts of the above lemma.

3.1 A lemma and some corollaries

Lemma 4. *Let G be a finite group and let $\chi: G\to\mathbb{C}^*$ be a 1-dimensional representation of G. Then, if $\beta\in\mathbb{C}(G)$, $\sum_{g\in G}\chi(g)(g\otimes g)(\beta\otimes\mathrm{id})=0$ in $\mathbb{C}(G\times G)=\mathbb{C}(G)\otimes\mathbb{C}(G)$ iff $\beta=0$.*

Proof. If $\beta = 0$ then clearly $\sum_{g\in G}\chi(g)(g\otimes g)(\beta\otimes \mathrm{id}) = 0$. For the implication in the opposite direction, let us see what $\sum_{g\in G}\chi(g)(g\otimes g)$ does to $\mathbb{C}(G\times G) = \oplus\,\mathrm{End}(V_x\otimes V_y)$ where the V_x are the irreducible representations of G. Let e_i be a basis for V_x and let f_j be a basis of V_y. Suppose that $g(e_i) = \sum_{j=1}^{j=\dim(V_x)} g_{ij}^x e_j$ and that $g(f_k) = \sum_{l=1}^{l=\dim(V_y)} g_{kl}^y f_l$ for all $i \in \{1,\dots,\dim(V_x)\}$ and for all $k \in \{1,\dots,\dim(V_y)\}$. Then,

$$\sum_g \chi(g)(g\otimes g)(e_i\otimes f_j) = \sum_g\sum_{k,l} g_{ik}{}^x g_{jl}{}^y \chi(g)(e_k\otimes f_l)$$

$$= \sum_{k,l}(e_k\otimes f_l)(\sum_g \chi(g) g_{ik}{}^x g_{jl}{}^y) = \sum_{k,l}(e_k\otimes f_l)(\sum_g g_{ik}{}^z g_{jl}{}^y)$$

where $V_z = V_x\otimes\chi$.

Note that $\sum_g(g\otimes g)\in \mathrm{End}(V_z\otimes V_y)$ is a G-module homomorphism. In fact, G acts trivially on $(\sum_g g\otimes g).(V_z\otimes V_y)$. Thus, $\frac{1}{|G|}\sum_g(g\otimes g)$ acts as a projection to the trivial part of $V_z\otimes V_y$. Note that $V_z\otimes V_y$ has a contains precisely $\langle\chi_z,\bar{\chi}_y\rangle$ copies of the trivial representation of G. In particular, it contains one copy of the trivial representation of G iff V_z and V_y are dual representations. In that case, the projection to that copy of the trivial representation is given by $v\otimes w\mapsto \frac{1}{d}w(v)\sum e_i\otimes f_i$ where d is the dimension of V_z. Here, $\{e_i\}$ is a basis for V_z and $\{f_i\}$ is the basis for V_y dual to $\{e_i\}$. This tells us that $\sum_g g_{ik}{}^z g_{jl}{}^y = \frac{|G|}{d}\delta_{y\bar{z}}\delta_{ij}\delta_{kl}$.

Therefore, in $\mathrm{End}(V_x\otimes V_y)$, if V_z is not dual to V_y, then $\sum_{g\in G}\chi(g)(g\otimes g) = 0$. Assume that V_z is dual to V_y. Let $\{e_i\}$ be a basis for V_z and let $\{f_i\}$ be the basis of V_y dual to $\{e_i\}$. If $\{\tilde{e}_i\}$ is the basis of V_x corresponding to $\{e_i\}$, then with respect to the ordered basis $\tilde{e}_1\otimes f_1, \tilde{e}_2\otimes f_1,\dots,\tilde{e}_d\otimes f_1, \tilde{e}_1\otimes f_2,\dots,\tilde{e}_d\otimes f_2,\dots,\tilde{e}_1\otimes f_d,\dots,\tilde{e}_d\otimes f_d$ of $V_x\otimes V_y$, $\frac{d}{|G|}\sum_{g\in G}\chi(g)(g\otimes g)$ corresponds to the matrix M such that $M_{ij} = 1$ if $i,j\in\{kd+k+1 | 0\le k\le d-1\}$ and $M_{ij} = 0$ otherwise. On the other hand, $\beta\otimes\mathrm{id}$ in $\mathrm{End}(V_x\otimes V_y)$ is given by a block diagonal matrix each of whose diagonal blocks is the matrix representing β in $\mathrm{End}(V_x)$. This proves the desired lemma.

□

In fact, in the above proof, we have also proven the following lemma.

Lemma 5. *Let G be a finite group, and let $\chi : G\to\mathbb{C}^*$ be a 1-dimensional representation of G. Let V_x and V_y be irreducible representations of G such that $V_x\otimes\chi$ is dual to V_y. Then, if $\beta\in\mathbb{C}(G)$, $\sum_{g\in G}\chi(g)(g\otimes g)(\beta\otimes\mathrm{id}) = 0$ in $\mathrm{End}(V_x\otimes V_y)$ iff $\beta = 0$ in $\mathrm{End}(V_x)$.*

In our problem, the group in question is S_k. We note that these lemmas give us the precise description of $\mathrm{B}(G(r,n)$ when $\mathbb{K} = \mathbb{C}$. Let $\mathbb{S}_\lambda$ denote the Schur-functor associated with the partition λ of k. In other words, if V is any vector space $\mathbb{S}_\lambda(V) = V^{\otimes k} \otimes_{\mathbb{K}S_k} V_\lambda$ where V_λ is the irreducible representation of S_k corresponding to the partition λ. We know that if V is a vector space of rank m, $\mathbb{S}_\lambda(V) = 0$ iff λ has more than m parts. Therefore, if Q has rank r, then $\mathbb{S}_\lambda(Q) = 0$ iff $|\lambda| > r$ and $\mathbb{S}_{\bar\lambda}(S) = 0$ iff $|\bar\lambda| > n-r$. Moreover, if λ and μ are two partitions of k, then $V^{\otimes k} \otimes W^{\otimes k} \otimes_{\mathbb{K}(S_k\times S_k)} V_\lambda \otimes V_\mu = \mathbb{S}_\lambda(V) \otimes \mathbb{S}_\mu(W)$. If $\gamma \in \mathbb{K}(S_k \times S_k) \neq 0$ in $\mathrm{End}(V_\lambda \otimes V_\mu)$, then $\mathbb{K}(S_k \times S_k).\gamma$ contains $V_\lambda \otimes V_\mu$. Therefore, $V^{\otimes k} \otimes W^{\otimes k} \otimes_{\mathbb{K}(S_k\times S_k)} \gamma$ contains $\mathbb{S}_\lambda(V) \otimes \mathbb{S}_\mu(W)$. Lemma 5 therefore says the following when $\mathbb{K} = \mathbb{C}$.

Lemma 6. *If the rank of Q is r and that of S is $n-r$, then*

$$\sum_\sigma \mathrm{sgn}(\sigma)(\sigma \otimes \sigma)(\beta \otimes \mathrm{id}) = 0$$

as an element of $\mathrm{Hom}_{\mathbb{K}}(\Omega^{\otimes k}, \Omega^{\otimes k})$ *iff* $\beta = 0$ *as an element of* $\mathrm{End}(V_\lambda)$ *for all partitions λ such that $|\lambda| \leq r$ and $|\bar\lambda| \leq n-r$.*

Proof. Let $\gamma = \sum_\sigma \mathrm{sgn}(\sigma)(\sigma \otimes \sigma)(\beta \otimes \mathrm{id})$. Then, by Lemma 5, $\gamma = 0$ in $\mathrm{End}(V_\lambda \otimes V_\mu)$ if $\mu \neq \bar\lambda$. Therefore, γ kills $\mathbb{S}_\lambda(Q^*) \otimes \mathbb{S}_\mu(S)$ whenever $\mu \neq \bar\lambda$. On the other hand, if $\gamma \neq 0$ in $\mathrm{End}(V_\lambda \otimes V_{\bar\lambda})$, then $\Omega^{\otimes k}.\gamma$ contains a copy of $\mathbb{S}_\lambda(Q^*) \otimes \mathbb{S}_{\bar\lambda}(S)$. The desired lemma follows immediately. □

Since the irreducible representations of S_k over $\mathbb{C}$ can be realized over $\mathbb{Q}$ and hence over any field of characteristic 0,lemmas 4,5 and 6 thus hold for $\mathbb{K}S_k$ where $\mathbb{K}$ is *any* field of characteristic 0. *This proves the first part of Lemma 3 specifying the vector space structure of* $\mathrm{B}(G(r,n))$. We have so far also identified the right S_k module structure of $\mathrm{B}(G(r,n))$. To describe the ring structure completely, we need to be able to compute cup products explicitly under this identification.

We now show how one computes the cup product of two elements $X_k \in \mathrm{Hom}_{\mathbb{K}}(\wedge^k\Omega, \Omega^{\otimes k}) \subset \mathrm{H}^k(G(r,n), \Omega^{\otimes k})$ and $Y_l \in \mathrm{Hom}_{\mathbb{K}}(\wedge^l\Omega, \Omega^{\otimes l}) \subset \mathrm{H}^l(G(r,n), \Omega^{\otimes l})$. Let $X_k = (\gamma_k \otimes \mathrm{id}) \circ i_k \in \mathrm{End}_{\mathbb{K}}(Q^{*\otimes k}) \otimes \mathrm{End}_{\mathbb{K}}(S^{\otimes k})$ and $Y_l = (\delta_l \otimes \mathrm{id}) \circ i_l \in \mathrm{End}_{\mathbb{K}}(Q^{*\otimes l}) \otimes \mathrm{End}_{\mathbb{K}}(S^{\otimes l})$ where i_k and i_l are the standard inclusions $\wedge^k\Omega \to \Omega^{\otimes k}$ and $\wedge^l\Omega \to \Omega^{\otimes l}$ respectively. $\mathrm{End}_{\mathbb{K}}(\Omega^{\otimes *})$ is identified with $\mathrm{End}_{\mathbb{K}}(Q^{*\otimes *}) \otimes \mathrm{End}_{\mathbb{K}}(S^{\otimes *})$ as usual. The following Lemma explicitly computes $X_k \cup Y_l$.

Lemma 7.

$$[(\gamma_k \otimes \mathrm{id}) \circ i_k] \cup [(\delta_l \otimes \mathrm{id}) \circ i_l] = [((\gamma_k \otimes \delta_l) \otimes \mathrm{id}) \circ i_{k+l}].$$

The element $(\gamma_k \otimes \delta_l) \in \mathbb{K}(S_k \times S_l) \subset \mathbb{K}(S_{k+l})$ where $S_k \times S_l$ is embedded in S_{k+l} in the natural way.

Before proving this Lemma, we note that *part 3 of Lemma 3 follows immediately from the above lemma.*

Proof. Let W be any $\mathbb{K}$-vector space $\mathrm{Sym}^*\Omega$ acts trivially. In other words, $y.w = 0$ for any $w \in W$ and any y in the ideal of $\mathrm{Sym}^*(\Omega)$ generated by Ω. Let $\phi \in \mathrm{End}(W)$. Let $\bar{\phi} : W \otimes \mathrm{Sym}^*(\Omega) \to W$ denote the map $\phi \otimes \eta$ where $\eta : \mathrm{Sym}^*\Omega \to \mathbb{K}$ canonical map from $\mathrm{Sym}^*\Omega$ to its quotient by the ideal generated by Ω.

Let $\alpha_j : \Omega^{\otimes j} \otimes \mathrm{Sym}^*(\Omega) \to \Omega^{\otimes j} \otimes \mathrm{Sym}^*(\Omega)$ denote the map

$$\omega_1 \otimes \cdots \otimes \omega_j \bigotimes Y \mapsto \omega_1 \otimes \cdots \otimes \omega_{j-1} \bigotimes \omega_j Y$$

for $\omega_1, \ldots, \omega_j \in \Omega$ and $Y \in \mathrm{Sym}^*(\Omega)$.

Let $d : \wedge^j \Omega \otimes \mathrm{Sym}^*(\Omega) \to \wedge^{j-1}\Omega \otimes \mathrm{Sym}^*(\Omega)$ denote the Koszul differential.

Note that the following diagram commutes.

$$\begin{array}{ccc}
\Omega^{\otimes j} \otimes \mathrm{Sym}^*(\Omega) & \xrightarrow{\alpha_j} & \Omega^{\otimes j-1} \otimes \mathrm{Sym}^*(\Omega) \\
{\scriptstyle i_j \otimes \mathrm{id}_{\mathrm{Sym}^*(\Omega)}} \downarrow & & \downarrow {\scriptstyle i_{j-1} \otimes \mathrm{id}_{\mathrm{Sym}^*(\Omega)}} \\
\wedge^j \Omega \otimes \mathrm{Sym}^*(\Omega) & \xrightarrow{d} & \wedge^{j-1}\Omega \otimes \mathrm{Sym}^*(\Omega)
\end{array}$$

We have the following commutative diagrams.

$$\begin{array}{ccccccc}
0 & \longrightarrow & \Omega^{\otimes k} & \longrightarrow & Z_1 & \longrightarrow & \ldots \\
 & & {\scriptstyle \bar{\gamma}_k} \uparrow & & {\scriptstyle \theta_1} \uparrow & & \uparrow \\
\ldots & \longrightarrow & \Omega^{\otimes k} \otimes \mathrm{Sym}^*\Omega & \xrightarrow{\alpha_k} & \Omega^{\otimes k-1} \otimes \mathrm{Sym}^*\Omega & \longrightarrow & \ldots
\end{array}$$

$$\begin{array}{ccccccc}
\ldots & \longrightarrow & Z_k & \longrightarrow & \mathbb{K} & \longrightarrow & 0 \\
\uparrow & & \uparrow & & \uparrow {\scriptstyle \mathrm{id}} & & \\
\ldots & \longrightarrow & \mathrm{Sym}^*\Omega & \longrightarrow & \mathbb{K} & \longrightarrow & 0
\end{array}$$

$$\begin{array}{ccccccc}
0 & \longrightarrow & \Omega^{\otimes l} & \longrightarrow & W_l & \longrightarrow & \ldots \\
 & & {\scriptstyle \bar{\delta}_l} \uparrow & & {\scriptstyle \theta_2} \uparrow & & \uparrow \\
\ldots & \longrightarrow & \Omega^{\otimes l} \otimes \mathrm{Sym}^*\Omega & \xrightarrow{\alpha_l} & \Omega^{\otimes l-1} \otimes \mathrm{Sym}^*\Omega & \longrightarrow & \ldots
\end{array}$$

$$\begin{array}{ccccccc}
\ldots & \longrightarrow & W_l & \longrightarrow & \mathbb{K} & \longrightarrow & 0 \\
\uparrow & & \uparrow & & \uparrow {\scriptstyle \mathrm{id}} & & \\
\ldots & \longrightarrow & \mathrm{Sym}^*\Omega & \longrightarrow & \mathbb{K} & \longrightarrow & 0
\end{array}$$

The top rows of the two commutative diagrams are exact sequences representing X_k and Y_l respectively. To compute the cup product $X_k \cup Y_l$ we only need to find vertical arrows making all squares in the following diagram commute.

$$\begin{tikzcd}
0 \arrow[r] & \Omega^{\otimes k+l} \arrow[r] & Z_1 \otimes \Omega^{\otimes l} \arrow[r] & \dots \\
\dots \arrow[r] & \Omega^{\otimes k+l} \otimes \operatorname{Sym}^* \Omega \arrow[r, "\alpha_{k+l}"] \arrow[u] & \Omega^{\otimes k+l-1} \otimes \operatorname{Sym}^* \Omega \arrow[r] \arrow[u] & \dots \arrow[u]
\end{tikzcd}$$

$$\begin{tikzcd}
\dots \arrow[r] & Z_k \otimes \Omega^{\otimes l} \arrow[r] & W_1 \arrow[r] & \dots \\
\dots \arrow[r] \arrow[u] & \Omega^{\otimes l} \otimes \operatorname{Sym}^* \Omega \arrow[r, "\alpha_l"'] \arrow[u] & \Omega^{\otimes l-1} \otimes \operatorname{Sym}^* \Omega \arrow[r] \arrow[u, "\theta_2"'] & \dots \arrow[u]
\end{tikzcd}$$

$$\begin{tikzcd}
\dots \arrow[r] & W_l \arrow[r] & \mathbb{K} \arrow[r] & 0 \\
\dots \arrow[r] \arrow[u] & \operatorname{Sym}^* \Omega \arrow[r] \arrow[u] & \mathbb{K} \arrow[r] \arrow[u, "\mathrm{id}"'] & 0
\end{tikzcd}$$

Note that the diagrams below commute.

$$\begin{tikzcd}
0 \arrow[r] & \Omega^{\otimes k+l} \arrow[r] & Z_1 \otimes \Omega^{\otimes l} \arrow[r] & \dots \\
\dots \arrow[r] & \Omega^{\otimes k+l} \otimes \operatorname{Sym}^* \Omega \arrow[r, "\alpha_{k+l}"] \arrow[u, "\gamma_k \bar{\otimes} \delta_l"] & \Omega^{\otimes k+l-1} \otimes \operatorname{Sym}^* \Omega \arrow[r] \arrow[u, "\theta_1 \otimes \delta_l"] & \dots \arrow[u]
\end{tikzcd}$$

$$\begin{tikzcd}
\dots \arrow[r] & Z_k \otimes \Omega^{\otimes l} \\
\dots \arrow[r] \arrow[u] & \Omega^{\otimes l} \otimes \operatorname{Sym}^* \Omega \arrow[u]
\end{tikzcd}$$

$$\begin{tikzcd}
Z_k \otimes \Omega^{\otimes l} \arrow[r] & \Omega^{\otimes l} \\
\Omega^{\otimes l} \otimes \operatorname{Sym}^* \Omega \arrow[r] \arrow[u, "- \otimes \delta_l"] & \Omega^{\otimes l} \arrow[u, "\delta_l"']
\end{tikzcd}$$

These diagrams prove the desired lemma.

□

3.2 An example.

Lemma 3 tells us that if $X = G(\infty, \infty) = \varinjlim G(r, \infty)$ then $\mathrm{R}(X) = \oplus_k \mathbb{K} S_k$ with $\sigma_* \alpha = \operatorname{sgn}(\sigma) \sigma^{-1} \alpha \sigma$ for all $\sigma \in S_k, \alpha \in \mathbb{K} S_k$. Thus, by Lemma 3 and

Proposition 1, if $\alpha \in S_k$, and $\beta \in S_l$, then $\alpha \odot \beta = \sum_{\sigma \in \mathrm{Sh}_{k,l}} \sigma(\alpha \times \beta)\sigma^{-1}$. In other words, $\mathrm{R}(X)$ is the commutative algebra generated by symbols x_γ for all $\gamma \in S_k$, for all k modulo the relations $x_\alpha x_\beta = \sum_{\sigma \in \mathrm{Sh}_{k,l}} x_{\sigma(\alpha\times\beta)\sigma^{-1}}$. This can be seen to be larger than the usual cohomology ring of this space.

4 The big Chern Classes t_k and a ring homomorphism from $K(X) \otimes \mathbb{Q}$ to $\mathrm{R}(X)$

Let V be a locally free coherent sheaf on a scheme X/S with X smooth over S. An algebraic connection on V is defined as an $\mathcal{O}_S$ linear sheaf homomorphism $D : V \to \Omega_{X/S} \bigotimes_{\mathcal{O}_X} V$ satisfying the Leibniz rule, i.e,

$$D(fv) = df \otimes v + fDv \; \forall \; f \in \Gamma(U, \mathcal{O}_X) \;,\; v \in \Gamma(U, V),$$

for every U open in X. Note that a connection on V by itself is not $\mathcal{O}_X$ linear. However, if D_1 and D_2 are two connections on $V|_U$ with $U \subseteq X$ open, then $D_1 - D_2 \in \Gamma(U, \mathrm{End}(V) \otimes \Omega_{X/S})$.

For each open $U \subseteq X$, let $C_V(U)$ denote the set of connections on $V|_U$. This gives us a sheaf of sets on X on which $\mathrm{End}(V) \otimes_{\mathcal{O}_X} \Omega_{X/S}$ acts simply transitively. Consider a covering of X by open affines U_i such that V is trivial on U_i, and pick an element $D_i \in C_V(U_i) \; \forall \; i$ (D_i exists as $d^n : \mathcal{O}_X^n \to \Omega_X^n$ is a connection and thus gives a connection on $V|_{U_i} \cong \mathcal{O}_X^n$, where n is the rank of V). The D_i together give rise to a well defined element $\theta_V \in \mathrm{H}^1(X, \mathrm{End}(V) \otimes \Omega)$.

Lemma 8. *$\theta_{V\otimes W} = A_V + B_W$, where A_V and B_W are the elements in $\mathrm{H}^1(X, \mathrm{End}(V) \otimes \mathrm{End}(W) \otimes \Omega)$ induced from θ_V and θ_W respectively by the maps $\mathrm{End}(V) \to \mathrm{End}(V) \otimes \mathrm{End}(W)$ ($m \mapsto m \otimes \mathrm{id}_W$) and $\mathrm{End}(W) \to \mathrm{End}(V) \otimes \mathrm{End}(W)$, ($m' \mapsto \mathrm{id}_V \otimes m'$) respectively.*

Corollary 4. *$\theta_{V\otimes V}$ is induced from θ_V by the map $\mathrm{End}(V) \to \mathrm{End}(V) \otimes \mathrm{End}(V)$, ($m \mapsto m \otimes \mathrm{id}_V + \mathrm{id}_V \otimes m$).*

Proof. Since V and W are locally free, we can cover X by open sets U_i so that V and W are free over U_i for each i. Let $D_i \in C_V(U_i)$, and $E_i \in C_W(U_i)$ for each i. The desired result follows from the fact that $\mathrm{id}_V \otimes E_i + D_i \otimes \mathrm{id}_W \in C_{(V\otimes W)}(U_i)$. $\square$

4.1 The big Chern Classes t_k

Given any two locally free coherent sheaves $\mathcal{F}$ and $\mathcal{G}$ on X, one has a cup product $\cup : \mathrm{H}^i(X, \mathcal{F}) \otimes \mathrm{H}^j(X, \mathcal{G}) \to \mathrm{H}^{i+j}(X, \mathcal{F} \otimes \mathcal{G})$. Hence, we can consider the cup product of θ_V with itself k times -

$$\theta_V \cup \cdots \cup \theta_V =: \theta_V^k \in \mathrm{H}^k(X, \mathrm{End}(V)^{\otimes k} \bigotimes \Omega^{\otimes k}).$$

The composition map $\varphi : \mathrm{End}(V)^{\otimes k} \to \mathrm{End}(V)$ induces a map $\varphi_* : \mathrm{H}^k(X, \mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k}) \to \mathrm{H}^k(X, \mathrm{End}(V) \otimes \Omega^{\otimes k})$. Let $\widetilde{\mathrm{t}_k(V)} := \varphi_* \theta_V^k$. The trace map $tr : \mathrm{End}(V) \to \mathcal{O}_X$ is $\mathcal{O}_X$ -linear and induces $tr_* : \mathrm{H}^k(X, \mathrm{End}(V) \otimes \Omega^{\otimes k}) \to \mathrm{H}^k(X, \Omega^{\otimes k})$. By definition, $\mathrm{t}_k(V) := tr_* \widetilde{\mathrm{t}_k(V)}$. The classes t_k are referred to in Kapranov [6] as the *big Chern classes*. The projection $\Omega^{\otimes k} \to \wedge^k \Omega$ when applied to $\mathrm{t}_k(V)$ gives us $k!\,\mathrm{ch}_k(V)$ where $\mathrm{ch}_k(V)$ is the degree k part of the Chern character of V. The appropriate reference for the construction of the Atiyah class and the construction of the components of the Chern character as done here is Atiyah [12].

4.2 Basic properties of the big Chern classes

Firstly, t_k is a characteristic class. In other words,

Lemma 9. *If $0 \to V' \to V \to V'' \to 0$ is an exact sequence of locally free coherent sheaves on X, then $\mathrm{t}_k(V) = \mathrm{t}_k(V') + \mathrm{t}_k(V'')$.*

Proof. Let V,V' and V'' be as in the statement of this lemma. We first prove this lemma for the case when k = 1. Consider a cover of X by affine open sets U_i such that V and V' are trivial over the U_i. On each U_i, choose a connection D_i, so that the restriction $D_i|V'$ of D_i to V' is a connection on V'. In other words, $D_i(\Gamma(U_i, V')) \subset \Gamma(U_i, \Omega \otimes V')$. On the other hand,for each $U \subset X$ open, one can consider the $\mathbb{K}$-vector space $C_{V,V'}(U)$ of connections on $V|U$ that give rise to a connection on $V'|U$. Note that the difference between any two elements of $C_{V,V'}(U)$ is an element of $\Gamma(U, \mathcal{P} \otimes \Omega)$, which acts simply transitively on $C_{V,V'}(U)$. Here, $\mathcal{P}$ is the subsheaf of sections of $\mathrm{End}(V)$ that preserve V'.

Let $C_V(U_i)$ denote the space of connections on $V|U_i$. Thinking of the $\Pi_i D_i$ as an element of $\Pi_i C_V(U_i)$ we see that the Cech 1-cocycle $\Pi_{i<j}(D_i - D_j)$ of $\Pi_{i<j}\Gamma(U_i \cap U_j, \mathrm{End}(V) \otimes \Omega)$ yields the Atiyah class θ_V of V in $\mathrm{H}^1(X, \mathrm{End}(V) \otimes \Omega)$. On the other hand, when the D_i are thought of as elements of $C_{V,V'}(U_i)$, they similarly give rise to an element $\theta_{V,V'}$ of $\mathrm{H}^1(X, \mathcal{P} \otimes \Omega)$. If $i : P \to \mathrm{End}(V)$ is the natural inclusion, then clearly, $(i \otimes id)_* \theta_{V,V'} = \theta_V$. We shall denote $(i \otimes id)$ by i henceforth. Note that $tr \circ i = tr$. Hence, $tr_* \theta_{V,V'} = tr_* \theta_V = t_1(V)$. On the other hand, restriction to V' gives us a map $p_1 : P \to \mathrm{End}(V')$. Then $p_{1*}\theta_{V,V'}$ is the cohomology class obtained by looking at $D_i|V'$ as elements of $C_{V'}(Ui)$ which is $\theta_{V'}$. We also have a projection $p_2 : P \to \mathrm{End}(V'')$. Note that since the D_i are connections on V that restrict to connections on V', they induce connections on V'' (all restricted to U_i) which we will again denote by D_i. Note that $p_{2*}\theta_{V,V'}$ is the cohomology class obtained by thinking of D_i as elements of $C_{V''}(U_i)$, i.e, $\theta_{V''}$. Now, $tr|_{\mathcal{P}} = tr \circ p_1 + tr \circ p_2$. This proves the lemma for k = 1.

Let $\theta^k_{V,V'} := \theta_{V,V'} \cup \cdots \cup \theta_{V,V'} \in \mathrm{H}^k(X, \mathcal{P}^{\otimes k} \bigotimes \Omega^{\otimes k})$. Let $\varphi : \mathcal{P}^{\otimes k} \to \mathcal{P}$ denote the composition map. Let $\mathrm{t}_k\widetilde{(V,V')} := \varphi_* \theta^k(V,V') \in \mathrm{H}^k(X, \mathcal{P} \otimes \Omega^{\otimes k})$. The following observations prove the lemma in general.

1. $i_* \mathrm{t}_k\widetilde{(V,V')} = \widetilde{\mathrm{t}_k(V)}$. This follows from the commutativity of the following diagram.

$$\begin{array}{ccc}
\mathcal{P}^{\otimes k} & \xrightarrow{i^{\otimes k}} & \mathrm{End}(V)^{\otimes k} \\
\Big\downarrow{\scriptstyle\varphi} & & {\scriptstyle\varphi}\Big\downarrow \\
\mathcal{P} & \xrightarrow{i} & \mathrm{End}(V)
\end{array}$$

2. $p_{1*} \mathrm{t}_k\widetilde{(V,V')} = \widetilde{\mathrm{t}_k(V')}$ and $p_{2*} \mathrm{t}_k\widetilde{(V,V')} = \widetilde{\mathrm{t}_k(V'')}$. This is because the two diagrams below commute.

$$\begin{array}{ccc}
\mathcal{P}^{\otimes k} & \xrightarrow{{p_1}^{\otimes k}} & \mathrm{End}(V')^{\otimes k} \\
\Big\downarrow{\scriptstyle\varphi} & & {\scriptstyle\varphi}\Big\downarrow \\
\mathcal{P} & \xrightarrow{p_1} & \mathrm{End}(V')
\end{array}$$

$$\begin{array}{ccc}
\mathcal{P}^{\otimes k} & \xrightarrow{{p_2}^{\otimes k}} & \mathrm{End}(V'')^{\otimes k} \\
\Big\downarrow{\scriptstyle\varphi} & & {\scriptstyle\varphi}\Big\downarrow \\
\mathcal{P} & \xrightarrow{p_2} & \mathrm{End}(V'')
\end{array}$$

From this and the additivity of trace, we see that $\mathrm{t}_k(V) = \mathrm{t}_k(V') + \mathrm{t}_k(V'')$.

$\square$

Lemma 10. *If $f : Y \to X$ is a morphism of varieties and V is a vector bundle on X, then $\mathrm{t}_k(f^*V) = f^*\mathrm{t}_k(V)$.*

Lemma 11. *If $V = V' \oplus V''$ as $\mathcal{O}_X$-modules and p_1 and p_2 are the natural projections $\mathrm{End}(V) \to \mathrm{End}(V')$ and $\mathrm{End}(V) \to \mathrm{End}(V'')$ respectively, then $p_{1*}\widetilde{\mathrm{t}_k(V)} = \widetilde{\mathrm{t}_k(V')}$ and $p_{2*}\widetilde{\mathrm{t}_k(V)} = \widetilde{\mathrm{t}_k(V'')}$.*

Lemmas 10 and 11 are fairly straightforward to verify and we shall skip their verification. Another important property that we prove here is that $\oplus\, \mathrm{t}_k : K(X) \otimes \mathbb{Q} \to \mathrm{R}(X)$ is a ring homomorphism.

Lemma 12. *If V and W are two locally free coherent sheaves on X, then,*

$$\mathrm{t}_k(V \otimes W) = \sum_{l+m=k} \mathrm{t}_l(V) \odot \mathrm{t}_m(W)$$

where $\odot$ *is the product* $\mathrm{H}^l(X, \Omega^{\otimes l}) \otimes \mathrm{H}^m(X, \Omega^{\otimes m}) \to \mathrm{H}^k(X, \Omega^{\otimes k})$ *appearing in Proposition 1. In other words,* $\oplus\, \mathrm{t}_k : K(X) \otimes \mathbb{Q} \to \mathrm{R}(X)$ *is a ring homomorphism.*

Proof. We know that $\theta_{V\otimes W} = \theta_V \otimes \mathrm{id}_W + \mathrm{id}_V \otimes \theta_W$. Therefore,

$$\theta_{V\otimes W}{}^k = (A_V + B_W) \cup \cdots \cup (A_V + B_W)$$

where $A_V = \theta_V \otimes \mathrm{id}_W$ and $B_W = \mathrm{id}_V \otimes \theta_W$. Thus,

$$\theta_{V\otimes W}{}^k = (A_V + B_W)^k = \sum_{l+m=k} \sum_{\sigma \in \mathrm{Sh}_{l,m}} \mathrm{sgn}(\sigma) \sigma^{-1}{}_*(A_V{}^l \cup B_W{}^m).$$

Here, a given permutation $\mu \in S_k$ acts on $\mathrm{End}(V \otimes W)^{\otimes k} \bigotimes \Omega^{\otimes k}$ by

$$v_1 \otimes \cdots \otimes v_k \bigotimes w_1 \otimes \cdots \otimes w_k \mapsto v_{\mu(1)} \otimes \cdots \otimes v_{\mu(k)} \bigotimes w_{\mu(1)} \otimes \cdots \otimes w_{\mu(k)}$$

and therefore induces a map from $\mathrm{H}^k(X, \mathrm{End}(V \otimes W)^{\otimes k} \otimes \Omega^{\otimes k})$ to itself.

To verify that

$$(A_V + B_W)^k = \sum_{l+m=k} \sum_{\sigma \in \mathrm{Sh}_{l,m}} \mathrm{sgn}(\sigma) \sigma^{-1}{}_*(A_V{}^l \cup B_W{}^m),$$

note that in $(A_V + B_W)^k$, terms having l A_V's cupped with m B_W's are in one-one correspondence with sequences $b_1 < \cdots < b_m$, $b_i \in \{1, 2, 3, \ldots, l+m\}$ $\forall i$ (the b_i's being the positions of the B_W's). Such sequences are in $1-1$ correspondence with (l, m) shuffles. The sequence $\mathrm{B} := b_1, .., b_m$ corresponds to the (l, m)-shuffle σ_{B} such that $\sigma_{\mathrm{B}}(l+i) = b_i, 1 \leq i \leq m$. Note that $\mathrm{sgn}(\sigma_{\mathrm{B}}) \sigma_{\mathrm{B}}{}_*^{-1} A_V{}^l \cup B_W{}^m$ is exactly the term in $(A_V + B_W)^k$ where the B_W's are in positions $b_1, \ldots, b_m$. The lemma is now proven by recognizing that $tr_* \circ \varphi_* \sigma_*^{-1}(A_V{}^l \cup B_W{}^m) = \sigma_*^{-1}\, \mathrm{t}_l(V) \cup \mathrm{t}_m(W)$ if σ is any (l, m)-shuffle. This is because the inverse of an (l, m)-shuffle does not change the order of composition among the $\mathrm{End}(V)$-terms and among the $\mathrm{End}(W)$ terms respectively. □

Not only that, the ring homomorphism $\oplus\, \mathrm{t}_k$ is also a homomorphism of special λ-rings. In other words, the big Chern classes commute with Adams operations. Indeed, the following lemma proves this fact. Note that in any special λ-ring A, the eigenspace corresponding to the eigenvalue p^l of the Adams operation ψ^p coincides with that corresponding to the eigenvalue 2^l of the operation ψ^2 for any $p \geq 1$. Therefore, to verify that $\oplus\, \mathrm{t}_k$ commutes with the Adams operations, it suffices to verify that $\oplus\, \mathrm{t}_k$ commutes with ψ^2. This is done in the lemma below.

Lemma 13. $\mathrm{t}_k(\psi^2 V) = \psi^2\, \mathrm{t}_k(V)$.

Proof. By the corollary to Lemma 8 (Corollary 4), $\theta_{V\otimes V}$ is induced from θ_V by the map $\beta : \mathrm{End}(V) \to \mathrm{End}(V)$ given by $m \to m \otimes \mathrm{id}_V + \mathrm{id}_V \otimes m$ i.e, $\theta_{V\otimes V} = \beta_*\theta_V$. Therefore,

$$\theta_{V\otimes V}{}^k = \beta_*\theta_V \cup \cdots \cup \beta_*\theta_V = (\beta \otimes \cdots \otimes \beta)_*\theta_V{}^k.$$

By abuse of notation, we shall refer to $\beta \otimes \cdots \otimes \beta$ as β. Then, $\theta^k_{V\otimes V} = \beta_*\theta^k_V$, where $\beta : \mathrm{End}(V)^{\otimes k} \to \mathrm{End}(V)^{\otimes k}$ is given by

$$m_1 \otimes \cdots \otimes m_k \mapsto \bigotimes_{i=1}^k (m_i \otimes \mathrm{id}_V + \mathrm{id}_V \otimes m_i).$$

Further, a direct computation shows that if W is a vector space over a field F, with $charF \neq 2$, $W \otimes W = \mathrm{Sym}^2 W \oplus \wedge^2 W$. Let p_1 and p_2 denote the resulting projections from $\mathrm{End}(W) \otimes \mathrm{End}(W) = \mathrm{End}(W \otimes W)$ onto $\mathrm{End}(\mathrm{Sym}^2 W)$ and $\mathrm{End}(\wedge^2 W)$ respectively. If $M, N \in \mathrm{End}(W)$, then

$$tr(p_1(M \otimes N)) - tr(p_2(M \otimes N)) = tr(M \circ N).$$

By this fact, and Lemma 11, we see that

$$\begin{aligned}\mathrm{t}_k(\psi^2 V) = \mathrm{t}_k(\mathrm{Sym}^2 V) - \mathrm{t}_k(\wedge^2 V) &= tr_* p_{1*}\mathrm{t}_k(\widetilde{V \otimes V}) - tr_* p_{2*}\mathrm{t}_k(\widetilde{V \otimes V})\\ &= tr_*\alpha_*\mathrm{t}_k(\widetilde{V \otimes V})\end{aligned}$$

where $\alpha : \mathrm{End}(V) \otimes \mathrm{End}(V) \to \mathrm{End}(V)$ is the composition map.

Let $\varphi : \mathrm{End}(V \otimes V)^{\otimes k} \to \mathrm{End}(V \otimes V)$ be the composition map. Observe that $\alpha \circ \varphi \circ \beta : \mathrm{End}(V)^{\otimes k} \to \mathrm{End}(V)$ is the map given by

$$m_1 \otimes \cdots \otimes m_k \mapsto \sum_{p+q=k} \sum_{\sigma \in \mathrm{Sh}_{p,q}} m_{\sigma(1)} \circ \cdots \circ m_{\sigma(k)}$$

($\circ$ denoting the usual matrix multiplication on the right hand side of the last equation). Consider the map $\gamma : \mathrm{End}(V)^{\otimes k} \to \mathrm{End}(V)^{\otimes k}$ given by

$$m_1 \otimes \cdots \otimes m_k \mapsto \sum_{p+q=k} \sum_{\sigma \in \mathrm{Sh}_{p,q}} m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(k)}.$$

Then, we see that

$$tr_* \circ \varphi_* \circ \gamma_*\theta_V{}^k = tr_* \circ \alpha_*\mathrm{t}_k(\widetilde{V \otimes V}) = \mathrm{t}_k(\psi^2 V).$$

Also observe that $\psi^2\,\mathrm{t}_k(V) = tr_*\varphi_*\psi^2_*\theta_V{}^k$ since the following diagram commutes.

$$\begin{array}{ccc}\mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k} & \xrightarrow{\mathrm{id} \otimes \psi^2} & \mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k}\\ {\scriptstyle tr\circ(\varphi\otimes\mathrm{id})}\downarrow & & \downarrow{\scriptstyle tr\circ(\varphi\otimes\mathrm{id})}\\ \Omega^{\otimes k} & \xrightarrow{\psi^2} & \Omega^{\otimes k}\end{array}$$

Here, ψ^2_* on $\mathrm{H}^k(X, \mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k})$ is by definition induced on co-homology by the endomorphism $\mathrm{id} \otimes \psi^2$ of $\mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k}$. Thus, the following lemma remains to be proven.

□

Lemma 14. $\gamma_* {\theta_V}^k = \psi^2_* {\theta_V}^k$

Proof. Note that the cup-product is anti-commutative. Therefore, if $\sigma \in S_k$, then the map given by

$$\sigma : m_1 \otimes \cdots \otimes m_k \bigotimes v_1 \otimes \cdots \otimes v_k \mapsto \mathrm{sgn}(\sigma) m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(k)} \bigotimes v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)}$$

preserves ${\theta_V}^k$.
If $\sigma \in S_k$ let $\sigma \otimes \mathrm{id}$ denote the endomorphism of $\mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k}$ such that

$$m_1 \otimes \cdots \otimes m_k \bigotimes v_1 \otimes \cdots \otimes v_k \mapsto m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(k)} \bigotimes v_1 \otimes \cdots \otimes v_k.$$

Similarly, let $\mathrm{id} \otimes \sigma$ denote the endomorphism of $\mathrm{End}(V)^{\otimes k} \otimes \Omega^{\otimes k}$ such that

$$m_1 \otimes \cdots \otimes m_k \bigotimes v_1 \otimes \cdots \otimes v_k \mapsto m_1 \otimes \cdots \otimes m_k \bigotimes v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)}.$$

It now suffices to note that

$$\begin{aligned}
\gamma = \sum_{p+q=k} \sum_{\sigma \in \mathrm{Sh}_{p,q}} \sigma \otimes \mathrm{id} &= \sum_{p+q=k} \sum_{\sigma \in \mathrm{Sh}_{p,q}} \mathrm{sgn}(\sigma)(\mathrm{id} \otimes \sigma^{-1}) \circ (\sigma) \\
\implies \gamma_* \theta_V^k &= \sum_{p+q=k} \sum_{\sigma \in \mathrm{Sh}_{p,q}} \mathrm{sgn}(\sigma)(\mathrm{id} \otimes \sigma^{-1})_* \circ (\sigma)_* \theta_V^k \\
&= \sum_{p+q=k} \sum_{\sigma \in \mathrm{Sh}_{p,q}} \mathrm{sgn}(\sigma)(\mathrm{id} \otimes \sigma^{-1})_* \theta_V^k \\
&= \psi^2_* \theta_V^k.
\end{aligned}$$

□

Recalling that $\alpha_l(V) = \mathrm{ch}^{-1}(\mathrm{ch}_l(V))$, where ch is the Chern character map, we now have the following corollary of Lemma 13 below.

Corollary 5. $\mathrm{t}_k(\alpha_l(V)) = {e_k}^{(l)}_* \, \mathrm{t}_k(V)$ *where* ${e_k}^{(l)}$ *is the idempotent described in Lemma 1.*

Proof. Note that $\psi^2 = \sum e^{(l)} 2^l$. The fact that the ${e_k}^{(l)}$ are mutually orthogonal idempotents adding upto id tells us that $\psi^2 \circ {e_k}^{(l)} = 2^l {e_k}^{(l)}$. Therefore, $\psi^2 \, \mathrm{t}_k(V) = \sum 2^l {e_k}^{(l)}_* \, \mathrm{t}_k(V) = \mathrm{t}_k(\psi^2 V) = \mathrm{t}_k(\sum 2^l \alpha_l(V)) = \sum 2^l \, \mathrm{t}_k(\alpha_l(V))$. Since eigenvectors corresponding to different eigenvalues of a linear operator on a finite dimensional vector space over a field of characteristic 0 are linearly independent, the desired result follows. □

Remark. More conceptually, if TV is the graded tensor algebra over a vector space V, (with usual tensor product giving the multiplication, and coproduct dictated by the fact that $V \subset TV$ are primitive elements), then T^*V is the graded Hopf algebra dual to TV. The map $\psi^2 = \mu \circ \Delta : T^*V \to T^*V$ has as its dual the map $\mu \circ \Delta : TV \to TV$. The 2^l-eigenspace of this map is seen to be $''\operatorname{Sym}^l(L(V))''$. Thus, the 2^l-eigenspace of $\psi^2 : T^*V \to T^*V$ is dual to the space $''\operatorname{Sym}^l(L(V))''$. Thus, $\mathrm{t}_k(\alpha_l(V))$ lands in k-cohomology with coefficients in a space dual to $''\operatorname{Sym}^l(L(\Omega))''$. Moreover, the last corollary explicitly describes the projector that gives $\mathrm{t}_k(\alpha_l(V))$ from $\mathrm{t}_k(V)$ as the action on $\mathrm{t}_k(V)$ of a certain idempotent in $\mathbb{K}(S_k)$. Thus, one can recover $\mathrm{t}_k(\alpha_l(V))$ from $\mathrm{t}_k(V)$ combinatorially.

5 Calculating $\mathrm{t}_k(Q)$, Q the universal quotient bundle of a Grassmannian $G(r,n)$

We remark that $Q_{G(r,n)}$ is often denoted by just Q in this and subsequent sections. The Grassmannian whose universal quotient bundle we are referring to is usually clear by the context.

5.1 Alternative construction for $\widetilde{\mathrm{t}_k(V)}$ and $\mathrm{t}_k(V)$

Let V be a locally free coherent sheaf on a (separated) scheme X/S. It is a fact that θ_V is the element in $\operatorname{Ext}^1(V, V \otimes \Omega) \cong \mathrm{H}^1(X, \operatorname{End}(V) \otimes \Omega)$ corresponding to the exact sequence $0 \to V \otimes \Omega \to J_1(V) \to V \to 0$ where $J_1(V)$ is the first jet bundle of V. Suppose that $\alpha \in \mathrm{H}^i(X, \mathcal{F}) = \operatorname{Ext}^i(\mathcal{O}_X, \mathcal{F})$ is given by an exact sequence

$$0 \to \mathcal{F} \to Y_1 \to \ldots \to Y_i \to \mathcal{O}_X \to 0$$

and that $\beta \in \mathrm{H}^j(X, \mathcal{G}) = \operatorname{Ext}^j(\mathcal{O}_X, \mathcal{G})$ is given by an exact sequence

$$0 \to \mathcal{G} \to Z_1 \to \ldots \to Z_j \to \mathcal{O}_X \to 0.$$

Let $\alpha * \beta$ be the element in $\mathrm{H}^{i+j}(X, \mathcal{F} \otimes \mathcal{G}) = \operatorname{Ext}^{i+j}(\mathcal{O}_X, \mathcal{F} \otimes \mathcal{G})$ defined by the exact sequence which is the tensor product of the exact sequences representing α and β respectively. We note that the product

$$* : \mathrm{H}^i(X, \mathcal{F}) \otimes \mathrm{H}^j(X, \mathcal{G}) \to \mathrm{H}^{i+j}(X, \mathcal{F} \otimes \mathcal{G})$$

$$\alpha \otimes \beta \mapsto \alpha * \beta$$

has the linearity and anticommutativity properties required of the cup product. Since all the cohomology classes we are dealing with are represented by exact sequences of $\mathcal{O}_X$-modules, we can define the cup product to be the product $*$. With this definition of the cup product, it will follow that $\widetilde{\mathrm{t}_k(V)} \in \operatorname{Ext}^k(V, V \otimes \Omega^{\otimes k})$ is given by $(\theta_V \otimes {\mathrm{id}_\Omega}^{k-1}) \circ \cdots \circ \theta_V$ where $\circ$ denotes the Yoneda product and θ_V is treated as an element in $\operatorname{Ext}^1(V, V \otimes \Omega)$.

5.2 Computation of $\widetilde{\mathrm{t}_1(Q)}$

Recall that Ω is identified with $Q^* \otimes S$. Let $\Delta : S \to Q \otimes \Omega$ be the map whose dual $\Delta^* : Q^* \otimes Q \otimes S^* \to S^*$ is $ev \otimes \mathrm{id}_{S^*}$, where $ev : Q^* \otimes Q \to \mathbb{K}$ is the evaluation map. Also, $ev \otimes \mathrm{id}_S$ is a map from $Q \otimes \Omega$ to S.

Lemma 15. *The element of* $\mathrm{End}_{\mathbb{K}}(Q \otimes \Omega)$ *representing* θ_Q *is* $\Delta \circ (ev \otimes \mathrm{id}_S)$.

Proof. We note that the following diagram commutes.

$$\begin{tikzcd}
0 \arrow[r] & S \arrow[r, "\varphi"] \arrow[d, "\Delta"'] & V \arrow[r, "\epsilon"] \arrow[d] & Q \arrow[r] \arrow[d, "\mathrm{id}"] & 0 \\
0 \arrow[r] & Q \otimes Q^* \otimes S \arrow[r] & J_1(Q) \arrow[r, "\gamma"] & Q \arrow[r] & 0
\end{tikzcd}$$

The bottom row of this diagram is the exact sequence giving θ_V. By the universal property of push-forwards, we see that the following diagram commutes (F denotes the pushforward $V \amalg_S Q^* \otimes Q \otimes S$).

$$\begin{tikzcd}
0 \arrow[r] & S \arrow[r, "\varphi"] \arrow[d, "\Delta"'] & V \arrow[r, "\epsilon"] \arrow[d] & Q \arrow[r] \arrow[d, "\mathrm{id}"] & 0 \\
0 \arrow[r] & Q^* \otimes Q \otimes S \arrow[r] \arrow[d, "\mathrm{id}"] & F \arrow[r] \arrow[d] & Q \arrow[r] \arrow[d, "\mathrm{id}"'] & 0 \\
0 \arrow[r] & Q^* \otimes Q \otimes S \arrow[r] & J_1(Q) \arrow[r] & Q \arrow[r] & 0
\end{tikzcd}$$

Therefore, θ_Q can be represented by the second row of the above diagram in $\mathrm{Ext}^1(Q, Q \otimes \Omega)$. Observe, however, that every arrow in this exact sequence is a P-module homomorphism (of course, $Q^* \otimes Q \otimes S$, V and therefore, F are all P-modules). Thus θ_Q can be represented by an exact sequence in the category of P-representations. It follows that for all $k \geq 1$, $\widetilde{\mathrm{t}_k(Q)}$ and $\mathrm{t}_k(Q)$ can be represented by exact sequences in the category of P-representations. Therefore, to find θ_Q, we need to find arrows α and β so that all squares in the following diagram commute.

$$\begin{tikzcd}
0 \arrow[r] & Q^* \otimes Q \otimes S \arrow[r] & F \arrow[r] & Q \arrow[r] & 0 \\
0 \arrow[r] & S \arrow[r, "\varphi"] \arrow[u, "\Delta"] & V \arrow[r, "\epsilon"] \arrow[u] & Q \arrow[r] \arrow[u, "\mathrm{id}"'] & 0 \\
\ldots \arrow[r] & Q \otimes \Omega \otimes \mathrm{Sym}^* \Omega \arrow[r] \arrow[u, "\alpha"] & Q \otimes \mathrm{Sym}^* \Omega \arrow[r] \arrow[u, "\beta"] & Q \arrow[r] \arrow[u, "\mathrm{id}"'] & 0
\end{tikzcd}$$

Observe that $\Omega = \mathrm{Hom}_{\mathbb{K}}(Q, S) \subseteq \mathrm{End}(V)$ (here, we have chosen a $\mathbb{K}$-vector space splitting $0 \to S \to V \leftrightarrows Q \to 0$. Choosing such a splitting describes

Ω as the subspace of elements in $\mathrm{End}(V)$ consisting of matrices whose "upper right block" is the only nonzero block. Note that the product of two such matrices is 0. Thus, any element of $\mathrm{Sym}^*\Omega$ can be thought of as an element of $\mathrm{Hom}(Q,V) \subset \mathrm{End}(V)$. In this scheme of things, we choose β to be the natural evaluation map, and α the restriction of β to $Q \otimes \Omega \otimes \mathrm{Sym}^*\Omega$. Note that β and α are $\mathrm{Sym}^*\Omega$-module homomorphisms by construction. Note that $\alpha : Q \otimes \Omega \otimes \mathrm{Sym}^*\Omega$ is the $\mathrm{Sym}^*\Omega$-module homomorphism induced by $\tilde{\alpha} := ev \in \mathrm{Hom}_{\mathbb{K}}(Q \otimes \Omega, S)$, where ev is the natural evaluation map. It follows that as an element in $\mathrm{Hom}_{\mathbb{K}}(Q \otimes \Omega, Q \otimes \Omega)$, θ_Q is given by $\Delta \circ (ev \otimes \mathrm{id}_S)$.

□

Let $\{e_i\}, 1 \le i \le r$ be a basis for Q. Let $\{f_i\}$ be the basis of Q^* dual to $\{e_i\}$. Let $\{u_i\}, 1 \le i \le n-r$ be a basis for S, and $\{v_i\}$ the basis for S^* dual to $\{u_i\}$. The following is a restatement of Lemma 15.

Lemma 16. *With the notation just fixed, as an element of* $\mathrm{End}_{\mathbb{K}}(Q \otimes \Omega) \doteq \mathrm{End}(Q) \otimes \mathrm{End}(\Omega) \doteq Q^* \otimes Q \bigotimes Q \otimes S^* \bigotimes Q^* \otimes S$,

$$\theta_Q = \sum_{l_1, m_1, r_1} f_{m_1} \otimes e_{l_1} \bigotimes e_{m_1} \otimes v_{r_1} \bigotimes f_{l_1} \otimes u_{r_1}$$

(l_1, m_1 running from 1 to r, r_1 running from 1 to $n-r$).

Proof. $ev(e_i \otimes f_j \otimes u_k) = \delta_{ij} u_k$ and $\Delta(u_k) = \sum_{l=1}^{r} e_l \otimes f_l \otimes u_k$. Therefore, $\theta_Q(e_i \otimes f_j \otimes u_k) = \delta_{ij} \sum_{l=1}^{r} e_l \otimes f_l \otimes u_k$. On the other hand,

$$f_{m_1} \otimes e_{l_1} \bigotimes e_{m_1} \otimes v_{r_1} \bigotimes f_{l_1} \otimes u_{r_1}(e_i \otimes f_j \otimes u_k) = \delta_{im_1}\delta_{jm_1}\delta_{kr_1} e_{l_1} \otimes f_{l_1} \otimes u_{r_1}.$$

This is nonzero iff $i = j = m_1$ and $k = r_1$. This proves the desired result. □

5.3 Computing $\widetilde{\mathrm{t}_k(Q)}$ for $k > 1$

This is done inductively. The method by which Yoneda products are computed is very similar to the cup product computation in the previous section. We therefore omit the details and state the key results.

If $i : \wedge^k\Omega \to \Omega^{\otimes k}$ is the natural inclusion, $\widetilde{\mathrm{t}_k(Q)}$ is given by $\gamma_k \circ i$ where $\gamma_k \in \mathrm{End}_{\mathbb{K}}(Q \otimes \Omega^{\otimes k})$ is as described in the following lemma.

Lemma 17. *Identifying* $\mathrm{End}_{\mathbb{K}}(Q \otimes \Omega^{\otimes k})$ *with* $\mathrm{End}_{\mathbb{K}}(Q) \bigotimes \Omega^{*\otimes k} \bigotimes \Omega^{\otimes k}$, *we have*

$$\gamma_k = \sum_{\substack{l_1,\dots,l_k \\ m_1,\dots,m_k \\ r_1,\dots,r_k}} \left(\begin{array}{c} (f_{m_1} \otimes e_{l_1}) \circ \cdots \circ (f_{m_k} \otimes e_{l_k}) \bigotimes (e_{m_1} \otimes v_{r_1}) \otimes \cdots \otimes (e_{m_k} \otimes v_{r_k}) \\ \bigotimes (f_{l_1} \otimes u_{r_1}) \otimes \cdots \otimes (f_{l_k} \otimes u_{r_k}) \end{array} \right).$$

Here, the $l_i, 1 \le i \le k$ and the $m_i, 1 \le i \le k$ run from 1 to r , while the $r_i, 1 \le i \le k$ run from 1 to $n-r$.

Having computed $\widetilde{\mathrm{t}_k(Q)}$ we compute $\mathrm{t}_k(Q)$. For this, we note that $\mathrm{t}_k(Q) = (tr\otimes\mathrm{id})_*\widetilde{\mathrm{t}_k(Q)}$ where $\widetilde{\mathrm{t}_k(Q)} \in \mathrm{End}(Q)\otimes\mathrm{Hom}_{\mathbb{K}}(\wedge^k\Omega, \Omega^{\otimes k})$ and $tr : \mathrm{End}(Q) \to \mathbb{K}$ is the trace map. Calculating $\mathrm{t}_k(Q)$ is then easy. In the formula in the previous lemma, we see that

$$(f_{m_1} \otimes e_{l_1}) \circ \cdots \circ (f_{m_k} \otimes e_{l_k})(e_i) = \delta_{im_k}\delta_{l_k m_{k-1}} \ldots \delta_{l_2 m_1} e_{l_1}.$$

From this, we see that $(f_{m_1} \otimes e_{l_1}) \circ \cdots \circ (f_{m_k} \otimes e_{l_k})$ has trace 1 iff $m_k = l_1, l_k = m_{k-1}, \ldots, l_2 = m_1$ and has trace 0 otherwise. From this it follows that if $i : \wedge^k\Omega \to \Omega^{\otimes k}$ is the natural inclusion, $\mathrm{t}_k(Q)$ is given by $\mu_k \circ i$ where $\mu_k \in \mathrm{Hom}_{\mathbb{K}}(\Omega^{\otimes k}, \Omega^{\otimes k})$ is as described in the following lemma.

Lemma 18. *Identifying* $\mathrm{End}_{\mathbb{K}}(\Omega^{\otimes k})$ *with* $\Omega^{*\otimes k} \bigotimes \Omega^{\otimes k}$ *we have*

$$\mu_k = \sum_{\substack{l_1,\ldots,l_k \\ r_1,\ldots,r_k}} (e_{l_2} \otimes v_{r_1}) \otimes \cdots \otimes (e_{l_k} \otimes v_{r_{k-1}}) \otimes (e_{l_1} \otimes v_{r_k}) \bigotimes (f_{l_1} \otimes u_{r_1}) \otimes \cdots \otimes (f_{l_k} \otimes u_{r_k})$$

$$= \sum_{\substack{m_1,\ldots,m_k \\ r_1,\ldots,r_k}} (e_{m_1} \otimes v_{r_1}) \otimes \cdots \otimes (e_{m_k} \otimes v_{r_k}) \bigotimes (f_{m_k} \otimes u_{r_1}) \otimes (f_{m_1} \otimes u_{r_2}) \otimes \cdots \otimes (f_{m_{k-1}} \otimes u_{r_k}).$$

As a consequence, the basis element $f_{i_1} \otimes \cdots \otimes f_{i_k} \bigotimes u_{j_1} \otimes \cdots \otimes u_{j_k}$ of $\Omega^{\otimes k}$ is mapped by $\mathrm{t}_k(Q)$ to $f_{i_k} \otimes f_{i_1} \otimes \cdots \otimes f_{i_{k-1}} \bigotimes u_{j_1} \otimes \cdots \otimes u_{j_k}$. Therefore, if we identify $\mathrm{End}_{\mathbb{K}}(\Omega^{\otimes k})$ with $Q^{*\otimes k} \otimes S^{\otimes k}$, $\mathrm{t}_k(Q)$ can be thought of as $(k\ k-1\ k-2\ ..\ 2\ 1) \otimes \mathrm{id}_{S^{\otimes k}}$ where $(k\ k-1\ k-2\ ..\ 2\ 1)$ is the k-cycle acting on $Q^{*\otimes k}$ by the usual action of S_k on $V^{\otimes k}$ for a vector space V. We denote this k-cycle by τ_k.

Let $P_{r,n}$ be as in Lemma 3. By Lemma 18 and the above paragraph,

Lemma 18'.
$$\mathrm{t}_k(Q) = P_{r,n}(\tau_k).$$

6 Proofs of Theorems 2 and 3

We recall that $S_{k,j}$ denotes the set $S_{k,j} = \{\sigma \in S_k | card\{i|\sigma(i) > \sigma(i+1)\} = j-1\}$, i.e, the set of permutations of $\{1, \ldots, k\}$ with $j-1$ descents. By part 2 of Lemma 3, if $\sum a_\sigma \,\mathrm{sgn}(\sigma)\sigma \in \mathbb{K}(S_k)$, we have

$$\sum a_\sigma \,\mathrm{sgn}(\sigma)\sigma_*(\mathrm{t}_k(Q)) = P_{r,n}(\sum a_\sigma \sigma^{-1}\tau_k\sigma).$$

The following lemma now follows immediately from Corollary 5.

Lemma 19.
$$\mathrm{t}_k(\alpha_l(Q)) = P_{r,n}(\sum_{j=1}^{n} \sum_{\sigma \in S_{k,j}} (a_k^{l,j} \sigma\tau_k\sigma^{-1}))$$

A remark and some notation. $\sum_{j=1}^{k}\sum_{\sigma\in S_{k,j}} \mathrm{sgn}(\sigma)a_k^{l,j}\sigma^{-1}$ is the operator $e_k^{(l)}$ for the graded commutative Hopf-algebra T^*V. In fact, $\sum_{j=1}^{k}\sum_{\sigma\in S_{k,j}} a_k^{l,j}\sigma$ is the operator $e_k^{(l)}$ for the co-commutative ordinary Hopf-algebra TV. We henceforth denote this idempotent by $\tilde{e}_k^{(l)}$. Let $*$ denote the conjugation action of $\mathbb{K}S_k$ on itself. If $a\in S_k$ and $b\in\mathbb{K}S_k$ then $a*b=aba^{-1}$ and $(\sum c_g g)*h=\sum c_g ghg^{-1}, h\in\mathbb{K}S_k$. Then, Lemma 19 can be concisely restated as

$$\mathrm{t}_k(\alpha_l(Q)) = P_{r,n}(\tilde{e}_k^{(l)} * \tau_k).$$

Note that $*$ is a left action.

6.1 Proofs of Corollary 2 and Corollary 3

Recall the definitions of the projections P_r and $P_{r,n}$ from Section 3. Assume for now that n is large enough so that $P_r = P_{r,n}$ for all values of k that we shall use. Let $I(k,r,l)$ denote the annihilator in $\mathbb{K}S_k$ of $\mathrm{t}_k(\alpha_l(Q))$. By Lemma 3 and Lemma 19 this is precisely the subspace

$$I(k,r,l) = \{\sum_g c_g\,\mathrm{sgn}(g) g | P_r((\sum_g c_g g^{-1}) * \tilde{e}_k^{(l)} * \tau_k) = 0\}.$$

If $\langle\alpha\rangle$ denotes the subspace of $\mathbb{K}S_k$ spanned by conjugates of α by elements of S_k where $\alpha\in\mathbb{K}S_k$, then

$$\dim(I(k,r,l)) = \dim(\langle\tilde{e}_k^{(l)} * \tau_k\rangle) - \dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k)\rangle).$$

Note that since P_{r-1} factors through P_r, $I(k,r,l)\subseteq I(k,r-1,l)$. It follows that this inclusion is strict if

$$\dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k)\rangle) > \dim(\langle P_{r-1}(\tilde{e}_k^{(l)} * \tau_k)\rangle).$$

We will prove the following lemma.

Lemma 20. *For a fixed l, there exists a constant C and infinitely many r such that there exists a $k < Cr^2$ so that*

$$\dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k)\rangle) > \dim(\langle P_{r-1}(\tilde{e}_k^{(l)} * \tau_k)\rangle).$$

Note that in such a situation, if $n > Cr^2 + r$ then $P_r = P_{r,n}$ as projection operators on $\mathbb{K}S_k$. We can then pick an element β in $\mathbb{K}S_k$ such that $\beta_*\,\mathrm{t}_k(\alpha_l(Q_{G(r-1,\infty)})) = 0$ and $\beta_*\,\mathrm{t}_k(\alpha_l(Q_{G(r,n)})) \neq 0$. If Corollary 2 were false there would be a morphism $f: G(r,n)\to G(r-1,\infty)$ so that $f^*\left(\beta_*\,\mathrm{t}_k(\alpha_l(Q_{G(r-1,\infty)}))\right) = \beta_*\,\mathrm{t}_k(\alpha_l(Q_{G(r,n)}))$. This gives us a contradiction. Therefore, Corollary 2 follows immediately from Lemma 20.

We will prove Lemma 20 by a simple counting argument. We however, need the following lemma.

Lemma 21.

$$\tilde{e}_k^{(l)} * \tau_k = \tilde{e}_{k-1}^{(l-1)} * \tau_k$$

where $S_{k-1} \subset S_k$ is embedded as the subgroup fixing k.

Proof. Let α be a permutation of $\{1, 2, 3, \ldots, k-1\}$ with $j-1$ descents. Then, among the permutations $\alpha, \alpha\tau_k, \ldots, \alpha{\tau_k}^{k-1}$, we see that j of the permutations have $j-1$ descents, while the remaining $k-j$ have j descents. For, $\alpha{\tau_k}^i$ has j descents or $j-1$ descents depending on whether $\alpha(k-i) < \alpha(k-i+1)$ or not, for $2 \le i \le k-1$. For $j-1$ such i, $\alpha(k-i) > \alpha(k-i+1)$ (corresponding to the descents of α). These $j-1$ elements together with α have $j-1$ descents. The remaining $k-j$ permutations have j descents. As ${\tau_k}^i \tau_k {\tau_k}^{-i} = \tau_k$, the coefficient of $\alpha\tau_k\alpha^{-1}$ in $\tilde{e}_k^{(l)} * \tau_k$ is given by $ja_k^{l,j} + (k-j)a_k^{l,j+1}$, since among the elements $\alpha, \alpha\tau_k, \ldots, \alpha{\tau_k}^{k-1}$, those with $j-1$ descents contribute $a_k^{l,j}$ and those with j descents contribute $a_k^{l,j+1}$ to the coefficient of $\alpha\tau_k\alpha^{-1}$ in $\tilde{e}_k^{(l)} * \tau_k$. The desired lemma follows from observing that $ja_k^{l,j} + (k-j)a_k^{l,j+1} = ja_{k-1}^{l-1,j}$, since $j\binom{X-j+k}{k} + (k-j)\binom{X-j-1-k}{k} = X\binom{X-j-1-k}{k-1}$.

□

Proof. (Proof of Lemma 20). Suppose we have shown that there exists a constant C such that for a fixed l and r,

$$\dim(\langle \tilde{e}_k^{(l)} * \tau_k \rangle) > \dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k) \rangle)$$

if $k \ge Cr^2$. Then,

$$\text{there exists } s \ge r \text{ so that } \dim(\langle P_s(\tilde{e}_k^{(l)} * \tau_k) \rangle) < \dim(\langle P_{s+1}(\tilde{e}_k^{(l)} * \tau_k) \rangle).$$

Therefore, for any l and r, there exists $s \ge r$ so that

$$\dim(\langle P_s(\tilde{e}_k^{(l)} * \tau_k) \rangle) < \dim(\langle P_{s+1}(\tilde{e}_k^{(l)} * \tau_k) \rangle).$$

With l,r and s as above, pick $k = Cr^2$. Then $k < C(s+1)^2$ as well. This proves the lemma provided we actually show that there exists a constant C such that for a fixed l and r,

$$\dim(\langle \tilde{e}_k^{(l)} * \tau_k \rangle) > \dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k) \rangle)$$

whenever $k > Cr^2$. This is what we will do now.

1. Observe that the stabilizer of τ_k under conjugation is the cyclic subgroup generated by τ_k. Thus, S_{k-1} acts freely on the conjugates of τ_k and $\beta * \tau_k = 0$ for some $\beta \in \mathbb{K}S_{k-1}$ iff $\beta = 0$. It follows from this remark and the Lemma 21 that $\dim(\langle \tilde{e}_k^{(l)} * \tau_k \rangle)$ is the dimension of the representation $\mathbb{K}S_{k-1}.\tilde{e}_{k-1}^{(l-1)}$ of $\mathbb{K}S_{k-1}$. By exercise 4.5 in Loday[2] that this space has dimension equal to the coefficient of q^{l-1} in $q(q+1)\ldots(q+k-2)$.

2. On the other hand, look at $\dim(\oplus_{|\lambda|\leq r}\operatorname{End}(V_\lambda)$ for a fixed r. Note that if $\lambda : k = \lambda_1 + \cdots + \lambda_{r'}$ is a partition of k, and if Π denotes the product of the hook lengths of the Young diagram corresponding to λ, then $\dim(V_\lambda) = \frac{k!}{\Pi} \leq \frac{k!}{\lambda_1!\lambda_2!\ldots\lambda_{r'}!}$. Thus, $\dim(\operatorname{End}(V_\lambda)) \leq (\frac{k!}{\lambda_1!\lambda_2!\ldots\lambda_{r'}!})^2$. Hence,

$$\dim(\oplus_{|\lambda|\leq r}\operatorname{End}(V_\lambda)) \leq \sum_{\substack{\lambda_1+\cdots+\lambda_r=k\\ \lambda_i\geq 0}} (\frac{k!}{\lambda_1!\lambda_2!\ldots\lambda_r!})^2$$

$$\leq (\sum_{\substack{\lambda_1+\cdots+\lambda_r=k\\ \lambda_i\geq 0}} \frac{k!}{\lambda_1!\lambda_2!\ldots\lambda_r!})^2 = r^{2k}.$$

Therefore, for a fixed r,

$$\dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k)\rangle) \leq \dim(\oplus_{|\lambda|\leq r}\operatorname{End}(V_\lambda)) \leq r^{2k}.$$

On the other hand,

$$\dim(\langle \tilde{e}_k^{(l)} * \tau_k\rangle) = \text{ coefficient of } q^{l-1} \text{ in } q(q+1)\ldots(q+k-2) \geq \frac{(k-2)!}{(l-2)!}.$$

We need to find k large enough so that $\frac{(k-2)!}{(l-2)!} > r^{2k}$. To see this we need to find k large enough so that

$$\ln((k-2)!) - \ln((l-2)!) > 2k\ln r.$$

Note that

$$\ln((k-2)!) > (k-2)\ln(k-2) - (k-3).$$

We therefore, only need to find k large enough so that

$$(k-2)\ln(k-2) > k-3+\ln((l-2)!) + (k-2)\ln(r^2) + 2\ln(r^2).$$

Put $D = \ln(r^4(l-2)!)$. We then need k so that

$$(k-2)\ln(k-2) > k-3+D+(k-2)\ln(r^2).$$

Certainly, there exists $N \in \mathbb{N}$ so that $N(k-2) > (k-3)+D$ To see this, note that we can pick $N > D+1$ if $k > 3$ for instance. In fact, picking $N > 5+\ln((l-2)!)$ works as well. The latter choice of N is independent of r. If $k-2 > e^N r^2$, then we see that

$$(k-2)\ln(k-2) > k-3+D+(k-2)\ln(r^2).$$

Certainly, $k > e^{N+1}r^2$ would do for our purposes.

Thus, if l and r are fixed, we have shown that there is a constant C so that when $k > Cr^2$, then

$$\dim(\langle \tilde{e}_k^{(l)} * \tau_k \rangle) > \dim(\langle P_r(\tilde{e}_k^{(l)} * \tau_k) \rangle).$$

If $l = 2$, in particular, we need

$$(k-2)\ln(k-2) > k - 3 + (k-2)\ln(r^2) + 2\ln(r^2)$$

We see that this happens if $k - 2 > 7r^2$.

$\square$

This completes the proof of Corollary 2. In addition, we have shown in Lemma 20 and hence in Corollary 2 that if $l = 2$, $C = 7$ works.

To complete the proof of Corollary 3, we make some observations.

Observation 1. *By Lemma 21,*

$$\tau_k = \sum_{l \geq 2} \tilde{e}_{k-1}^{(l-1)} * \tau_k = \sum_{l \geq 2} \tilde{e}_k^{(l)} * \tau_k$$

$$\implies \mathrm{t}_k(Q) = \sum_{l \geq 2} \mathrm{t}_k(\alpha_l(Q)) \implies \mathrm{t}_k(\alpha_1(Q)) = 0 \ \forall k \geq 2.$$

Observation 2. *Since* $\oplus \mathrm{t}_k : K(X) \otimes \mathbb{Q} \to \oplus \mathrm{H}^k(X, \Omega^{\otimes k})$ *is a ring homomorphism, is follows that*

$$\mathrm{t}_k(\alpha_1(Q)^2) = 0$$

if $k \neq 2$.

If $f : G(s+1, N) \to G(s, M)$ is a morphism, then one sees that

$$f^*(\alpha_2(Q')) = A\alpha_1(Q)^2 + B\alpha_2(Q)$$

where Q and Q' are the universal quotient bundles of $G(s+1, N)$ and $G(s, M)$ respectively. By Observation 2,

$$\mathrm{t}_k\left(f^*\left(\alpha_2(Q_{G(s,M)})\right)\right) = B\,\mathrm{t}_k(\alpha_2(Q_{G(s+1,N)})).$$

If $B \neq 0$, one sees that $I(k, s, 2) \subseteq I(k, s+1, 2)$ (a contradiction). This finally proves Corollary 3.

To prove Theorem 2, we need the following lemma.

Lemma 22. *X a smooth (projective) scheme. Suppose that $[V] \in K(X)$ is given by $[V] = \sum a_i[V_i]$, where V_i's are of rank $\leq r$. Then, $I(k, r, l)$ annihilates $\mathrm{t}_k(\alpha_l([V]))$.*

Proof. There exists $N \in \mathbb{N}$ so that for each $m > N$ there exist surjections $\mathbb{G}_i \to V_i(m)$ where $\mathbb{G}_i$ is a free $\mathcal{O}_X$ module for each i. Let K_i denote the rank of $\mathbb{G}_i$. This is equivalent to saying that for each i there exists a morphism $f_i : X \to G(rank(V_i), K_i)$ so that $V_i(m) = {f_i}^* Q_i$, Q_i being the universal quotient bundle of $G(rank(V_i), K_i)$. Thus for each i, $I(k,r,l)$ kills $\mathrm{t}_k\left(\alpha_l(V_i \otimes \mathcal{O}(m))\right)$ for each $m > N$.

To prove this lemma, it suffices to show that $I(k,r,l)$ kills $\mathrm{t}_k(\alpha_l(V_i))$ for each i. For this, we note that $\oplus\, \mathrm{t}_k(\mathcal{O}(1)) = e^{\mathrm{t}_1(\alpha_1(\mathcal{O}_1))}$, with the understanding that $\mathrm{t}_1(\alpha_1(\mathcal{O}_1))^{D+1} = 0$ where D is the dimension of the ambient projective space. Thus, $\oplus\, \mathrm{t}_k(\mathcal{O}(m)) = e^{m\, \mathrm{t}_1(\alpha_1(\mathcal{O}_1))}$. Since the Vandermonde determinant $\Delta(N+1, .., N+D+1) \neq 0$, we can find a linear combination W of $\mathcal{O}(N+1), \dots, \mathcal{O}(N+D+1)$ so that $\mathrm{t}_k(W) = 0$ for every $k \geq 1$ and $\mathrm{t}_0(W) = 1$. Clearly, $\mathrm{t}_k(\alpha_l(V_i \otimes W)) = \mathrm{t}_k(\alpha_l(V_i))$ is killed by $I(k,r,l)$. $\square$

Proof of Theorem 2. Lemma 20 implies that given any fixed $l \geq 2$, there exists a constant C such that there exist infinitely many r such that given any $n > Cr^2 + r$,

$$I(k,r,l) \subsetneq I(k,r-1,l).$$

Lemma 22 implies that $I(k,r-1,l)$ annihilates $\mathrm{t}_k(x)$ for any element x of $F_{r-1}\mathrm{CH}^l(Q_{G(r,n)}) \otimes \mathbb{Q}$. Theorem 2 now follows immediately from the fact that $I(k,r,l)$ is the annihilator of $\mathrm{t}_k(\alpha_l(Q_{G(r,n)}))$ by definition.

6.2 Outline of proof of Theorem 3

Originally, the hope was for a stronger result saying that for fixed l and r, there exists a k satisfying $I(k,r,l) \subsetneq I(k,r-1,l)$. In fact, there was the hope of being able to show that $I(2r,r,l) \subsetneq I(2r,r-1,l)$. This would have shown that there is no morphism $f : G(r,2r) \to G(r-1,M)$ so that $f^*(\alpha_l(Q')) = \alpha_l(Q)$. We have so far been unable to do this in general. However, we have found (by means of a computer program) that $I(6,3,2) \subsetneq I(6,2,2)$ thus proving that if $f : G(3,6) \to G(2,M)$ is a morphism, then $f^*(\alpha_2(Q')) = C\alpha_1(Q)^2$. This we do by showing that $\oplus_{|\lambda|=3} \mathrm{End}(V_\lambda)$ contains an irreducible representation V_μ of S_6 not contained in $\oplus_{|\lambda| \leq 2} \mathrm{End}(V_\lambda)$, and that if π_μ denotes the projection from $\mathbb{K}S_k$ to V_μ, then $\pi_\mu * \tilde{e}_6^{(2)} * \tau_6 \neq 0$. This is achieved using a Mathematica program.

7 Proof of Theorem 1

7.1 A certain decomposition of $\mathbb{K}S_k$

Observe that $\mathbb{K}S_k = \oplus W_\lambda$ where W_λ is the $\mathbb{K}$-span of elements of S_k in the conjugacy class corresponding to the partition λ. We shall break each of the spaces W_λ further into a direct sum of $\mathbb{K}$-vector spaces in a specific manner. The significance of the new decomposition shall become clear as we proceed. First, let us decompose the conjugacy class $C_{(k)}$ which is the conjugacy class of the cycle τ_k. Note that $\tau_k = \sum_{l\geq 2} \tilde{e}_k^{(l)} * \tau_k$ and that $\tilde{e}_k^{(l)}\tilde{e}_k^{(l')} = \delta_{ll'}\tilde{e}_k^{(l)}$. Define operators Π_l on $C_{(k)}$ by $\Pi_l(\beta\tau_k\beta^{-1}) = \beta * (\tilde{e}_k^{(l)} * \tau_k)$ for $\beta \in S_k$ and extend this by linearity to $C_{(k)}$. Note that $\sum_{l\geq 2} \Pi_l(\beta * \tau_k) = \beta * \tau_k$. First, we need to check that we actually have a well defined operator here. It suffices to show that if $\beta, \gamma \in S_k$ with $\beta * \tau_k = \gamma * \tau_k$ then $\Pi_l(\beta * \tau_k) = \Pi_l(\gamma * \tau_k)$. In other words, we need to show that $\beta * (\tilde{e}_k^{(l)} * \tau_k) = \gamma * (\tilde{e}_k^{(l)} * \tau_k)$ which is equivalent to showing that $(\beta^{-1}\gamma) * (\tilde{e}_k^{(l)} * \tau_k) = \tilde{e}_k^{(l)} * \tau_k$. But $\beta * \tau_k = \gamma * \tau_k$ iff $\beta^{-1}\gamma = \tau_k^s$ for some s. Therefore, the fact that Π_l is well defined follows from the following lemma.

Lemma 23.
$$\tau_k^s * (\tilde{e}_k^{(l)} * \tau_k) = \tilde{e}_k^{(l)} * \tau_k$$
for any integer s.

Proof. This really follows from the fact that for any smooth scheme X, and for any vector bundle V on X,

$$\mathrm{sgn}(\tau_k)\tau_{k*}\,\mathrm{t}_k(V) = \mathrm{t}_k(V).$$

After all, $\mathrm{sgn}(\tau_k)\tau_{k*}{\theta_V}^k = {\theta_V}^k$ (by the properties of the cup product). Hence,

$$tr_*\varphi_*\,\mathrm{sgn}(\tau_k)\tau_{k*}{\theta_V}^k = tr_*\varphi_*{\theta_V}^k$$

where $\varphi : \mathrm{End}(V)^{\otimes k} \to \mathrm{End}(V)$ is k-fold composition. The right hand side of this equation is $\mathrm{t}_k(V)$ by definition. The left hand side is $\mathrm{sgn}(\tau_k)\tau_{k*}\,\mathrm{t}_k(V)$ since
$$tr \circ \varphi \circ \tau_k = \tau_k \circ tr \circ \varphi.$$
This tells us that $\mathrm{sgn}(\tau_k^s)\tau_{k*}^s\,\mathrm{t}_k(V) = \mathrm{t}_k(V)$. To finish the proof of the lemma, we observe that by Lemma 19,

$$\tau_k^s * (\tilde{e}_k^{(l)} * \tau_k) = \mathrm{sgn}(\tau_k^s)\tau_{k*}^s\,\mathrm{t}_k(\alpha_l(Q'))$$

and that
$$\tilde{e}_k^{(l)} * \tau_k = \mathrm{t}_k(\alpha_l(Q'))$$
where Q' is the universal quotient bundle of the Grassmannian $G(r', 2r')$ with r' chosen to be greater than k. □

The other detail to be verified is the fact that the operators Π_l are mutually orthogonal projections. For this, we see that

$$\Pi_l(\beta * \tau_k) = \beta * (\tilde{e}_k^{(l)} * \tau_k) = (\beta\tilde{e}_k^{(l)}) * \tau_k \implies \Pi_l \circ \Pi_m(\beta * \tau_k)$$
$$= (\beta\tilde{e}_k^{(m)}\tilde{e}_k^{(l)}) * \tau_k = (\beta\delta_{lm}\tilde{e}_k^{(l)}) * \tau_k.$$

We therefore, have a direct sum decomposition $W_{(k)} = \oplus_{l\geq 2}\Pi_l(W_{(k)})$.

We now proceed to breakup W_λ into a direct sum of $\mathbb{K}$-vector spaces in an analogous manner. Note that C_λ is the conjugacy class of $\tau_\lambda := \tau_{\lambda_1}\tau_{\lambda_2}\dots\tau_{\lambda_s}$ where the partition λ is given by $\lambda : k = \lambda_1 + .. + \lambda_s$, the λ_i 's arranged in decreasing order and where τ_{λ_i} is the cycle $(\lambda_1+\cdots+\lambda_i, \lambda_1+\cdots+\lambda_i-1, \dots, \lambda_1+\cdots+\lambda_{i-1})$ which is after all the cycle τ_{λ_i} embedded in S_k under the composition $S_{\lambda_i} \subset S_{\lambda_1}\times\cdots\times S_{\lambda_s} \subset S_k$. Call the map $S_{\lambda_1}\times\cdots\times S_{\lambda_s} \subset S_k$ as φ. Note that φ extends to a $\mathbb{K}$-algebra homomorphism $\varphi : \mathbb{K}(S_{\lambda_1}\times\cdots\times S_{\lambda_s}) \to \mathbb{K}(S_k)$. Identify $\mathbb{K}(S_{\lambda_1})\otimes\cdots\otimes\mathbb{K}(S_{\lambda_s})$ with $\mathbb{K}(S_{\lambda_1}\times\cdots\times S_{\lambda_s})$ and consider $(\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s})*\tau_\lambda$. By this we mean that we are looking at $\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s}$ as an element of $\mathbb{K}S_k$ through the homomorphism φ. We now make the following observations that give a step by step, explicit construction of the decomposition of $\mathbb{K}S_k$ that we are interested in.

Observation 1. *The elements* $\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s}$ *are mutually orthogonal idempotents in* $\mathbb{K}(S_k)$ *adding up to* id. This follows from the fact that the above statement is true in $\mathbb{K}(S_{\lambda_1}\times\cdots\times S_{\lambda_s})$.

Observation 2. *As* $\tau_\lambda = \tau_{\lambda_1}\otimes\cdots\otimes\tau_{\lambda_s}$,

$$(\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s})*\tau_\lambda = (\tilde{e}^{(l_1)}_{\lambda_1}*\tau_{\lambda_1})\otimes\cdots\otimes(\tilde{e}^{(l_s)}_{\lambda_s}*\tau_{\lambda_s})$$

It follows that if for some i, $\lambda_i \geq 2$ *and* $l_i = 1$, *then*

$$(\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s})*\tau_\lambda = 0.$$

Observation 3. *Let*

$$\tilde{e}^{(l)}_\lambda := \sum_{l_1+\cdots+l_s=l}\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s}.$$

Then $\tilde{e}^{(l)}_\lambda$ *is an idempotent with*

$$\tilde{e}^{(l)}_\lambda.(\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s}) = (\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s})$$

if $l_1+\cdots+l_s = l$ *and*

$$\tilde{e}^{(l)}_\lambda.(\tilde{e}^{(l_1)}_{\lambda_1}\otimes\cdots\otimes\tilde{e}^{(l_s)}_{\lambda_s}) = 0$$

otherwise.

Let Π_l be defined by $\Pi_l(\beta*\tau_\lambda) = (\beta\tilde{e}^{(l)}_\lambda)*\tau_\lambda$ for every $\beta\in C_\lambda$. We then have

Lemma 24. *The* Π_l *are well-defined mutually orthogonal projection operators on* W_λ.

Proof. Note that it suffices to show that if γ is a permutation in the stabilizer of τ_λ under conjugation, then $\gamma * (\tilde{e}^{(l)}_\lambda * \tau_\lambda) = \tilde{e}^{(l)}_\lambda * \tau_\lambda$. Note that if γ stabilizes τ_λ under conjugation, then γ is of the form $\zeta(\tau^{r_1}_{\lambda_1} \otimes \cdots \otimes \tau^{r_s}_{\lambda_s})$ where ζ permutes blocks of equal lengths among $[1,\ldots,\lambda_1], [\lambda_1+1,\ldots,\lambda_1+\lambda_2], \ldots, [\lambda_1+\cdots+\lambda_{s-1}+1,\ldots,k]$ while preserving order within such blocks. Now we need to show that $\gamma * (\tilde{e}^{(l)}_\lambda * \tau_\lambda) = \tilde{e}^{(l)}_\lambda * \tau_\lambda$. Observe that

$$(\tau^{r_1}_{\lambda_1} \otimes \cdots \otimes \tau^{r_s}_{\lambda_s}) * (\tilde{e}^{(l_1)}_{\lambda_1} \otimes \cdots \otimes \tilde{e}^{(l_s)}_{\lambda_s}) * \tau_\lambda = (\tau^{r_1}_{\lambda_1} * \tilde{e}^{(l_1)}_{\lambda_1} * \tau_{\lambda_1}) \otimes \cdots \otimes (\tau^{r_s}_{\lambda_s} * \tilde{e}^{(l_s)}_{\lambda_s} * \tau_{\lambda_s})$$
$$= (\tilde{e}^{(l_1)}_{\lambda_1} \otimes \cdots \otimes \tilde{e}^{(l_s)}_{\lambda_s}) * \tau_\lambda$$

(the last equality by Lemma 23). So, we only need to show that

$$\zeta * \tilde{e}^{(l)}_\lambda * \tau_\lambda = \tilde{e}^{(l)}_\lambda * \tau_\lambda.$$

But this is true since ζ induces a permutation ζ' of $1,2,..,s$ and we see that $\zeta.(\tilde{e}^{(l_1)}_{\lambda_1} \otimes \cdots \otimes \tilde{e}^{(l_s)}_{\lambda_s}) = (\tilde{e}^{(l_{\zeta'(1)})}_{\lambda_{\zeta'(1)}} \otimes \cdots \otimes \tilde{e}^{(l_{\zeta'(s)})}_{\lambda_{\zeta'(s)}})$. □

Observation 4. *It now follows from this and the fact that the* Π_l *are mutually orthogonal idempotents adding upto* id *that*

$$W_\lambda = \oplus \Pi_l(W_\lambda).$$

Also, Observation 2 tells us that $\Pi_1(W_\lambda) = 0$ *and that* $\Pi_2(W_\lambda) = 0$ *if* $\lambda \neq (k)$. *Therefore, this direct sum decomposition runs over* $l \geq 2$. *Combining this with the decomposition* $\mathbb{K}S_k = \oplus_\lambda W_\lambda$, *we see that*

$$\mathbb{K}S_k = \oplus_\lambda \oplus_{l\geq 2} \Pi_l(W_\lambda) = \oplus_{l\geq 2} \Pi_l(\mathbb{K}S_k).$$

7.2 Proof of Corollary 1

Definition :Define an *elementary functor of type* (k,l) to be a map v (not necessarily linear) from $K(X) \otimes \mathbb{Q}$ to $\mathrm{R}_k(X)$ such that

$$w(x) = \beta_* \,\mathrm{t}_{\lambda_1}(\alpha_{l_1}(x)) \cup \cdots \cup \mathrm{t}_{\lambda_s}(\alpha_{l_s}(x))$$

for some $\beta \in \mathbb{K}S_k$, some s-tuple $(\lambda_1,..,\lambda_s)$ of non-negative integers adding up to k and some s-tuple $(l_1,\ldots,l_s)$ of non-negative integers adding up to l.

Define a *functor of type* (k,l) to be a map from $K(X) \otimes \mathbb{Q}$ to $\mathrm{R}_k(X)$ given by a "linear combination of elementary functors of type (k,l)". In other words, a functor of type (k,l) is a map v from $K(X) \otimes \mathbb{Q}$ to $\mathrm{R}_k(X)$ such that

$$v(x) = \sum_{j=1}^{j=p} c_j w_j(x)$$

where $p \in \mathbb{N}$, and $w_1, .., w_p$ are elementary functors of type (k,l).

Define a *vector of type* (k,l) in $P_{r,n}(\mathbb{K}S_k)$ to be an element of the form $v(Q)$, where v is a functor of type (k,l) and Q is the universal quotient bundle of the Grassmannian $G(r,n)$.

Note that if v is a functor of type (k,l), then

$$v(\psi^p x) = p^l v(x)$$

for any $x \in K(X) \otimes \mathbb{Q}$. Also note that functors of type (k,l) respect pullbacks.

We now try to understand what the decomposition of $\mathbb{K}S_k$ given in the Section 7.1 means. Lemma 19 together with Lemma 3 part 3 tells us that

$$\mathrm{t}_{\lambda_1}(\alpha_{l_1}(Q_{G(r,n)})) \cup \cdots \cup \mathrm{t}_{\lambda_s}(\alpha_{l_s}(Q_{G(r,n)})) = P_{r,n}(\tilde{e}_\lambda^{(l)} * \tau_\lambda).$$

Also, by Lemma 3 part 2

$$\mathrm{sgn}(\beta)\beta_*^{-1}\,\mathrm{t}_{\lambda_1}(\alpha_{l_1}(Q_{G(r,n)})) \cup \cdots \cup \mathrm{t}_{\lambda_s}(\alpha_{l_s}(Q_{G(r,n)})) = P_{r,n}(\beta * \tilde{e}_\lambda^{(l)} * \tau_\lambda).$$

Let $l = \sum_i l_i$. Thus the space spanned by

$$\{\beta_*\,\mathrm{t}_{\lambda_1}(\alpha_{l_1}(Q_{G(r,n)})) \cup \cdots \cup \mathrm{t}_{\lambda_s}(\alpha_{l_s}(Q_{G(r,n)})) \mid \sum_i l_i = l, \sum_i \lambda_i = k\},$$

which is $P_{r,n}(\Pi_l(\mathbb{K}S_k))$, is precisely the space of vectors of type (k,l).

If both r and $n-r$ are larger than k, then $P_{r,n} = \mathrm{id}$. What we did in Section 7.1 shows that in this case, $\mathbb{K}S_k$ decomposes into the direct sum of the spaces $\Pi_l(\mathbb{K}S_k)$. The space $\Pi_l(\mathbb{K}S_k)$ is stable under conjugation and is the space of vectors of type (k,l). However, if k is not too large, something very interesting happens primarily because the projection $P_{n,r}$ "behaves badly" with the projections Π_l. Let $n \geq 2r+1$ and let $k = 2r$. Then, $P_{r,n} = P_r$. Also, $\mathrm{t}_j(Q_{G(r,n)}) = \mathrm{t}_j(Q_{G(r,M)})$ for every $M \geq n$ and every $j \leq k$. It follows that $v_l(Q_{G(r,n)}) = v_l(Q_{G(r,M)})$ for all $M \geq n$ if v_l is any functor of type $(2r,l)$. Let Q denote $Q_{G(r,n)}$. The following claim holds in this situation.

Claim: There exists a nontrivial linear dependence relation of the form

$$\sum_l v_l(Q) = 0$$

such that v_l *is a functor of type* $(2r,l)$ *for each* l.

The above claim is proven in Section 7.3. This leads to Corollary 1 as follows. Choose a shortest nontrivial linear dependence relation of the form

$$\sum_l v_l(Q) = 0$$

with v_l a functor of type $(2r, l)$. Then, suppose that there exists a map $f : G(r, n) \to G(r, M)$ with $f^*([Q_{G(r,M)}]) = \psi^p[Q_{G(r,n)}]$, we can assume without loss of generality that $M \geq n$. Thus,

$$0 = f^*(\sum_l v_l(Q_{G(r,M)})) = \sum_l v_l(f^* Q_{G(r,M)}) = \sum_l v_l(\psi^p Q) = \sum_l p^l v_l(Q).$$

Since $p \geq 2$, comparing this linear dependence relation with the previous one would enable us to extract a linear dependence relation of the same form but of shorter length than the one we began with. This yields a contradiction.

The proof of theorem 1 requires a little more work which we do in Section 7.4.

7.3 A linear dependence relation between functors of type $(2r, l)$

First, we observe that if V is a vector space with $V = V_1 \oplus V_2$ and also $V = \oplus W_i$, with p_i being the projections to V_i and π_i being the projections to W_i, then

$$\dim\ p_1(W_1) + \cdots + \dim\ p_1(W_m) \geq \dim\ V_1.$$

To see this, suppose that equality holds. Then,

$$\dim\ p_1(W_i) = \dim\ W_i - \dim\ W_i \cap V_2$$

$$\implies \dim\ W_1 \cap V_2 + \cdots + \dim\ W_m \cap V_2 = \dim\ V_2.$$

From this, we see that $\pi_i(V_2) = W_i \cap V_2$ for all $i \in \{1, 2, .., , \}$. In particular, if $\pi_i(V_2) \neq W_i \cap V_2$, then

$$\dim\ p_1(W_1) + \cdots + \dim\ p_1(W_m) > \dim\ V_1.$$

Having said this, we will prove that for $V = \mathbb{K}S_{2r}$ ($V = V_1 \oplus V_2$ where $V_1 = \oplus_{|\lambda| \leq r} \operatorname{End}(V_\lambda)$ and $V_2 = \oplus_{|\lambda| > r} \operatorname{End}(V_\lambda)$ also $V = \oplus_{l \geq 2} \Pi_l(V)$)

$$\Pi_2(V_2) \neq \Pi_2(V) \cap V_2.$$

This will prove that

$$\sum_{l \geq 2} \dim\ P_r(\Pi_l(V)) > \dim\ V_1.$$

Observation 4 of Section 7.1 tells us that $\Pi_2(V) = \Pi_2(W_{(2r)})$. Any element in this space is a linear combination of conjugates of τ_{2r}. It follows that if such a linear combination is nonzero in $\operatorname{End}(V_\lambda)$ it is also nonzero as an element of $\operatorname{End}(V_{\bar{\lambda}})$, where $\bar{\lambda}$ is the partition conjugate to λ. Thus $\Pi_{(2r)}(V) \cap V_2 = 0$. It therefore, suffices to prove that $\Pi_2(V_2) \neq 0$.

Lemma 25. *To prove that* $\Pi_2(V_2) \neq 0$, *it suffices to show that*

$$\Pi_2((1\ 2r) \sum_{g \in C_\mu} g) \neq 0$$

where $(1\ 2r)$ *is the transposition interchanging* 1 *with* $2r$ *and* μ *is some partition among* $\{(2r-1,1), \dots, (r,r)\}$.

Proof. Consider the matrix $M = (\chi_\lambda(C_\mu))$ where λ runs over all partitions of $2r$ that satisfy $\lambda \geq (r,r)$ (recall that there is a lexicographic ordering among the partitions,enabling one to compare them), and $\mu \in \{(2r-1,1), \dots, (r,r)\}$. Note that if λ is such a partition and $\lambda \neq (r,r)$ then $\lambda_1 \geq r+1$. We claim that M is of rank r. To prove this, it suffices to show that N is of rank r where $N = (\psi_\lambda(C_\mu))$, where

$$\psi_\lambda = Ind_{S_\lambda}^{S_{2r}}(triv) = \chi_\lambda + \sum_{\mu > \lambda} K_{\mu\lambda}\chi_\mu.$$

However,

$$\psi_\lambda(C_\mu) = \frac{1}{|C_\mu|}[S_{2r} : S_\lambda]|C_\mu \cap S_\lambda|.$$

Therefore, $\psi_\lambda(C_\mu) = 0$ if $\mu > \lambda$. This lexicographic order is a total order. Consider the restriction of N to the rows given by the partitions in $\{(2r-1,1), \dots, (r,r)\}$. This restriction of N is then a lower triangular matrix with nonzero diagonal entries if the rows are arranged in the correct order (since $\psi_\lambda(C_\lambda) \neq 0$). It follows that N and therefore, M are matrices of rank r.

We further claim that if we restrict M to rows corresponding to $\lambda > (r,r)$, we still get a matrix of rank r. To see this, we need to show that for some scalars a_λ,

$$\chi_{(r,r)}(C_\mu) = \sum_{\lambda > (r,r)} a_\lambda \chi_\lambda(C_\mu)$$

for all $\mu \in \{(2r-1,1), \dots, (r,r)\}$. For this, it is enough to show that

$$\psi_{(r,r)}(C_\mu) = \sum_{\lambda > (r,r)} b_\lambda \psi_\lambda(C_\mu)$$

for all $\mu \in \{(2r-1,1), \dots, (r,r)\}$, for some scalars b_λ. In fact, we claim that there are scalars b_i,$0 \leq i \leq r-1$, so that

$$\psi_{(r,r)}(C_\mu) = \sum_{0 \leq i \leq r-1} b_i \psi_{(2r-i,i)}(C_\mu).$$

Note that $|C_{(2r-s,s)} \cap S_{(2r-t,t)}| = 0$ if $s \neq t$ and both are nonzero. Also note that $\psi_{(2r)}(C_{(r,r)}) \neq 0$. Thus the vector $(\psi_{(2r)}(C_\mu)), \mu \in \{(2r-1,1), \dots, (r,r)\}$

is given by $(a_1, .., a_r)$, where $a_r \neq 0$. The vector $\psi_{(2r-s,s)}(C_\mu), \mu \in \{(2r-1,1), \ldots, (r,r)\}$ is given by $(0, .., 0, d_s, \ldots, 0)$, $d_s \neq 0$ for $1 \leq s \leq r-1$. Thus,

$$\psi_{(2r)}(C_\mu) - \sum \frac{a_s}{d_s} \psi_{(2r-s,s)}(C_\mu) = (0, .., 0, a_r)$$

which is a nonzero multiple of $\psi_{(r,r)}(C_\mu)$. This shows that the matrix $M = \chi_\lambda(C_\mu)$ where $\lambda > (r,r)$ and $\mu \in \{(2r-1,1), \ldots, (r,r)\}$ is of rank r. Since $\chi_{\bar{\lambda}} = \chi_\lambda . \mathrm{sgn}$, and $|\bar{\lambda}| \geq r+1$ iff $\lambda > (r,r)$, the matrix $M' = \chi_\lambda(C_\mu)$ where $|\bar{\lambda}| \geq r+1$ and $\mu \in \{(2r-1,1), \ldots, (r,r)\}$ is obtained from M by multiplying some columns by -1 and is therefore of rank r.

Now suppose that $\Pi_2((1\ 2r) \sum_{g \in C_{(2r-s,s)}} g) \neq 0$ for some $1 \leq s \leq r$. Since M' is of rank r, we can find a linear combination of rows of M' that gives us the vector e_s i.e, $\sum_{|\lambda| > r+1} a_\lambda \chi_\lambda(C_\mu) = 0$ if $\mu \neq (2r-s,s)$ and $\sum_{|\lambda|>r+1} a_\lambda \chi_\lambda(C_\mu) = 1$ if $\mu = (2r-s,s)$. So,

$$\Pi_2((1\ 2r)(\sum_{\substack{g \in S_{2r} \\ |\lambda| > r+1}} a_\lambda \chi_\lambda(g) g)) = \Pi_2((1\ 2r) \sum_{g \in C_{(2r-s,s)}} g) \neq 0.$$

The first equality is because only the $2r$ cycles contribute to $\Pi_2(V)$. Note that since $\sum \chi_\lambda(g) g \in \mathrm{End}(V_\lambda)$ it follows that

$$(\sum_{\substack{g \in S_{2r} \\ |\lambda| > r+1}} a_\lambda \chi_\lambda(g) g) \in V_2$$

and hence

$$(1\ 2r)(\sum_{\substack{g \in S_{2r} \\ |\lambda| > r+1}} a_\lambda \chi_\lambda(g) g) \in V_2.$$

It follows that $\Pi_2(V_2) \neq 0$. $\square$

Lemma 26. *For some s, $1 \leq s \leq r$, we have* $\Pi_2((1\ 2r) \sum_{g \in C_{(2r-s,s)}} g) \neq 0$.

Proof. Every $2r$ cycle that arises in $(1\ 2r) \sum_{g \in C_{(2r-s,s)}} g$ arises with coefficient 1. We therefore need to identify the $2r$ cycles that do arise. They are those of the form $(1\ a_2\ ..\ a_s\ 2r\ a_{s+2}\ \ldots)$ or $(1\ a_2\ \ldots\ a_{2r-s}\ 2r \ldots)$. For this proof, denote the subgroup of S_{2r} fixing the elements i and j by $S(i,j)$ for any $1 \leq i < j \leq 2r$. We note that

$$(1\ 2r) \sum_{g \in C_{(2r-s,s)}} g$$

$$= \sum_{\alpha \in S(1,2r)} \alpha * (2r\ 2r-s\ 2r-s-1 ....1\ 2r-1\ 2r-2 \ldots\ 2r-s+1)$$

$$+ \alpha * (2r\ s\ s-1 \ldots\ 1\ 2r-1 ..\ s+1)$$

$$
\begin{aligned}
&= \sum_{\alpha\in S(1,2r)} \alpha * (\tau_{2r-1}^{s-1} + \tau_{2r-1}^{2r-s-1}) * \tau_{2r}\\
&= (\tau_{2r-1}^{-1}(\sum_{\beta\in S(2r-1,2r)} \beta)\tau_{2r-1}) * (\tau_{2r-1}^{s-1} + \tau_{2r-1}^{2r-s-1}) * \tau_{2r}\\
&= (\tau_{2r-1}^{-1} \sum_{\beta\in S(2r-1,2r)} \beta) * (\tau_{2r-1}^{s} + \tau_{2r-1}^{2r-s}) * \tau_{2r}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Pi_2((1\ 2r)\sum_{g\in C_{(2r-s,s)}} g) &= (\tau_{2r-1}^{-1} \sum_{\beta\in S(2r-1,2r)} \beta) * (\tau_{2r-1}^{s} + \tau_{2r-1}^{2r-s}) * (\tilde{e}_{2r}^{(2)} * \tau_{2r})\\
&= (\tau_{2r-1}^{-1} \sum_{\beta\in S(2r-1,2r)} \beta) * (\tau_{2r-1}^{s} + \tau_{2r-1}^{2r-s}) * (\tilde{e}_{2r-1}^{(1)} * \tau_{2r}),
\end{aligned}
$$

the last equality following from Lemma 21.

For this proof, denote the subgroup of S_{2r-1} fixing the element i by $S(i)$ for any $1 \le i \le 2r-1$. It therefore, suffices to show that

$$
(\tau_{2r-1}^{-1} \sum_{\beta\in S(2r-1)} \beta)(\tau_{2r-1}^{s} + \tau_{2r-1}^{2r-s})(\tilde{e}_{2r-1}^{(1)}) \neq 0
$$

for some s, $1 \le s \le r$. It therefore, suffices to show that

$$
W_s := (\sum_{\beta\in S(2r-1)} \beta)(\tau_{2r-1}^{s} + \tau_{2r-1}^{2r-s})(\tilde{e}_{2r-1}^{(1)}) \neq 0
$$

for some s, $1 \le s \le r$. Consider a vector space V of finite dimension, and let u and v be two basis vectors of V. We will show that the right action of W_s on $u^{\otimes 2r-2} \otimes v$ is nonzero. Note that

$$
\begin{aligned}
\frac{1}{(2r-2)!}(u^{\otimes 2r-2} \otimes v)W_s &= (u^{\otimes 2r-2} \otimes v)(\tau_{2r-1}^{s} + \tau_{2r-1}^{2r-s})\tilde{e}_{2r-1}^{(1)}\\
&= (u^{\otimes s-1} \otimes v \otimes u^{\otimes 2r-1-s} + u^{\otimes 2r-1-s} \otimes v \otimes u^{\otimes s-1})\tilde{e}_{2r-1}^{(1)}.
\end{aligned}
$$

Therefore, it is enough to show that

$$
(u^{\otimes s-1} \otimes v \otimes u^{\otimes 2r-1-s} + u^{\otimes 2r-1-s} \otimes v \otimes u^{\otimes s-1})\tilde{e}_{2r-1}^{(1)} \neq 0
$$

for some s, $1 \le s \le r$. For this, we note that

$$
0 \neq ad(u)^{2r-2}(v) = (l_u - r_u)^{2r-2}(v) = \sum_i \binom{2r-2}{i} u^{\otimes i} \otimes v \otimes u^{2r-2-i}.
$$

Now, $ad(u)^{2r-2}(v)$ is an element of the free Lie algebra generated by V. The idempotent $\tilde{e}_{2r-1}^{(1)}$ therefore acts as the identity on this vector, which is a linear combination of $(u^{\otimes s-1} \otimes v \otimes u^{\otimes 2r-1-s} + u^{\otimes 2r-1-s} \otimes v \otimes u^{\otimes s-1})$ where s runs from 1 to r.

□

7.4 Final step to the proof of Theorem 1

Suppose that $[\psi^p Q] = [Y]$ for some genuine vector bundle Y. Then Y is of rank r, and for all sufficiently large m, $Y \otimes \mathcal{O}(m)$ is a quotient of $\mathcal{O}_G{}^s$ for some s. It follows that $Y \otimes \mathcal{O}(m) = f^*Q'$ for some morphism $f : G(r,n) \to G(r,n')$, where Q' is the universal quotient bundle of $G(r,n')$. Without loss of generality we may assume that $n' \geq 2r+1$. Let Q denote the universal quotient bundle of $G(r,n)$. As in Section 7.2, choose a shortest linear dependence relation of the form

$$\sum_l v_l(Q) = 0$$

where v_l is a functor of type $(2r,l)$.

Then, $\sum_l v_l(Q') = 0$. Since the v_l's respect pullbacks,

$$\sum_l v_l(Y \otimes \mathcal{O}(m)) = 0$$

for all sufficiently large m. Note that $\oplus\, \mathrm{t}_k(\mathcal{O}(m)) = exp(\mathrm{t}_1(\alpha_1(\mathcal{O}(1))))$. Therefore,

$$\mathrm{t}_{\lambda_i}(\alpha_{l_i}(Y \otimes \mathcal{O}(m))) = \mathrm{t}_{\lambda_s}(\alpha_{l_s}(Y) + m\alpha_{l_s-1}(Y)\alpha_1(\mathcal{O}(1)) + \dots).$$

Therefore,

$$v_l(Y \otimes \mathcal{O}(m)) = v_l(Y) + m.A_1(Y) + \dots + m^s A_s(Y)$$

for all l with $A_i(Y) \in \mathrm{R}(G(r,n))$. In other words, $v_l(Y \otimes \mathcal{O}(m))$ is a polynomial in m with coefficients in $\mathrm{R}(G(r,n))$ whose constant term is $v_l(Y)$. It follows that $\sum_l v_l(Y \otimes \mathcal{O}(m))$ is a polynomial in m with coefficients in $\mathrm{R}(G(r,n))$ whose constant term is $\sum_l v_l(Y)$. The fact that $\sum_l v_l(Y \otimes \mathcal{O}(m))$ vanishes for all sufficiently large m implies that $\sum_l v_l(Y) = 0$. Thus,

$$\sum_l v_l(\psi^p Q) = \sum_l p^l v_l(Q) = 0$$

as well. As in Section 7.2, since $p \geq 2$, this together with the linear dependence relation $\sum_l v_l(Q) = 0$ yields a linear dependence relation of the same form but of shorter length, thereby giving a contradiction. This finally proves Theorem 1.

Appendix

This appendix if for sketching a proof of Observation 1 of Section 3. This material is by and large reproduced from notes by Jinhyun Park [13] of a course taught by Madhav Nori at the University of Chicago in Fall 2004.

Recall that given a morphism $f : Y \to X$ of schemes, a sheaf $\mathcal{F}$ on Y is said to have descent data if it satisfies the following three properties.

$[D_1]$. Given any two morphisms $g_1, g_2 : Z \to Y$ such that $f \circ g_1 = f \circ g_2$, there is an isomorphism $c(g_1, g_2) : g_1^*\mathcal{F} \cong g_2^*\mathcal{F}$.
$[D_2]$. (Functoriality). Given any morphisms $h : W \to Z$ and $g_1, g_2 : Z \to Y$ such that $f \circ g_1 = f \circ g_2$, the following diagram commutes.

$$\begin{array}{ccc} h^* \circ g_1^*\mathcal{F} & \xrightarrow{h^*c(g_1,g_2)} & h^* \circ g_2^*\mathcal{F} \\ \simeq\Big\downarrow & & \Big\downarrow\simeq \\ (g_1 \circ h)^*\mathcal{F} & \xrightarrow{c(g_1\circ h, g_2 \circ h)} & (g_2 \circ h)^*\mathcal{F} \end{array}$$

$[D_3]$. Given any three morphisms $g_1, g_2, g_3 : Z \to Y$ such that $f \circ g_1 = f \circ g_2 = f \circ g_3$ the following diagram commutes.

$$\begin{array}{ccc} g_1^*\mathcal{F} & \xrightarrow{c(g_1,g_2)} & g_2^*\mathcal{F} \\ {\scriptstyle c(g_1,g_3)}\Big\downarrow & & \Big\downarrow{\scriptstyle c(g_2,g_3)} \\ g_3^*\mathcal{F} & \xrightarrow{\mathrm{id}} & g_3^*\mathcal{F} \end{array}$$

We now recall a theorem of Grothendieck [15].

Theorem 4. *Let $f : Y \to X$ be a flat surjective morphism of schemes. There is an equivalence of categories*

$\{$*Quasicoherent sheaves on* $X\} \longleftrightarrow$

$\{$*quasicoherent sheaves on* Y *with descent data*$\}$

$$\mathcal{G} \mapsto f^*\mathcal{G}.$$

The following construction due to Grothendieck [15] gives the inverse to the above equivalence of categories.

Construction 1. Let $\mathcal{F}$ be a quasicoherent sheaf on Y with descent data. Note that for every open $U \subset X$, $\mathcal{F}|_{f^{-1}(U)}$ is a quasicoherent sheaf with descent data for the morphism $f|_{f^{-1}(U)} : f^{-1}(U) \to U$. Let $\overline{\mathcal{F}}$ denote the sheafification of the presheaf

$$U \mapsto \{s \in \Gamma(f^{-1}(U), \mathcal{F}) \mid c(g_1, g_2)g_1^*s = g_2^*s \text{ for all } g_1, g_2 : Z \to f^{-1}(U)\}.$$

The inverse to the equivalence of categories in Theorem 4 is given by $\mathcal{F} \mapsto \overline{\mathcal{F}}$. For example, $\overline{\mathcal{O}_Y} = \mathcal{O}_X$.
Let P be an affine group scheme over $\mathbb{K}$. Let $f : Y \to X$ be a principal P-bundle on X. Then, descent data for f on a sheaf $\mathcal{F}$ is indeed equivalent to a P-action on $\mathcal{F}$. Theorem 4 therefore implies the following theorem.

Theorem 5. *Let $f : Y \to X$ be a principal P-bundle. There is an equivalence of categories*

$$\{\text{ Quasicoherent sheaves on } X\} \longleftrightarrow \{\text{Quasicoherent sheaves on } Y \text{ with } P \text{ action}\}$$

$$\mathcal{G} \mapsto f^*\mathcal{G}.$$

Corollary 6. *The functor*

$$F : \{\text{ } P\text{-representations}\} \longrightarrow \{\text{locally free Quasicoherent sheaves on } X\}$$

$$F(V) = \overline{\mathcal{O}_Y \otimes_{\mathbb{K}} V}$$

is an exact functor commuting with $\otimes$.

Proof. $\mathcal{O}_Y$ is naturally a P-sheaf on Y. A representation V of P is a H-sheaf on Spec $\mathbb{K}$. Therefore, $\mathcal{O}_Y \otimes_{\mathbb{K}} V$ is a P-sheaf on $Y = Y \times_{\text{Spec } \mathbb{K}} \text{Spec } \mathbb{K}$. By Theorem 5, $F(V)$ is a quasicoherent sheaf on X. Clearly, $F(V)$ is locally free. It can also be verified without difficulty that $V \mapsto \mathcal{O}_Y \otimes_{\mathbb{K}} V$ is an exact functor commuting with $\otimes$. Since the functor from Theorem 5 is an equivalence of categories, the desired corollary follows. □

We can now sketch the proof of the following theorem. Let G be a affine algebraic group and let P be a closed subgroup of G. Let $\mathcal{P}$ denote the category of P-representations. With these assumptions, we have the following theorem of Bott [4]. This theorem has been referred to in Section 3 as Observation 1.

Theorem 6. *Let G be reductive. If $\mathbb{K}$ is regarded as the trivial P-representation,*

$$H^i(G/P, F(V))^G \simeq Ext^i_{\mathcal{P}}(\mathbb{K}, V).$$

Proof. For any $V \in \mathcal{P}$, let $T^i(V) = \mathrm{H}^i(G/P, F(V))^G$. We shall show that in the language of Grothendieck [14], $T^0(V) = \mathrm{Hom}_{\mathcal{P}}(\mathbb{K}, V)$ and $T^i(V) = R^i T^0(V)$. This will prove the desired theorem. To do this, we need to verify the following list of properties.

(a) $T^i : \mathcal{P} \to \mathbb{K}-$ vector spaces is a functor.

(b) Given a short exact sequence $0 \to V' \to V \to V'' \to 0$ in $\mathcal{P}$, there is a long exact sequence

$$\dots \xrightarrow{\delta} T^i(V') \longrightarrow T^i(V) \longrightarrow T^i(V'') \xrightarrow{\delta} T^{i+1}V' \longrightarrow \dots$$

The given short exact sequence gives a long exact sequence $\mathrm{H}^i(G/P, -)$. Now, for any exact sequence $W' \to W \to W''$ of G-representations, the sequence ${W'}^G \to W^G \to {W''}^G$ is exact. This verifies (b).

(c) The data in (b) is functorial.

(d) $T^0(V) = V^P$.

(e) (effaceability) For all $i > 0$, for all $\alpha \in T^i(V)$, there is a monomorphism

$j : V \to W$ in $\mathcal{P}$ such that $T^i(j)(\alpha) = 0$.

We check (e), the only nontrivial assertion above. Put $W = \Gamma(P, \mathcal{O}_P)$. Then, $F(W) = f_*\mathcal{O}_G$ where $f : G \to G/P$ is the natural morphism. Note that G is affine and f ia an affine morphism. Therefore, for any quasicoherent sheaf $\mathcal{F}$ on G, $\mathrm{H}^i(G, \mathcal{F}) = 0$ for every $i > 0$ and $R^i f_*\mathcal{F} = 0$ for every $i > 0$. The Leray spectral sequence then tell us that $\mathrm{H}^i(G/P, f_*\mathcal{F}) \simeq \mathrm{H}^i(G, \mathcal{F}) = 0$ for all $i > 0$. In particular, $\mathrm{H}^i(G/P, F(W)) = 0$ for every $i > 0$. Let V be any P-representation. We have an isomorphism

$$\mathrm{Hom}_{\mathcal{P}}(V, \Gamma(P, \mathcal{O}_P)) \simeq V^* \tag{1}$$

$$L \mapsto \mathrm{ev}_{\mathrm{id}} \circ L.$$

Here, $\mathrm{ev}_{\mathrm{id}} \circ L$ is the composite

$$V \xrightarrow{L} \Gamma(P, \mathcal{O}_P) \xrightarrow{\mathrm{ev}_{\mathrm{id}}} \mathbb{K}.$$

Denote the inverse of the isomorphism (1) by S. Choose linear functionals $u_1, \dots, u_i, \dots$ on V such that $\cap \mathrm{Ker}(u_i) = 0$. Then, $S(u_i) \in \mathrm{Hom}_{\mathcal{P}}(V, \Gamma(P, \mathcal{O}_P))$. Clearly, the morphism $\oplus_i S(u_i) : V \to \oplus_i \Gamma(P, \mathcal{O}_P)$ is a monomorphism in $\mathcal{P}$. Further, $T^p(\oplus_i \Gamma(P, \mathcal{O}_P)) = 0$ whenever $p > 0$ since we just showed that $T^p(\Gamma(P, \mathcal{O}_P)) = 0$ whenever $p > 0$. This completes the verification of (e) and therefore, the proof of the desired theorem.

□

References

[1] Nori, M. V. *Personal Communication*

[2] Loday, J- L. *Cyclic Homology*, Springer ,1997.

[3] Fulton, W. and Harris, J. *Representation Theory: a first course*, Springer Verlag, GTM 129, 1991.

[4] Bott, Raoul. *Homogenous Vector Bundles*, Annals of Mathematics, 2nd Ser., Vol 66,No.2 (Sep., 1957), pp 203-248.

[5] Reutenauer, Christophe. *Free Lie Algebras*, London MathematicalSociety Monographs, new series, no. 7, Oxford University Press, 1993.

[6] Kapranov, M. Rozansky-Witten invariants via Atiyah classes. *Compositio Math.* 115(1999), 71-113.

[7] Paranjape, K. H.; Srinivas, V. Self maps of homogeneous spaces . *Invent. Math.* 98(1989), no.2,425-444

[8] Paranjape, K. H.; Srinivas, V. Continuous self-maps of quadric hypersurfaces. *Proceedings of the Indo-French conference on Geometry* (*Bombay*, 1989), 135-148, *Hindustan Book Agency, Delhi*,1993.

[9] Fulton, W. and Lang, S. *Riemann-Roch Algebra*, Grundlehren der mathematischen Wissenschaften, Springer Verlag, 1985.

[10] Weyl H. *The Classical Groups*, University Press, Princeton, 1939.

[11] Grothendieck, A. and Dieudonne, J. *Elements de Geometrie Algebrique*, Publ. Math. IHES 28 (1966), pp 143-154.

[12] Atiyah, M. F. *Complex analytic connections on fibre bundles* , Trans. Amer. Math. Soc 85 (1957), 181-207.

[13] Nori, M. V.; Notes by Jinhyun Park. K-Theory and related topics. Course taught at the University of Chicago in Fall 2004. Available online at http://math.kaist.ac.kr/~jinhyun/note/nori/ktheory_aut2004.pdf.

[14] Grothendieck, A. *Sur quelques points d'algèbre homologique*, Tôhoku Math. J. (2) 9 (1957), 119-221.

[15] Grothendieck, A. *Technique de descent et théorèmes d'existence en géometrie algébraique.I.* Séminaire Bourbaki, Vol. 5, Exp. No. 190, 299-327.

Spectral Analysis of Relativistic Atoms – Interaction with the Quantized Radiation Field

Matthias Huber

Received: February 26, 2009

Communicated by Heinz Siedentop

Abstract. This is the the second part of a series of two papers, which investigate spectral properties of Dirac operators with singular potentials. We will provide a spectral analysis of a relativistic one-electron atom in interaction with the second quantized radiation field and thus extend the work of Bach, Fröhlich, and Sigal [5] and Hasler, Herbst, and Huber [19] to such systems. In particular, we show that the lifetime of excited states in a relativistic hydrogen atom coincides with the life time given by Fermi's Golden Rule in the non-relativistic case. We will rely on the technical preparations derived in the first part [25] of this work.

2000 Mathematics Subject Classification: Primary 81V10; Secondary 35J99, 35Q40, 81Q10, 81T99
Keywords and Phrases: Dirac-Operator, Quantum Electrodynamics, Spectral Analysis, Fermi Golden Rule

1 Introduction

We continue our study of resonances for relativistic electrons and apply the results about one-particle Dirac operators with singular potentials in [25] to a relativistic Pauli-Fierz model. We prove upper and lower bounds on the lifetime of excited states for a relativistic hydrogen (-like) atom coupled to the quantized radiation field and show that it is described by Fermi's Golden Rule and coincides with the non-relativistic result in leading order in the fine structure constant α.
The spectral analysis of non-relativistic atoms in interaction with the radiation field was initiated by Bach, Fröhlich, and Sigal [4, 5]. It was carried on by Griesemer, Lieb and Loss [16], by Fröhlich, Griesemer and Schlein (see for

example [15]) and many others (see for example Hiroshima [22], Arai and Hirokawa [3], Dereziński and Gérard [12], Hiroshima and Spohn [21]), Loss, Miyao and Spohn [32] or Hasler and Herbst [18, 17]). Recently, Miyao and Spohn [35] showed the existence of a groundstate for a semi-relativistic electron coupled to the quantized radiation field.

Bach, Fröhlich, and Sigal [5] proved a lower bound on the lifetime of excited states in non-relativistic QED. Later, an upper bound was proven by Hasler, Herbst, and Huber [19] (see also [24]) and by Abou Salem et al. [1]. As in [4, 5, 19] we use the method of complex dilation. Since the corresponding operators are not normal, we are going to apply the Feshbach projection method, which was introduced in non-relativistic QED by Bach et al. [4, 5].

We describe the electron by the Coulomb-Dirac operator, projected onto its positive spectral subspace. Note that this choice is not gauge invariant. Our analysis will work for other potentials as well, as long as condition (26) holds for the difference between fine structure components, and as long the eigenfunctions have a exponential decay uniform in the velocity of light.

On a technical level the relativistic model is more difficult to handle than the nonrelativistic Pauli-Fierz model. One reason is the fine structure splitting of the eigenvalues. Moreover, due to the use of complex dilation one has to make sense of the notion of a positive spectral subspace for a non-selfadjoint operator. Finally, a factor of α is missing in front of the radiation field.

We would like to mention that the Feshbach method is named after the physicist Herman Feshbach, which used the method to deal with resonances in nuclear physics [14, Equation (2.14)]. Also Howland [23] used the Feshbach operator calling it "Livšic matrix", since Livšic [31, 30] used the method in scattering theory. Moreover, the method is known under the name "Schur complement". This name is due to Haynsworth [20], who used it in honor of the Schur determinant formula. Also Menniken and Motovilov [34, 33] use the Schur complement to treat resonances of 2×2-operator matrices. They call it "transfer function", however. For a detailed overview over the history of the Schur complement, we refer the reader to [40]. For some more references about resonances in general and the spectral analysis of (non-relativistic QED) we refer the reader to [25].

2 Model and Definitions

The (initial) Hilbert space of our model is $\mathcal{H}' := \mathcal{H}_{\mathrm{el}} \otimes \mathcal{F}$, where $\mathcal{H}_{\mathrm{el}} := L^2(\mathbb{R}^3; \mathbb{C}^4)$ is the Hilbert space for a relativistic electron and

$$\mathcal{F} := \bigoplus_{N=0}^{\infty} \mathcal{S}_N L^2[(\mathbb{R}^3 \times \mathbb{Z}_2)]^N$$

is the Fock space (with vacuum Ω) of the quantized electromagnetic field taking into account the two polarizations of the photon. $\mathcal{S}_N$ is the projection onto the subspace of functions which are symmetric under exchange of variables.

The Coulomb-Dirac operator with velocity of light $\mathfrak{c}$, Planck constant $\hbar$, electron mass m, elementary charge e, atomic number $\mathfrak{Z}$ and permittivity of the vacuum ϵ_0 is in SI units

$$D' := -\mathrm{i}\,\hbar\mathfrak{c}\boldsymbol{\alpha}\cdot\nabla + \beta mc^2 - \frac{e^2\mathfrak{Z}}{4\pi\epsilon_0}\frac{1}{|\cdot|}.$$

This operator is self-adjoint on the domain $H^1(\mathbb{R}^3;\mathbb{C}^4)$ for $\frac{e^2\mathfrak{Z}}{4\pi\epsilon_0} < \frac{\sqrt{3}}{2}c$. In the following, we will always assume that this condition is fulfilled. Actually, for technical reasons, we are even going to impose some more restrictive conditions later on (see for example Theorem 3).
We denote the positive spectral projection of this operator by $\Lambda'^{(+)}$. We will restrict the operator to its positive spectral subspace and couple it to the quantized radiation field $A'_{\kappa'}(x) := A'_{\kappa'}(x)_+ + A'_{\kappa'}(x)_-$, where $A'_{\kappa'}(x)_+$ and $A'_{\kappa'}(x)_-$ are defined as in the non-relativistic case by

$$A'_{\kappa'}(x)_+ := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,\kappa'(|k|)\sqrt{\frac{\hbar}{2\epsilon_0\mathfrak{c}|k|(2\pi)^3}}\varepsilon'_\mu(k)e^{-ik\cdot x}a'^*_\mu(k) \tag{1}$$

$$A'_{\kappa'}(x)_- := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,\kappa'(|k|)\sqrt{\frac{\hbar}{2\epsilon_0\mathfrak{c}|k|(2\pi)^3}}\varepsilon'_\mu(k)e^{\mathrm{i}\,\alpha k\cdot x}a'_\mu(k). \tag{2}$$

Here $\varepsilon'_\mu(k)$, $\mu = 1, 2$ are the polarization vectors of the photons, which depend only on the direction of k.
If we add the operator H'_f for the kinetic energy of the photons

$$H'_\mathrm{f} := \hbar\mathfrak{c}\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,|k|a'^*_\mu(k)a'_\mu(k), \tag{3}$$

we obtain (cf. [11, B-V.1., Formula (35) through (39), page 431])

$$H' := \Lambda'^{(+)}[\mathfrak{c}\boldsymbol{\alpha}\cdot(-\mathrm{i}\,\hbar\nabla - eA'_{\kappa'}(x)) + \beta m\mathfrak{c}^2 - \frac{e^2\mathfrak{Z}}{4\pi\epsilon_0}\frac{1}{|\cdot|}]\Lambda'^{(+)} + H'_f.$$

In principle, one could define the operator without restriction to the positive spectral subspace. For this case it is at least known that selfadjoint realizations exist [2, Theorem 1.2], which are, however, not explicitly known. Moreover the expression for the inverse life lifetime (see equation (21)) without UV cutoff would diverge in this case so the investigation of this operator with regard to the lifetime of excited states would not make any sense. We would like to mention that for a certain class of potentials – which does not include the Coulomb potential – it is known that the operator without projections is essentially self-adjoint on a suitable domain. (see Stockmeyer and Zenk [38] and Arai [2]).
Similar to the non-relativistic case [19, 5] we set $a_0 := \alpha^{-1}(\frac{\hbar}{m\mathfrak{c}})$ (Bohr radius), $\zeta := a_0$ and $\xi^{-1} := \frac{\alpha}{a_0}$ and scale the operator according to $x \to \zeta x$ and $k \to \xi^{-1}k$. We denote the corresponding unitary transformation by U. In

this scaling we can expect to be able to treat the operator similarly as in the non-relativistic case. We have to make the replacements

$$\hbar\nabla \to \alpha m\mathfrak{c}\nabla \qquad eA'_{\kappa'}(x) \to \alpha^{5/2}m\mathfrak{c}A_\kappa(\alpha x)$$
$$\frac{e^2}{4\pi\epsilon_0}\frac{1}{|\cdot|} \to \alpha^2 m\mathfrak{c}^2\frac{1}{|\cdot|} \qquad H'_{\mathrm{f}} \to \alpha^2 m\mathfrak{c}^2 H_{\mathrm{f}}$$

and obtain

$$\tilde{H}'_{\alpha,\mathfrak{Z}} := UH'U^* = \alpha^2 m\mathfrak{c}^2\left[\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}[D_{\alpha^{-1},\mathfrak{Z}} - \sqrt{\alpha}\boldsymbol{\alpha}\cdot A_\kappa(\alpha x)]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} + H_{\mathrm{f}}\right]. \quad (4)$$

Here

$$D_{\alpha^{-1},\mathfrak{Z}} := -\mathrm{i}\,\alpha^{-1}\boldsymbol{\alpha}\cdot\nabla + \alpha^{-2}\beta - \frac{\mathfrak{Z}}{|\cdot|}$$

with $\alpha\mathfrak{Z} < \sqrt{3}/2$ is the scaled version of the Dirac operator D'. $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}$ is the positive spectral projection of the operator $D_{\alpha^{-1},\mathfrak{Z}}$, where α^{-1} plays the role of the velocity of light after the scaling and $\mathfrak{Z}$ the role of the coupling constant. We denote the eigenvalues of this operator by $\tilde{E}_{n,l}(\alpha^{-1},\mathfrak{Z})$, where n is the principal quantum number and l numbers the eigenvalues belonging to the principal quantum number n by size *not* counting multiplicities. We have $n \in \mathbb{N}$ and $l \in \mathbb{N}$ with $l \leq n$. We set

$$E_{n,l}(\alpha^{-1},\mathfrak{Z}) := \tilde{E}_{n,l}(\alpha^{-1},\mathfrak{Z}) - c^2, \quad E_n(\infty,\mathfrak{Z}) := -\frac{\mathfrak{Z}^2}{2n^2}, \quad (5)$$

where $E_n(\infty,\mathfrak{Z})$ is the n-th eigenvalue (not counting multiplicities) of the Schrödinger operator which we obtain in the limit $\alpha \to 0$ (see [25, Section 8]). We abbreviate $E_n := E_n(\infty,\mathfrak{Z})$ and $E_{n,l}(\alpha) := E_{n,l}(\alpha^{-1},\mathfrak{Z})$ for $n \in \mathbb{N}$ and for $1 \leq l \leq n$.
H_{f} and $A_\kappa(x)$ are given by

$$H_{\mathrm{f}} := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3} \mathrm{d}k\,|k|a^*_\mu(k)a_\mu(k) \quad (6)$$

and $A_\kappa(x) := A_\kappa(x)_+ + A_\kappa(x)_-$ with

$$A_\kappa(x)_+ := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k\,\kappa(|k|)}{\sqrt{4\pi^2|k|}}\varepsilon_\mu(k)e^{-ik\cdot x}a^*_\mu(k) \quad (7)$$
$$A_\kappa(x)_- := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k\,\kappa(|k|)}{\sqrt{4\pi^2|k|}}\varepsilon_\mu(k)e^{ik\cdot x}a_\mu(k) \quad (8)$$

as in the non-relativistic case.
In the following, we will consider the operator

$$H_\alpha := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}[D_{\alpha^{-1},\mathfrak{Z}} - \alpha^{-2} - \sqrt{\alpha}\boldsymbol{\alpha}\cdot A_\kappa(\alpha x)]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} + H_{\mathrm{f}} \quad (9)$$

on $\mathcal{H} := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} L^2(\mathbb{R}^3;\mathbb{C}^4) \otimes \mathcal{F}$, where we omit trivial factors $\otimes \mathbf{1}_{\mathrm{f}}$ or $\mathbf{1}_{\mathrm{el}}\otimes$.
In order to apply the methods of the non-relativistic case (see Bach, Fröhlich, and Sigal [5] and Hasler, Herbst, and Huber [19]) with a minimal amount of changes, and in order to apply the results about the non-relativistic limit obtained in [25], we subtract the rest energy α^{-2}. As in the non-relativistic case we define the perturbation parameter $g := \alpha^{3/2} > 0$ and the perturbation operator

$$W^{(\alpha)} := \sqrt{\alpha}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} \boldsymbol{\alpha} \cdot A_\kappa(\alpha x)\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}$$

as well as the free operator

$$H_{\alpha,0} := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} D_{\alpha^{-1},\mathfrak{Z}} \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} + H_{\mathrm{f}} - \alpha^{-2}$$

and the electronic operator

$$H^{(\alpha)}_{\mathrm{el}} := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}[D_{\alpha^{-1},\mathfrak{Z}} - \alpha^{-2}]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}.$$

We will always assume $\mathfrak{Z} > 0$.
We will prove the self-adjointness of these operators in Section 3. Note that contrary to the non-relativistic case also the free operator depends on α. We suppress the dependence of the operators H_α, $H_{\alpha,0}$ and $H^{(\alpha)}_{\mathrm{el}}$ on the atomic number $\mathfrak{Z}$, since we will treat it as a fixed parameter.
Note that the prefactor of the photonic field in (9) is $\sqrt{\alpha}$ only and not $\alpha^{3/2}$ as in the non-relativistic case. Moreover, $D_{\alpha,\mathfrak{Z}}$ depends on the fine structure constant. The limit $\alpha \to 0$ corresponds in this scaling to the non-relativistic limit. In the treatment of the resonances for this operator the distance of neighbouring eigenvalues may vanish as $\alpha \to 0$ so that the estimates on the Feshbach operator (see below) have to be improved. Nevertheless we will use the perturbation parameter $g = \alpha^{3/2}$.
As in [5, 19], we will make use of (complex) dilations of the above operators: We define

$$\begin{aligned} H^{(\alpha)}_{\mathrm{el}}(\theta) &:= \mathcal{U}_{\mathrm{el}}(\theta) H^{(\alpha)}_{\mathrm{el}} \mathcal{U}_{\mathrm{el}}(\theta)^{-1},\ H_g(\theta) := \mathcal{U}(\theta) H_g \mathcal{U}(\theta)^{-1} \text{ and} \\ W_g(\theta) &:= \mathcal{U}(\theta) W_g \mathcal{U}(\theta)^{-1} \end{aligned} \tag{10}$$

for real θ, where $\mathcal{U}(\theta)$ is the unitary group associated to the generator of dilations. It is defined in such a way that the coordinates of the electron are dilated as $x_j \mapsto e^\theta x_j$ and the momenta of the photons as $k \mapsto e^{-\theta}k$. In this way we obtain the operator

$$H^{(\alpha)}_{\mathrm{el}}(\theta) := \mathcal{U}_{\mathrm{el}}(\theta) H^{(\alpha)}_{\mathrm{el}} \mathcal{U}_{\mathrm{el}}(\theta)^{-1} = \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)[D_{\alpha^{-1},\mathfrak{Z}}(\theta) - \alpha^{-2}]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)$$

on $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta) L^2(\mathbb{R}^3;\mathbb{C}^4)$, which is selfadjoint on $\mathrm{Dom}(H^{(\alpha)}_{\mathrm{el}}(\theta)) =$

$= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) H^1(\mathbb{R}^3;\mathbb{C}^4)$, as well as the operators

$$\begin{aligned} H_\alpha(\theta) &:= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta)[D_{\alpha^{-1},\mathfrak{z}}(\theta) - \alpha^{-2} - \sqrt{\alpha}\boldsymbol{\alpha}\cdot A^{(\theta)}_\kappa(\alpha x)]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) + e^{-\theta} H_{\mathrm{f}} \\ W^{(\alpha)}(\theta) &:= \sqrt{\alpha}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta)[\boldsymbol{\alpha}\cdot A^{(\theta)}_\kappa(\alpha x)]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) \\ H_{\alpha,0}(\theta) &:= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) D_{\alpha^{-1},\mathfrak{z}}(\theta)\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) + e^{-\theta} H_{\mathrm{f}} - \alpha^{-2} \end{aligned}$$

on $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) L^2(\mathbb{R}^3;\mathbb{C}^4)\otimes\mathcal{F}$, where $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta)$ has been defined in [25] even for non-real θ. Here $A^{(\theta)}_\kappa(x) := A^{(\theta)}_\kappa(x)_+ + A^{(\theta)}_\kappa(x)_-$, where

$$A^{(\theta)}_\kappa(x)_+ := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3} \mathrm{d}k\, G^{(\bar\theta)}_x(k,\mu)^* a^*_\mu(k)$$

and

$$A^{(\theta)}_\kappa(x)_- := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3} \mathrm{d}k\, G^{(\theta)}_x(k,\mu) a_\mu(k)$$

with

$$G^{(\theta)}_x(k,\mu) := \frac{e^{-\theta}\kappa(e^{-\theta}|k|)}{\sqrt{4\pi^2|k|}} e^{\mathrm{i}\,k\cdot x}\epsilon_\mu(k).$$

We will show in Section 3 that these operators admit a holomorphic continuation to certain values of θ. Moreover, we define

$$W^{(\alpha)}_{1,0}(\theta) := \sqrt{\alpha}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta)\left[\boldsymbol{\alpha}\cdot A^{(\theta)}_\kappa(\alpha x)_+\right]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) \tag{11}$$

$$W^{(\alpha)}_{0,1} := \sqrt{\alpha}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta)\left[\boldsymbol{\alpha}\cdot A^{(\theta)}_\kappa(\alpha x)_-\right]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) \tag{12}$$

$$w_{0,1}(k,\mu;\theta) := \sqrt{\alpha}\boldsymbol{\alpha}\cdot G^{(\theta)}_{\alpha x}(k,\mu) \tag{13}$$

$$w_{1,0}(k,\mu;\theta) := w_{0,1}(k,\mu;\bar\theta)^*. \tag{14}$$

Using the notation from [25, Section 5] we define the projections

$$\begin{aligned} P^{(\alpha)}_{\mathrm{el},n,l}(\theta) &:= P_{n,l}(\alpha^{-1},\mathfrak{z};\theta) & P^{(\alpha)}_{\mathrm{el},n,l} &:= P_{n,l}(\alpha^{-1},\mathfrak{z};0) \\ P^{(\alpha)}_{\mathrm{el},n}(\theta) &:= P_{n}(\alpha^{-1},\mathfrak{z};\theta) & P^{(\alpha)}_{\mathrm{el},n} &:= P_{n}(\alpha^{-1},\mathfrak{z};0) \\ \overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta) &:= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) - P^{(\alpha)}_{\mathrm{el},n}(\theta) & \overline{P}^{(\alpha)}_{\mathrm{el},n} &:= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}} - P^{(\alpha)}_{\mathrm{el},n} \\ \overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) &:= P^{(\alpha)}_{\mathrm{el},n}(\theta) - P^{(\alpha)}_{\mathrm{el},n,l}(\theta) & \overline{P}^{(\alpha)}_{\mathrm{el},n,l} &:= P^{(\alpha)}_{\mathrm{el},n} - P^{(\alpha)}_{\mathrm{el},n,l} \\ \underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) &:= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) - P^{(\alpha)}_{\mathrm{el},n,l}(\theta) & \underline{P}^{(\alpha)}_{\mathrm{el},n,l} &:= \Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}} - P^{(\alpha)}_{\mathrm{el},n,l} \end{aligned}$$

as operators on $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}}(\theta) L^2(\mathbb{R}^3;\mathbb{C}^4)$ and $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{z}} L^2(\mathbb{R}^3;\mathbb{C}^4)$, respectively. Moreover, we need for a $\eta > 0$ such that $\tilde{E}_{\tilde n,\tilde n}(\alpha^{-1},\mathfrak{z}) < \alpha^{-2} - \eta$ and $\tilde{E}_{\tilde n+1,1}(\alpha^{-1},\mathfrak{z}) > \alpha^{-2} - \eta$ for some $\tilde n \in \mathbb{N}$ (see [25, Section 7]) the projections

$$P_{\mathrm{disc}}(\alpha;\theta) := P_{\mathrm{disc},\tilde n}(\alpha^{-1},\mathfrak{z};\theta) = \sum_{1\le n'\le\tilde n} P_{n'}(\alpha^{-1},\mathfrak{z};\theta) \tag{15}$$

and

$$\bar{P}_{\mathrm{disc}}(\alpha;\theta) := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta) - P_{\mathrm{disc}}(\alpha;\theta) \tag{16}$$

as operators on $\mathrm{Ran}\,\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)$ as well. η is chosen in such a way that $\tilde{n} > n$, where n is the principal quantum number whose life-time we are interested in.
For $\rho > 0$ (to be specified later) we define the projections

$$P_{n,l}(\theta) := P^{(\alpha)}_{\mathrm{el},n,l} \otimes \chi_{H_{\mathrm{f}} \leq \rho}, \quad \overline{P}_{n,l}(\theta) := \mathbf{1} - P_{n,l}(\theta)$$

and for $R > 0$

$$\overline{P}_{n,l}(\theta;R) := P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \chi_{H_{\mathrm{f}}+R>\rho} + \underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \mathbf{1}_{\mathrm{f}}$$

as operators on $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)\otimes\mathcal{F}$.
As in [5, 19], the main technical tool in our analysis is the Feshbach operator

$$\begin{aligned}\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta)-z) := {}& P_{n,l}(\theta)(H_\alpha(\theta)-z)P_{n,l}(\theta) - P_{n,l}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l}(\theta) \\ & \times [\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}\overline{P}_{n,l}(\theta)W^{(\alpha)}(\theta)P_{n,l}(\theta),\end{aligned} \tag{17}$$

which we define as an operator on $\mathrm{Ran}\,P_{n,l}(\theta)$. Note that we need the Feshbach operator for *each* fine structure component of the considered principal quantum number n, i.e. for all $1 \leq l \leq n$. Note moreover that we do not distinguish between the operators PAP and $PAP|_{\mathrm{Ran}\,P}$ when we write PAP, where A is a closed operator an P a projection with $\mathrm{Dom}\,A \subset \mathrm{Ran}\,P$. The meaning of this expression will be clear from the context.
We will show below that the Feshbach operator can be approximated in a certain sense by the operators

$$\begin{aligned}Z^{od}_{n,l,\pm}(\alpha) := {}& \lim_{\epsilon\downarrow 0}\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\, P^{(\alpha)}_{\mathrm{el},n,l}w_{0,1}(k,\mu;0) \\ & \times \underline{P}^{(\alpha)}_{\mathrm{el},n,l}\left[\underline{P}^{(\alpha)}_{\mathrm{el},n,l}H^{(\alpha)}_{\mathrm{el}} - E_{n,l}(\alpha) + |k| \pm \mathrm{i}\,\epsilon\right]^{-1}\underline{P}^{(\alpha)}_{\mathrm{el},n,l}w_{1,0}(k,\mu;0)P^{(\alpha)}_{\mathrm{el},n,l}\end{aligned} \tag{18}$$

and

$$Z^{d}_{n,l}(\alpha) := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k}{|k|}P^{(\alpha)}_{\mathrm{el},n,l}w_{0,1}(k,\mu;0)P^{(\alpha)}_{\mathrm{el},n,l}w_{1,0}(k,\mu;0)P^{(\alpha)}_{\mathrm{el},n,l} \tag{19}$$

as well as

$$Z_{n,l,\pm}(\alpha) := Z^{d}_{n,l}(\alpha) + Z^{od}_{n,l,\pm}(\alpha), \tag{20}$$

defined as operators on $\mathrm{Ran}\,P^{(\alpha)}_{\mathrm{el},n,l}$. These operators are the relativistic analoga of [19, Equations (3) and (4)]. Note that $\mathcal{U}_{\mathrm{el}}(\theta)$ restricted to $\mathrm{Ran}\,P^{(\alpha)}_{\mathrm{el},n,l}$ is a similarity transformation ([25, Lemma 9]).

It is easy to see that the imaginary part of $Z_{n,l,\pm}(\alpha)$ is given by (cf. Equation (11) in Remark 1 in [19])

$$\begin{aligned}\operatorname{Im} Z_{n,l,\pm}(\alpha) = \mp\, \pi \sum_{\substack{n',l':\\ E_{n',l'}(\alpha)<E_{n,l}(\alpha)}} \sum_{\mu=1,2} \int_{|\omega|=1} d\omega\, \left(E_{n',l'}(\alpha) - E_{n,l}(\alpha)\right)^2 \\ \times P^{(\alpha)}_{\mathrm{el},n,l} w_{0,1}((E_{n,l}(\alpha) - E_{n',l'}(\alpha))\omega, \mu; 0) P^{(\alpha)}_{\mathrm{el},n',l'} \\ \times w_{1,0}((E_{n,l}(\alpha) - E_{n',l'}(\alpha))\omega, \mu; 0) P^{(\alpha)}_{\mathrm{el},n,l}.\end{aligned}$$

It will turn out that the lifetime in lowest order in the fine structure constant α is given by the same expression as in the non-relativistic case (see Lemma 10). Therefore, we define (cf. [19, Equation (12)])

$$Z_{n,l,\mathrm{im}} = g^2 \frac{2}{3} \sum_{\substack{1\le n'<n\\ 1\le l\le n}} (-E_{n'} + E_n)^3 \times \kappa(|-E_{n'} + E_n|)^2 P^{(0)}_{\mathrm{el},n,l} x P^{(0)}_{\mathrm{el},n',l'} x P^{(0)}_{\mathrm{el},n,l} \tag{21}$$

and

$$Y_{n,l,\pm}(\alpha) := \mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0)^{-1} \operatorname{Re} Z_{n,l}(\alpha) \mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0) \mp \mathrm{i}\, Z_{n,l,\mathrm{im}} \tag{22}$$

as operators on $\operatorname{Ran} P^{(0)}_{\mathrm{el},n,l}$, where we defined $\operatorname{Ran} P^{(0)}_{\mathrm{el},n,l} := \mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0)^{-1} \times P^{(\alpha)}_{\mathrm{el},n,l} \mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0)$. $\mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0)$ is the unitary transformation which corresponds to taking the non-relativistic limit (see [25, Section 8]). We set

$$Z_{n,l}(\alpha) := Z_{n,l,-}(\alpha), \quad Y_{n,l}(\alpha) := Y_{n,l,-}(\alpha).$$

Note that contrary to [19] the coupling constant g is contained in the definition of the objects $Z_{n,l}(\alpha)$, $Y_{n,l}(\alpha)$ and so on. We see from Equation (21) that transitions between fine structure components of a principal quantum number do not play a role in lowest order in α.

Note that we remove the dependence on α only from the imaginary part, since a discussion of the real part, which yields the Lamb shift [28, 6], does not make sense without an UV renormalization. Moreover, the Lamb shift is not important for lifetime measurements using the so called "beam-foil"-method [10, 13, 7, 8].

We can now formulate our main result: Fix $n > 2$. Since $Z_{n,l,\mathrm{im}}$ is obtained from the corresponding matrix in the nonrelativistic case by restricting the corresponding quadratic form to $\operatorname{Ran} P^{(0)}_{\mathrm{el},n,l}$, we see immediately that in this case $Z_{n,l,\mathrm{im}}$ is strictly positive for all $1 \le l \le n$ (see [19, Appendix B.3]). Note that this is not the case for $n = 2$ due to the metastability of the 2s-sates of hydrogen. Indeed we will need in our proof the Feshbach operator and the matrices $Z_{n,l,\pm}(\alpha)$ and $Y_{n,l,\pm}(\alpha)$ for *all* fine structure components of the corresponding principal quantum number and not only for the fine structure component, whose lifetime we are interested in.

Theorem 1. *Let* $n > 2$ *and* $\phi(\alpha)$ *a normalized eigenvector of* $H_{\mathrm{el}}^{(\alpha)}$ *with eigenvalue* $E_{n,l}(\alpha)$, $\psi(\alpha) := \phi(\alpha)\otimes\Omega$ *and* $\phi(0) := \mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{Z}; 0)^{-1}\phi(\alpha)$. *Then there is a* $C > 0$ *such that for all* $\alpha > 0$ *small enough and all* $s \geq 0$

$$\langle\psi(\alpha), e^{-isH_\alpha}\psi(\alpha)\rangle = \langle\phi(0), e^{-is(E_{n,l}(\alpha)-Y_{n,l}(\alpha))}\phi(0)\rangle + b(g,s)$$

holds, where $|b(g,s)| \leq C\sqrt{\alpha}$.

We will prove Theorem 1 in Section 7.

Remark 1. *If we compare Definition* (22) *of* $Y_{n,l}(\alpha)$ *with [19, Formula (12)] we see that the lifetime of an excited state in the relativistic model is the same as in the Pauli-Fierz model. Thus relativistic effects play a minor role for electric dipole transitions. But there seems to be a small relativistic contribution for the decay of the metastable* $2s$*-state of hydrogen (see Breit and Teller [9]).*

3 Selfadjointness and Dilation Analyticity

Before we can turn to the operator H_α in the following sections we have to prove its selfadjointness and the holomorphicity properties of the operators $H_\alpha(\theta)$.

Theorem 2. *Let* $0 < \alpha\mathfrak{Z} < \sqrt{3}/2$. *Then the following holds: The operator*

$$H_\alpha : \mathcal{D} \subset (\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} L^2(\mathbb{R}^3;\mathbb{C}^4)) \otimes \mathrm{Dom}(H_{\mathrm{f}}) \to (\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} L^2(\mathbb{R}^3;\mathbb{C}^4)) \otimes \mathcal{F}$$

is on $\mathcal{D} := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} H^1(\mathbb{R}^3;\mathbb{C}^4))\hat{\otimes}\,\mathrm{Dom}(H_{\mathrm{f}})$ *essentially selfadjoint, where* $\hat{\otimes}$ *denotes the algebraic tensor product.*

Proof. Because of [39, Theorem 4.4] the operator $H_{\mathrm{el}}^{(\alpha)} + \alpha^{-2}$ is selfadjoint and positive on the domain $\mathrm{Dom}(H_{\mathrm{el}}^{(\alpha)}) = \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} H^1(\mathbb{R}^3;\mathbb{C}^4)$. Since H_{f} is selfadjoint and positive on a suitable domain $\mathrm{Dom}(H_{\mathrm{f}})$, it follows from [36, Theorem VIII.33] that $H_{\alpha,0} + \alpha^{-2}$ is essentially selfadjoint and positive on the (algebraic) tensor product $\mathcal{D} = \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}} H^1(\mathbb{R}^3;\mathbb{C}^4)\hat{\otimes}\,\mathrm{Dom}(H_{\mathrm{f}})$. We have for all $\psi \in \mathcal{D}$ and all $\epsilon > 0$ with a $C > 0$ (see for example [5, Proof of Lemma 1.1])

$$\begin{aligned}&\|W^{(\alpha)}\psi\| \leq C\sqrt{\alpha}\|(H_{\mathrm{f}}+1)^{1/2}\psi\| \leq C\sqrt{\alpha}(\|\psi\| + \sqrt{\|\psi\|\|H_{\mathrm{f}}\psi\|})\\ &\leq C\sqrt{\alpha}[(1+\frac{1}{2\epsilon})\|\psi\| + \frac{\epsilon}{2}\|H_{\mathrm{f}}\psi\|] \leq C\sqrt{\alpha}[(1+\frac{1}{2\epsilon})\|\psi\| + \frac{\epsilon}{2}\|(H_{\alpha,0}+\alpha^{-2})\psi\|].\end{aligned}$$

Thus $W^{(\alpha)}$ is infinitesimally $(H_{\alpha,0}+\alpha^{-2})$-bounded, and in turn $H_\alpha+\alpha^{-2}$ (and thus also H_α) is essentially selfadjoint on $\mathrm{Dom}(H_{\alpha,0})$. ◻

We denote the operators defined in Theorem 2 again by H_α and $H_{\alpha,0}$ respectively.

We turn to the operators $H_\alpha(\theta)$ and $H_{\alpha,0}(\theta)$ on the domain $\mathrm{Dom}(H_\alpha(\theta)) = \mathrm{Dom}(H_{\alpha,0}(\theta)) = \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)H^1(\mathbb{R}^3;\mathbb{C}^4)\hat{\otimes}\,\mathrm{Dom}(H_\mathrm{f})$. In the following theorem we show that the families of operators

$$\begin{aligned} &U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)H_\alpha(\theta)U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}, \\ &\qquad\qquad U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)H_{\alpha,0}(\theta)U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}, \end{aligned} \tag{23}$$

defined on the Hilbert space $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}L^2(\mathbb{R}^3;\mathbb{C}^4)\otimes\mathcal{F}$ with domain $U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta) \times \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)H^1(\mathbb{R}^3;\mathbb{C}^4)\hat{\otimes}\,\mathrm{Dom}(H_\mathrm{f})$, are holomorphic families of type (B) on a suitable domain. Here $U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)$ is the transformation function between positive spectral projections of $D_{\alpha^{-1},\mathfrak{Z}}$ and $D_{\alpha^{-1},\mathfrak{Z}}(\theta)$ defined in [25, Theorem 6]. We will write $U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)$ for the operator $U_\mathrm{DL}(\alpha^{-1},\mathfrak{Z};\theta)\otimes\mathbf{1}_\mathrm{f}$.

Theorem 3. *Let $\theta \in S_{\pi/4}$, $2\alpha\mathfrak{Z}C(\mathrm{Im}\,\theta) < 1$, $C_\mathrm{DL}|\theta| < q$ and $C_\mathrm{DLS}|\theta| < q$ for some $0 < q < 1$, where the constants C_DL and C_DLS are defined in [25, Section 6] and $C(\mathrm{Im}\,\theta)$ is defined in [25, Section 4]. Then there is a $\theta_0 > 0$ independent of $0 < \alpha \le 1$ such that for all $|\theta| \le \theta_0$ the operators* (23) *define holomorphic families of operators $\tilde{H}_\alpha(\theta)$ bzw. $\tilde{H}_{\alpha,0}(\theta)$ of type (B) on a suitable domain $\mathrm{Dom}(\tilde{H}_\alpha(\theta)) = \mathrm{Dom}(\tilde{H}_{\alpha,0}(\theta))$. These operators are m-sectorial.*

Proof. The expression $q_{\alpha^{-1},0}(\psi) := \langle\psi,(D_{\alpha^{-1},\mathfrak{Z}}\otimes\mathbf{1}+\mathbf{1}\otimes H_\mathrm{f})\psi\rangle$ for $\psi\in\mathcal{D}$ is a positive closable quadratic form whose closure $\tilde{q}_{\alpha^{-1},0}$ defines a selfadjoint operator which coincides with the operator $H_{\alpha,0}$ defined in Theorem 2. We have $\mathrm{Dom}(\tilde{q}_{\alpha^{-1},0}) = \mathrm{Dom}((H_{\alpha,0}+\alpha^{-2})^{1/2})$. In particular, for $\psi\in\mathrm{Dom}(\tilde{q}_{\alpha^{-1},0})$ the estimate

$$\begin{aligned} \||D_{\alpha^{-1},\mathfrak{Z}}|^{1/2}\psi\| &= \|(\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}D_{\alpha^{-1},\mathfrak{Z}}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}})^{1/2}\psi\| \\ &\le \|(\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}D_{\alpha^{-1},\mathfrak{Z}}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}\otimes\mathbf{1}+\mathbf{1}\otimes H_\mathrm{f})^{1/2}\psi\| < \infty \end{aligned}$$

holds, and in the same way we see $\|(H_\mathrm{f}+1)^{1/2}\psi\| < \infty$.
Thus, we find for $\psi\in\mathrm{Dom}(\tilde{q}_{\alpha^{-1},0})$

$$\begin{aligned} &\langle\psi, U_\mathrm{DL}(\alpha^{-1};\theta)D_{\alpha^{-1},\mathfrak{Z}}(\theta)U_\mathrm{DL}(\alpha^{-1};\theta)^{-1}\psi\rangle \\ &=\langle|D_{\alpha^{-1},\mathfrak{Z}}|^{1/2}\psi, |D_{\alpha^{-1},\mathfrak{Z}}|^{-1/2}|D_{\alpha^{-1},0}|^{1/2} \\ &\quad\times|D_{\alpha^{-1},0}|^{-1/2}U_\mathrm{DL}(\alpha^{-1};\theta)D_{\alpha^{-1},\mathfrak{Z}}(\theta)U_\mathrm{DL}(\alpha^{-1};\theta)^{-1}|D_{\alpha^{-1},0}|^{-1/2} \\ &\quad\times|D_{\alpha^{-1},0}|^{1/2}|D_{\alpha^{-1},\mathfrak{Z}}|^{-1/2}|D_{\alpha^{-1},\mathfrak{Z}}|^{1/2}\psi\rangle. \end{aligned} \tag{24}$$

[25, Lemma 5 and Lemma 6] imply

$$|\langle\psi,U_\mathrm{DL}(\alpha^{-1};\theta)D_{\alpha^{-1},\mathfrak{Z}}(\theta)U_\mathrm{DL}(\alpha^{-1};\theta)^{-1}\psi\rangle-\langle\psi,D_{\alpha^{-1},\mathfrak{Z}}\psi\rangle|\le C|\theta|\langle\psi,D_{\alpha^{-1},\mathfrak{Z}}\psi\rangle$$

with some $C>0$ independent of α and θ. Moreover, $|e^{-\theta}\langle\psi,H_\mathrm{f}\psi\rangle-\langle\psi,H_\mathrm{f}\psi\rangle| \le B|\theta|\langle\psi,H_\mathrm{f}\psi\rangle$ with $B := e^{\pi/4}$. Since $\|W^{(\alpha)}(\theta)(H_\mathrm{f}+1)^{-1/2}\| \le \sqrt{\alpha}C_1$ with some

$C_1 > 0$ independent of θ and α (see for example [5, Proof of Lemma 1.1]) we obtain for all $\epsilon > 0$

$$\begin{aligned}&|\langle\psi, U_{\mathrm{DL}}(\alpha^{-1};\theta)W^{(\alpha)}(\theta)U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}\psi\rangle|\\ &\qquad\le \sqrt{\alpha}C_1(1+C_{\mathrm{DL}}|\theta|)^2[(1/\epsilon^2+\epsilon^2)\|\psi\|^2+\epsilon^2\langle\psi, H_{\mathrm{f}}\psi\rangle]. \quad (25)\end{aligned}$$

It follows that the quadratic form $p_{\alpha^{-1};\theta}(\psi) := \langle\psi,(H_\alpha+\alpha^{-2})\psi\rangle$ for $\psi \in \mathrm{Dom}(\tilde{q}_{\alpha^{-1},0})$ is well defined for sufficiently small $|\theta|$. If we choose $|\theta|$ so small that $(C+B)|\theta| < 1$ holds, and then in (25) $\epsilon > 0$ small enough (depending on θ), we see that the quadratic form $p_{\alpha^{-1};\theta} - \tilde{q}_{\alpha,0}$ is relatively $\tilde{q}_{\alpha,0}$-bounded with form bound smaller than 1.
Because of [27, Theorem VI-1.33] the quadratic form $p_{\alpha^{-1};\theta}$ is closed with $\mathrm{Dom}(p_{\alpha^{-1};\theta}) = \mathrm{Dom}(\tilde{q}_{\alpha^{-1},0})$ and sectorial. Moreover,

$$\begin{aligned}&|D_{\alpha^{-1},0}|^{-1/2}U_{\mathrm{DL}}(\alpha^{-1};\theta)D_{\alpha^{-1},\mathfrak{Z}}(\theta)U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}|D_{\alpha^{-1},0}|^{-1/2}\\ =&|D_{\alpha^{-1},0}|^{-1/2}U_{\mathrm{DL}}(\alpha^{-1};\theta)|D_{\alpha^{-1},0}|^{1/2}|D_{\alpha^{-1},0}|^{-1/2}D_{\alpha^{-1},\mathfrak{Z}}(\theta)|D_{\alpha^{-1},0}|^{-1/2}\\ &\times|D_{\alpha^{-1},0}|^{1/2}U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}|D_{\alpha^{-1},0}|^{-1/2}.\end{aligned}$$

Using Equation (24) and [25, Theorem 6 c)] we see that the expression $\langle\psi, U_{\mathrm{DL}}(\alpha^{-1};\theta)D_{\alpha^{-1},\mathfrak{Z}}(\theta)U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}\psi\rangle$ for all $\psi \in \mathrm{Dom}(p_{\alpha^{-1};\theta})$ is a holomorphic function of θ. It is easy to see that

$$(H_{\mathrm{f}}+1)^{-1/2}U_{\mathrm{DL}}(\alpha^{-1};\theta)W^{(\alpha)}(\theta)U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}(H_{\mathrm{f}}+1)^{-1/2}$$

is bounded-holomorphic. Thus $\langle\psi, U_{\mathrm{DL}}(\alpha^{-1};\theta)W^{(\alpha)}(\theta)U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}\psi\rangle$ is holomorphic function of θ. It follows that $p_{\alpha^{-1};\theta}(\psi)$ is a holomorphic function of θ for all $\psi \in \mathrm{Dom}(p_{\alpha^{-1};\theta}) = \mathrm{Dom}(\tilde{q}_{\alpha^{-1},0})$. The family of m-sectorial operators defined by these quadratic forms is a holomorphic family of type (B) (see [27, Chapter VII-4.2]). The proof for the operator without interaction works analogously. Since $\|W^{(\alpha)}(\theta)(H_{\mathrm{f}}+1)^{-1/2}\| \le \sqrt{\alpha}C_1$ (see above), is infinitesimally operator bounded with respect to the free operator which implies the equality of the domains. ∎

Remark 2. *The above proof also shows that the operators* $U_{\mathrm{DL}}(\alpha^{-1};\theta)D_{\alpha^{-1},\mathfrak{Z}}(\theta)U_{\mathrm{DL}}(\alpha^{-1};\theta)^{-1}|_{\mathrm{Ran}\,\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}}$ *on the space* $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}L^2(\mathbb{R}^3;\mathbb{C}^4)$ *are sectorial for sufficiently small* $|\theta|$. *In particular, the assumptions of the Ichinose Lemma (see [37, Corollary 2 on page 183] or [26]) are fulfilled so that*

$$\begin{aligned}&\sigma(D_{\alpha^{-1},\mathfrak{Z}}(\theta)|_{\mathrm{Ran}\,\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)}\otimes\mathbf{1}_{\mathrm{f}}+e^{-\theta}\mathbf{1}_{\mathrm{el}}\otimes H_{\mathrm{f}})=\\ &\qquad\qquad=\sigma(D_{\alpha^{-1},\mathfrak{Z}}(\theta)|_{\mathrm{Ran}\,\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)})+e^{-\theta}\sigma(H_{\mathrm{f}})\end{aligned}$$

holds.

In the following, we will consider $U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}\tilde{H}_\alpha(\theta)U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)$ and $U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}\tilde{H}_{\alpha,0}(\theta)U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)$ on the respective domains $U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}\,\mathrm{Dom}(\tilde{H}_\alpha(\theta))$ and $U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}\,\mathrm{Dom}(\tilde{H}_{\alpha,0}(\theta))$. We will denote these operators by $H_\alpha(\theta)$ and $H_{\alpha,0}(\theta)$ again.

4 Technical Lemmata

In this section we will formulate and prove all technical statements which we will need to show the existence of the Feshbach operator and in order to approximate it by suitable operators.

Using the dilation analyticity we can restrict to $\theta = \mathrm{i}\,\vartheta$ with $0 < \vartheta < \theta_0$. We choose θ_0 so small that the statements of Theorem 3 as well as the statements in [25, Appendix A] hold. Moreover, we choose for this θ_0 a $\alpha_0 > 0$ so small that the statements about the "nonrelativistic limit" of the operator $D_{\alpha^{-1},\mathfrak{Z}}(\theta)$ (proven in [25, Section 8]) and inequality (26) hold. In particular, all projections occurring in the following are uniformly bounded in α and θ.

We put

$$\delta_{n,l,\pm}(\alpha) := \begin{cases} |E_{n,l}(\alpha) - E_{n,l\pm 1}(\alpha)|/2 & 1 < l < n \\ |E_{n,l}(\alpha) - E_{n,l+1}(\alpha)|/2 & l = 1 \\ |E_{n,l}(\alpha) - E_{n,l-1}(\alpha)|/2 & l = n \end{cases}$$

$$\delta_{n,l}(\alpha) := \min\{\delta_{n,l,+}(\alpha), \delta_{n,l,-}(\alpha)\}, \qquad \delta_{n,\pm} := |E_n - E_{n\pm 1}|/2,$$

$$\delta_n := \min\{\delta_{n,+}, \delta_{n,-}\}.$$

Note that $\delta_{n,l}(\alpha) = \delta_{n,l,\pm}(\alpha)$ holds for $l = 1$ or $l = n$. We will suppress the dependence of these quantities on α in certain places in order to simplify notation. It follows from the explicit form of the eigenvalues (see [29]) that for all $\alpha < \alpha_0$ with $\alpha_0 > 0$ small enough the inequality

$$c_1\alpha^2 \leq \delta_{n,l,\pm}(\alpha) \leq c_2\alpha^2 \tag{26}$$

holds with two constants $0 < c_1 < c_2$ independent of α and l.

We choose $\rho, \sigma > 0$ and define the sets (see Figure 1)

$$\mathcal{A}^{<}_{n,l}(\alpha,\sigma) := [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha), E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)] + \mathrm{i}\,[-\sigma,\infty), \quad 1 \leq l \leq n$$

and

$$\mathcal{A}_{n,l}(\alpha,\sigma) := \begin{cases} [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha), E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)] + \mathrm{i}\,[-\sigma,\infty) & 1 < l < n \\ [E_n - \delta_{n,-}, E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)] + \mathrm{i}\,[-\sigma,\infty) & l = 1 \\ [E_n - \delta_{n,l,-}(\alpha), E_n + \delta_{n,+}] + \mathrm{i}\,[-\sigma,\infty) & l = n \end{cases}.$$

Note that for $1 < l < n$ the identity $\mathcal{A}^{<}_{n,l}(\alpha,\sigma) = \mathcal{A}_{n,l}(\alpha,\sigma)$ holds. Moreover, following [5] we define $B_\theta(\rho) := \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)[H^{(\alpha)}_{\mathrm{el}}(\theta) - E_{n,l}(\alpha) + e^{-\theta}(H_{\mathrm{f}} + \rho)]\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)$ as an operator on the Hilbert space $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4) \otimes \mathcal{F}$ with domain $\mathrm{Dom}(B_\theta(\rho)) = U_{\mathrm{DL}}(\alpha^{-1},\mathfrak{Z};\theta)^{-1}\,\mathrm{Dom}(\tilde{H}_{\alpha,0}(\theta))$. The operator is a densely defined and closed operator (cf. Theorem 3 and the remarks following it). It follows that $B_\theta(\rho)^*$ is densely defined as well and we have $B_\theta(\rho)^{**} = B_\theta(\rho)$. Note that the adjoint is to be taken with respect to the scalar product on the Hilbert space $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)\otimes\mathcal{F}$. In particular, $B_\theta(\rho)^*\neq B_{\bar\theta}(\rho)$. As in the Pauli-Fierz model $B_\theta(\rho)$ is only an auxiliary object, which saves some combinatorics. In principle, one could prove all statements without using $B_\theta(\rho)$. Note that all norms, scalar products and adjoints are to be understood in the sense of $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)$ or $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)\otimes\mathcal{F}$. We will choose ρ and σ later on as suitable functions of the coupling constant g. At the moment, we assume only that σ and ρ are nonnegative and bounded by some constant from above.

In the proofs in this and the following section, C denotes a generic, positive constant, which does not depend on α and z, but perhaps on ϑ.

In the following lemmas we will prove some estimates on the resolvents of the free operator $H_{\alpha,0}$ and of the electronic operator $H^{(\alpha)}_{\mathrm{el}}$. The lemmas generalize similar statements and their proofs [5]. Due to the fine structure splitting and the missing power of α some additional difficulties have to be addressed.

Lemma 1. *Let $0<\vartheta<\theta_0$. Then the following statements hold:*

a) *There is a $C>0$ such that for all $\alpha\leq\alpha_0$, all $\sigma\leq\frac{\delta_{n,l}(\alpha)\sin\vartheta}{2\cos\vartheta}$, all $R>0$ and all $z\in\mathcal{A}_{n,l}(\alpha,\sigma)$*

$$\Big\|\big[\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_{\mathrm{f}}(H^{(\alpha)}_{\mathrm{el}}(\theta)+e^{-\theta}(H_{\mathrm{f}}+R)-z)\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_{\mathrm{f}}\big]^{-1}\\ \times\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_{\mathrm{f}}\Big\|\leq\frac{C}{\delta_{n,l}(\alpha)\sin\vartheta}\quad(27)$$

holds.

b) *There is a $C>0$ such that for all $\rho>0$, all $\sigma\leq\frac{\rho\sin\vartheta}{2}$, all $R\geq\rho$ and all $z\in\mathcal{A}_{n,l}(\alpha,\sigma)$*

$$\Big\|\big[\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_{\mathrm{f}}(H^{(\alpha)}_{\mathrm{el}}(\theta)+e^{-\theta}(H_{\mathrm{f}}+R)-z)\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_{\mathrm{f}}\big]^{-1}\\ \times\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_{\mathrm{f}}\Big\|\leq\frac{C}{R\sin\vartheta}\quad(28)$$

holds.

c) *There is a $C>0$ such that for all $\alpha\leq\alpha_0$, all $\sigma\leq\frac{\delta_n\sin\vartheta}{2\cos\vartheta}$, all $R>0$ and all $z\in\mathcal{A}_{n,l}(\alpha,\sigma)$*

$$\Big\|\big[\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes\mathbf{1}_{\mathrm{f}}(H^{(\alpha)}_{\mathrm{el}}(\theta)+e^{-\theta}(H_{\mathrm{f}}+R)-z)\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes\mathbf{1}_{\mathrm{f}}\big]^{-1}\\ \times\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes\mathbf{1}_{\mathrm{f}}\Big\|\leq\frac{C}{\delta_n\sin\vartheta}.\quad(29)$$

Proof.

a) We split the projection and the resolvent according to the formula $\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) = \sum_{1\le l'\le n,\, l'\neq l} P^{(\alpha)}_{\mathrm{el},n,l'}(\theta)$ and use the representation (spectral theorem) in which H_{f} is the multiplication with the variable r. In order to simplify the notation we will suppress the dependence of the eigenvalues $E_{n,l'}(\alpha)$ on α. Note that for $E_{n,l'} < E_{n,l}$

$$|E_{n,l'} - z + e^{-\theta}(r+R)| \geq \mathrm{Im}\,(e^{\theta}(z - E_{n,l'})) \geq \\ \geq -(\cos\vartheta)\sigma + \sin\vartheta(\mathrm{Re}\, z - E_{n,l'}) \geq \frac{\sin\vartheta\delta_{n,l}(\alpha)}{2} \tag{30}$$

and for $E_{n,l'} > E_{n,l}$

$$|E_{n,l'} - z + e^{-\theta}(r+R)| \;\geq\; \mathrm{Re}\,(E_{n,l'} - z + e^{-\theta}(r+R)) \;\geq\; \delta_{n,l}(\alpha) \tag{31}$$

holds, which proves the claim together with [25, Corollary 5]. For $l = 1$ and $l = n$ the estimates (30) and (31) respectively are not needed. We used in the first estimate that $(\cos\vartheta)\sigma \leq \frac{\sin\vartheta\delta_{n,l}(\alpha)}{2}$.

b) We estimate $\mathrm{Im}\,(-E_{n,l'} + z - e^{-\theta}(r+R)) \geq -\sigma + \sin\vartheta(r+R) \geq \frac{\sin\vartheta R}{2}$, where we used $\sigma \leq \frac{\sin\vartheta R}{2}$.

c) We split the projection $\overline{P}^{(\alpha)}_{\mathrm{el},n} = \bar{P}_{\mathrm{disc}}(\alpha;\theta) + \sum_{\substack{1\le n'\le\tilde{n}\\ n'\neq n}} P^{(\alpha)}_{\mathrm{el},n'}(\theta)$ according to (15) and obtain analogously to the proof of a) the estimate $|\frac{1}{E_{n,l'}-z+e^{-\theta}(r+R)}| \leq \frac{C}{\delta_n \sin\vartheta}$ and with [25, Corollary 4]

$$\left\| \left[\bar{P}_{\mathrm{disc}}(\alpha;\theta)(H^{(\alpha)}_{\mathrm{el}}(\theta) + e^{-\theta}(H_{\mathrm{f}} + R) - z)\bar{P}_{\mathrm{disc}}(\alpha;\theta)\otimes \mathbf{1}_{\mathrm{f}}\right]^{-1}\bar{P}_{\mathrm{disc}}(\alpha;\theta)\right\| \\ \leq \sup_{r>0} \frac{C}{-\eta - (\mathrm{Re}\, z - (r+R))} \leq \frac{C}{\delta_n}.$$

□

LEMMA 2. *Let $0 < \vartheta < \theta_0$. Then the following statements hold:*

a) There is a $C > 0$ such that for all $\alpha \leq \alpha_0$, all $R > 0$, all $\sigma \leq \min\{\frac{\delta_{n,l}(\alpha)\sin\vartheta}{2\cos\vartheta}, \frac{\delta_n \sin\vartheta}{2\cos\vartheta}, 1/2\rho\sin\vartheta\}$ and all $z \in \mathcal{A}_{n,l}(\alpha,\sigma)$

$$\left\| \left[\overline{P}_{n,l}(\theta;R)(H^{(\alpha)}_{\mathrm{el}}(\theta) + e^{-\theta}(H_{\mathrm{f}} + R) - z)\overline{P}_{n,l}(\theta;R)\right]^{-1}\overline{P}_{n,l}(\theta;R)\right\| \leq \\ \leq \frac{C}{\min\{\delta_n, \delta_{n,l}(\alpha), \rho\}\sin\vartheta} \tag{32}$$

b) There is a $C > 0$ such that for all $\alpha \leq \alpha_0$, all $\sigma \leq \min\{\frac{\delta_{n,l}(\alpha)\sin\vartheta}{2\cos\vartheta}$,

$\frac{\delta_n \sin\vartheta}{2\cos\vartheta}, 1/2\rho\sin\vartheta\}$ *and all* $z \in \mathcal{A}_{n,l}(\alpha,\sigma)$

$$\left\| \left[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\right]^{-1}\overline{P}_{n,l}(\theta)B_\theta(\rho)\right\| \leq \\ \leq \frac{C}{\sin\vartheta}\left(1+\frac{\rho}{\min\{\delta_{n,l}(\alpha),\delta_n\}}\right) \tag{33}$$

holds.

Proof.
a) We split the projection

$$\overline{P}_{n,l}(\theta;R) = \overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes \mathbf{1}_\mathrm{f} + \overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes \mathbf{1}_\mathrm{f} + P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes \chi_{H_\mathrm{f}+R\geq\rho}.$$

For $r+R\geq\rho$ we estimate as follows: $\mathrm{Im}\,(-E_{n,l}+z-e^{-\theta}(r+R)) \geq -\sigma + \sin\vartheta(r+R) \geq \frac{\sin\vartheta\rho}{2}$. We used here $\sigma \leq 1/2\rho\sin\vartheta$ and $r+R\geq\rho$. This shows the claim together with (27) and (29) in Lemma 1.
b) As before, we split $\overline{P}_{n,l}(\theta) = \overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes\mathbf{1}_\mathrm{f}+\overline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\mathbf{1}_\mathrm{f}+P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_\mathrm{f}\geq\rho}$.
We start with

$$\left\|\left[P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_\mathrm{f}\geq\rho}(H_{\alpha,0}(\theta)-z)\right]^{-1}P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_\mathrm{f}\geq\rho}B_\theta(\rho)\right\| \\ = \sup_{r\geq\rho}\left|\frac{e^{-\theta}(r+\rho)}{E_{n,l}-z+e^{-\theta}r}\right|\|P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\| \leq \frac{4}{\sin\vartheta}\|P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\|,$$

where we used the inequality

$$\mathrm{Im}\,(-E_{n,l}+z-e^{-\theta}r) \geq -\sigma+\sin\vartheta r \geq \frac{\sin\vartheta r}{2}, \tag{34}$$

which follows from $\sigma\leq\frac{\sin\vartheta\rho}{2}$ and $\rho\leq r$.
Using Equations (30) and (31) from the proof of Lemma 1 as well as Equation (34) we obtain with some $C>0$ (independent of α)

$$\begin{aligned}&\left|\frac{E_{n,l'}-E_{n,l}+e^{-\theta}(r+\rho)}{E_{n,l'}-z+e^{-\theta}r}\right| \leq \left|\frac{E_{n,l'}-E_{n,l}+e^{-\theta}\rho}{E_{n,l'}-z+e^{-\theta}r}\right|+\left|\frac{e^{-\theta}r}{E_{n,l'}-z+e^{-\theta}r}\right| \\ \leq & C\frac{2(\alpha^2+\rho)}{\sin\vartheta\delta_{n,l}(\alpha)}+\begin{cases}\frac{2\rho}{\sin\vartheta\delta_{n,l}(\alpha)}, & r\leq\rho\\ \frac{2r}{\sin\vartheta r}, & r>\rho\end{cases} \\ \leq & C\frac{4}{\sin\vartheta}(1+\frac{\rho}{\delta_{n,l}(\alpha)}).\end{aligned}$$

Analogously, we obtain for $n'\neq n$ the estimate $\left|\frac{E_{n',l'}-E_{n,l}+e^{-\theta}(r+\rho)}{E_{n',l'}-z+e^{-\theta}r}\right| \leq$

$C\frac{4}{\sin\vartheta}(1+\frac{\rho}{\delta_n})$. Eventually we find

$$\begin{aligned}
&\left\|\left[\bar{P}_{\mathrm{disc}}(\alpha;\theta)(H_{\alpha,0}(\theta)-z)\right]^{-1}\bar{P}_{\mathrm{disc}}(\alpha;\theta)B_\theta(\rho)\right\| \\
\leq &\|\bar{P}_{\mathrm{disc}}(\alpha;\theta)\| + \sup_{r\geq 0}\left\|\frac{z-E_{n,l}-e^{-\theta}\rho}{\bar{P}_{\mathrm{disc}}(\alpha;\theta)(H_{\mathrm{el}}^{(\alpha)}(\theta)-z+e^{-\theta}(r+\rho_0))}\bar{P}_{\mathrm{disc}}(\alpha;\theta)\right\| \\
\leq &\|\bar{P}_{\mathrm{disc}}(\alpha;\theta)\| + \sup_{r\geq 0}\frac{\max\{\delta_{n,-},\delta_{n,+}\}+\rho}{\mathrm{Re}\,(-\eta-z)+\cos\vartheta r} \leq \frac{C}{\delta_n},
\end{aligned}$$

using [25, Corollary 4]. □

Part b) of the above lemma and the following lemmas are preparations for the proof of relative bounds on the interaction.

COROLLARY 1. *Let $0<\vartheta<\theta_0$. Then there is a $C>0$ such that for all $\alpha\leq\alpha_0$, all $\sigma\leq\min\{\frac{\delta_{n,l}(\alpha)\sin\vartheta}{2\cos\vartheta},\frac{\delta_n\sin\vartheta}{2\cos\vartheta},1/2\rho\sin\vartheta\}$ and all $z\in\mathcal{A}_{n,l}(\alpha,\sigma)$*

$$\||B_\theta(\rho)|^{1/2}[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho)^*|^{1/2}\| \leq \frac{C}{\sin\vartheta}$$

holds.

Proof. We find

$$\begin{aligned}
&\left\||B_\theta(\rho)|\left[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\right]^{-1}\overline{P}_{n,l}(\theta)\right\| \\
&\quad=\left\|\left[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\right]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho)^*|\right\|.
\end{aligned}$$

The claim follows by complex interpolation and using Lemma 2 b). □

LEMMA 3. *Let $0<\vartheta<\theta_0$. Then there is a $C>0$ such that for all $0<\alpha<\alpha_0$ and all $\rho>0$ the following statements hold:*

a)

$$\left\|\overline{P}_{\mathrm{el},n}^{(\alpha)}(\theta)B_\theta(\rho)^{-1}\right\| \leq \frac{C}{\sin\vartheta} \tag{35}$$

b)

$$\left\|B_\theta(\rho)^{-1}\right\| \leq \frac{C}{\sin\vartheta}\left(1+\frac{1}{\rho}\right) \tag{36}$$

c)

$$\left\|H_{\mathrm{f}}B_\theta(\rho)^{-1}\right\| \leq \frac{C}{\sin\vartheta} \tag{37}$$

Proof.
a) We estimate using [25, Corollary 4]

$$\begin{aligned}&\|\bar{P}_{\rm disc}(\alpha;\theta)B_\theta(\rho)^{-1}\|\\ &\le \sup_{r\ge 0}\frac{C}{-\eta - E_{n,l}+\cos\vartheta(r+\rho)} \le \frac{C}{\delta_n}\end{aligned}$$

and note that analogously to the Formulas (30) and (31) we find for $n'<n$

$$\begin{aligned}|E_{n',l'}-E_{n,l}+e^{-\theta}(r+\rho)| &\ge \mathrm{Im}\,(e^{\theta}(z-E_{n,l}))\\ &\ge -(\cos\vartheta)\sigma+\sin\vartheta(E_{n,l}-E_{n',l'}) \ge \sin\vartheta\delta_n \end{aligned} \quad (38)$$

and for $n'>n$

$$|E_{n',l'}-E_{n,l}+e^{-\theta}(r+\rho)| \ \ge\ \mathrm{Re}\,(E_{n',l'}-E_{n,l}+e^{-\theta}(r+\rho)) \ \ge\ \delta_n, \quad (39)$$

which proves the claim.
b) In view of part a) it suffices to show the estimate on $\mathrm{Ran}\, P_{\mathrm{el},n}^{(\alpha)}(\theta)$. We find for all $1\le n'\le\tilde{n}$ and all $1\le l\le n'$, in particular for $n'=n$,

$$|E_{n',l'}(\alpha)-E_{n,l}(\alpha)+e^{-\theta}(r+\rho)| \ge \sin\vartheta(r+\rho)\ge\rho\sin\vartheta, \quad (40)$$

which proves the claim.
c) Using Formula (40) we obtain for all $1\le n'\le\tilde{n}$ and all $1\le l\le n'$

$$\frac{r}{|E_{n',l'}(\alpha)-E_{n,l}(\alpha)+e^{-\theta}(r+\rho)|}\le\frac{r}{\sin\vartheta(r+\rho)}\le\frac{1}{\sin\vartheta},$$

which prove the claim on $\mathrm{Ran}\, P_{\rm disc}(\alpha;\theta)$. Using [25, Corollary 4] we find on $\mathrm{Ran}\,\bar{P}_{\rm disc}(\alpha;\theta)$

$$\left\|H_{\rm f}B_\theta(\rho)^{-1}\bar{P}_{\rm disc}(\alpha;\theta)\right\| \le C\sup_{r\ge0}\frac{r}{\eta-E_{n,l}+\cos\vartheta(r+\rho)}\le C\frac{1}{\cos\vartheta}.$$

Note that $|\sin\vartheta|<\cos\vartheta$ for $|\vartheta|<\pi/4$. ∎

Corollary 2. *Let $0<\vartheta<\theta_0$. Then there is a $C>0$ such that for all $0<\alpha<\alpha_0$ and all $\rho>0$ the following estimates hold:*

a)

$$\|\overline{P}_{\mathrm{el},n}^{(\alpha)}(\theta)|B_\theta(\rho)|^{-1/2}\|\le\frac{C}{\sqrt{\sin\vartheta}},\quad \||B_\theta(\rho)^*|^{-1/2}\overline{P}_{\mathrm{el},n}^{(\alpha)}(\theta)\|\le\frac{C}{\sqrt{\sin\vartheta}}$$

b)

$$\||B_\theta(\rho)|^{-1/2}\|\le\frac{C}{\sqrt{\sin\vartheta}}\left(1+\frac{1}{\sqrt{\rho}}\right),\quad \||B_\theta(\rho)^*|^{-1/2}\|\le\frac{C}{\sqrt{\sin\vartheta}}\left(1+\frac{1}{\sqrt{\rho}}\right)$$

c)

$$\|H_{\mathrm{f}}^{1/2}|B_\theta(\rho)|^{-1/2}\| \le \frac{C}{\sqrt{\sin\vartheta}}, \qquad \|H_{\mathrm{f}}^{1/2}|B_\theta(\rho)^*|^{-1/2}\| \le \frac{C}{\sqrt{\sin\vartheta}}$$

Proof.
a) We find $\|\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)|B_\theta(\rho)|^{-1/2}\|^2 \le \|\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\|\|\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)B_\theta(\rho)^{-1}\|$ as well as $\||B_\theta(\rho)^*|^{-1/2}\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\|^2 \le \|\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)^*\|\|B_\theta(\rho)^{-1}\overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\|$. The claim follows now from Lemma 3.
b) This follows immediately from the spectral theorem for self-adjoint operators.
c) From Formula (37) in Lemma 3 we obtain for all $\psi \in \mathrm{Dom}(B_\theta(\rho))$ the estimate $\|H_{\mathrm{f}}\psi\| \le \frac{C}{\min\{\sin\vartheta,\cos\vartheta\}}\|B_\theta(\rho)\psi\|$. Taking the square root of this operator inequality, the claim follows. The second inequality follows analogously using the identity $\|H_{\mathrm{f}}B_\theta(\rho)^{-1}\| = \|[B_\theta(\rho)^*]^{-1}H_{\mathrm{f}}\| = \|H_{\mathrm{f}}[B_\theta(\rho)^*]^{-1}\|$. □

In the last two lemmas in this section, we prove relative bounds on the interaction. In comparison to the non-relativistic case, we have the additional difficulty that the factor in front of the interaction is $\sqrt{\alpha}$ only. To circumvent this problem, we use the statements about the non-relativistic limit shown in [25].

LEMMA 4. *Let $0 < \vartheta < \theta_0$. Then there is a $C > 0$ such that for all $0 < \alpha < \alpha_0$ and all $\rho > 0$ the estimate*

$$\left\||B_\theta(\rho)^*|^{-1/2}W^{(\alpha)}(\theta)|B_\theta(\rho)|^{-1/2}\right\| \le \frac{C}{\sin\vartheta}\sqrt{\alpha}\left[1+\alpha\left(1+\frac{1}{\rho^{1/2}}\right)\right]$$

holds.

Proof. We split the projection according to $\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta) = P_1(\theta) + P_2(\theta)$, where $P_1(\theta) = P^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes\mathbf{1}_{\mathrm{f}},\ P_2(\theta) = \overline{P}^{(\alpha)}_{\mathrm{el},n}(\theta)\otimes\mathbf{1}_{\mathrm{f}}$. Since the estimate with $A^{(\theta)}_\kappa(\alpha x)_+$ works analogously, we consider $A^{(\theta)}_\kappa(\alpha x)_-$ only. We find for $\psi,\psi' \in \Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)\otimes\mathcal{F}$ and $i,j\in\{1,2\}$

$$\begin{aligned}\Big|\Big\langle\psi', |B_\theta(\rho)^*|^{-1/2}P_i(\theta)\boldsymbol{\alpha}\cdot A^{(\theta)}_\kappa(\alpha x)_- P_j(\theta)|B_\theta(\rho)|^{-1/2}\psi\Big\rangle\Big|& \\ \le \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k\,|\kappa(e^{-\theta}|k|)|}{\sqrt{4\pi^2|k|}}\left\|P_i(\theta)^*|B_\theta(\rho)^*|^{-1/2}\psi'\right\|& \\ \times\left\|P_i(\theta)\boldsymbol{\alpha}\cdot\varepsilon_\mu(k)e^{\mathrm{i}\,\alpha k\cdot x}P_j(\theta)\right\|\left\|a_\mu(k)P_j(\theta)|B_\theta(\rho)|^{-1/2}\psi\right\|.& \quad (41)\end{aligned}$$

We have to make a case distinction:

Case 1: $i = j = 2$. Using Corollary 2 a) we find $\||B_\theta(\rho)^*|^{-1/2}P_i(\theta)\| \leq C$. Moreover $\|P_i(\theta)\boldsymbol{\alpha}\cdot\varepsilon_\mu(k)e^{\mathrm{i}\,\alpha k\cdot x}P_j(\theta)\| \leq C$. The r.h.s. of Formula (41) can be estimated by

$$C\|\psi'\|\|\psi\|\|H_{\mathrm{f}}^{1/2}|B_\theta(\rho)|^{-1/2}\| \leq \frac{C}{\sqrt{\sin\vartheta}}\|\psi'\|\|\psi\| \tag{42}$$

with a generic $C > 0$, where we used Corollary 2 c) in the last step.
Case 2: All other combinations of i and j. ¿From [25, Lemma 10 or Theorem 11] it follows that $\|P_i(\theta)\boldsymbol{\alpha}\cdot\varepsilon_\mu(k)e^{\mathrm{i}\,\alpha k\cdot x}P_j(\theta)\| \leq C\alpha(1+\alpha|k|)$, and from Corollary 2 a) and b) that $\||B_\theta(\rho)^*|^{-1/2}P_i(\theta)\| \leq \frac{C}{\sqrt{\sin\vartheta}}(1+\frac{1}{\rho^{1/2}})$. The r.h.s. of Formula (41) can be estimated by

$$\alpha\frac{C}{\sqrt{\sin\vartheta}}\left(1+\frac{1}{\rho^{1/2}}\right)\|\psi'\|\sqrt{\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k\,|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|^2}} \tag{43}$$

$$\times\sqrt{\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,|k|\|a_\mu(k)P_j(\theta)|B_\theta(\rho)|^{-1/2}\psi\|^2} \leq \alpha\frac{C}{\sin\vartheta}\left(1+\frac{1}{\rho^{1/2}}\right)\|\psi'\|\|\psi\|$$

in this case with a generic $C > 0$. □

Lemma 5. *Let $0 < \vartheta < \theta_0$. Then there is a $C > 0$ such that for all $0 < \alpha < \alpha_0$ and all $\rho > 0$ the following estimates hold:*

a)

$$\left\||B_\theta(\rho)^*|^{-1/2}W_{1,0}^{(\alpha)}(\theta)P_{n,l}(\theta)\right\| \leq \frac{C}{\sqrt{\sin\vartheta}}g \tag{44}$$
$$\left\|P_{n,l}(\theta)W_{0,1}^{(\alpha)}(\theta)|B_\theta(\rho)|^{-1/2}\right\| \leq \frac{C}{\sqrt{\sin\vartheta}}g$$

b)

$$\left\||B_\theta(\rho)^*|^{-1/2}W_{0,1}^{(\alpha)}(\theta)P_{n,l}(\theta)\right\| \leq \frac{C}{\sqrt{\sin\vartheta}}g\rho^{1/2} \tag{45}$$
$$\left\|P_{n,l}(\theta)W_{1,0}^{(\alpha)}(\theta)|B_\theta(\rho)|^{-1/2}\right\| \leq \frac{C}{\sqrt{\sin\vartheta}}g\rho^{1/2}$$

c)

$$\left\|W_{0,1}^{(\alpha)}(\theta)P_{n,l}(\theta)\right\| \leq Cg\rho, \qquad \left\|P_{n,l}(\theta)W_{1,0}^{(\alpha)}(\theta)\right\| \leq Cg\rho \tag{46}$$

Proof. We begin with

$$\left\||B_\theta(\rho)^*|^{-1/2}W_{0,1}^{(\alpha)}(\theta)P_{n,l}(\theta)\right\|$$
$$\leq \sqrt{\alpha}\left\||B_\theta(\rho)^*|^{-1/2}\right\|\left\|\Lambda_{\alpha^{-1},\mathfrak{Z}}^{(+)}(\theta)\boldsymbol{\alpha}\cdot A_\kappa^{(\theta)}(\alpha x)_- P_{n,l}(\theta)\right\|$$

and find with [25, Theorem 11] similarly as in [4, Lemma IV.9.]

$$\begin{aligned}&\left|\left\langle\psi', \Lambda^{(+)}_{\alpha^{-1},3}(\theta)\boldsymbol{\alpha}\cdot A^{(\theta)}_{\kappa}(\alpha x)_{-}P_{n,l}(\theta)\psi\right\rangle\right|\\
&\leq \sum_{\mu=1}^{2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k\,|\kappa(e^{-\theta}|k|)|}{\sqrt{4\pi^2|k|}}\left|\left\langle\psi', \Lambda^{(+)}_{\alpha^{-1},3}(\theta)\boldsymbol{\alpha}\cdot\varepsilon_{\mu}(k)e^{\mathrm{i}\,\alpha k\cdot x}P^{(\alpha)}_{\mathrm{el},n,l}a_{\mu}(k)\chi_{H_{\mathrm{f}}\leq\rho}\psi\right\rangle\right|\\
&\leq C\alpha\sqrt{\sum_{\mu=1}^{2}\int_{|k|\leq\rho}\mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|^2}}\sqrt{\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}|k|\|a_{\mu}(k)\chi_{H_{\mathrm{f}}\leq\rho}\psi\|^2}\\
&\leq \alpha\rho\|\psi\|\|\psi'\|. \qquad (47)\end{aligned}$$

For $\|P_{n,l}(\theta)W^{(\alpha)}_{1,0}(\theta)|B_\theta(\rho)|^{-1/2}\|$ one shows a similar estimate such that the claim in b) follows from Corollory 2. Formula (47) and an analogous calculation for $W^{(\alpha)}_{1,0}(\theta)$ prove the claim in c).
To show a) we estimate similarly as in Formula (47)

$$\begin{aligned}&|\langle\psi', P_{n,l}(\theta)W^{(\alpha)}_{0,1}(\theta)|B_\theta(\rho)|^{-1/2}\psi\rangle\\
&\qquad\leq C\sqrt{\alpha}\alpha\|\psi'\|\|H_{\mathrm{f}}^{1/2}|B_\theta(\rho)|^{-1/2}\|\|\psi\|\leq\frac{C}{\sqrt{\sin\theta}}g\|\psi'\|\|\psi\|.\end{aligned}$$

The estimate on $\||B_\theta(\rho)^*|^{-1/2}W^{(\alpha)}_{1,0}(\theta)P_{n,l}(\theta)\|$ follows analogously. □

5 Existence and approximation of the Feshbach Operator

We set now $\rho_0 = g^{4/3} = \alpha^2$ and $\sigma_0 = g^{5/3} = \alpha^{5/2}$ and use the estimates from Section 4 for $\rho = \rho_0$ and $\sigma = \sigma_0$.
We apply the strategy from [5], but have to overcome additional difficulties. First, we generalize [5, Lemma 3.14] to the relativistic case and show the existence of the inverse $[\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}$.

Lemma 6. *Let $0<\vartheta<\theta_0$. Then there is a $C>0$ such that for all sufficiently small $\alpha>0$ the following holds: The operator $\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)$ is for all $z\in\mathcal{A}_{n,l}(\alpha,\sigma_0)$ invertible on $\operatorname{Ran}\overline{P}_{n,l}(\theta)$, and we have*

$$\|[\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}\overline{P}_{n,l}(\theta)\|\leq\frac{C}{\sin^2\vartheta\rho_0}.$$

Proof. The claim follows from the series expansion

$$\begin{aligned}&\left\|\left[\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)\right]^{-1}\overline{P}_{n,l}(\theta)\right\|\\
=&\left\|\sum_{n=0}^{\infty}\left[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\right]^{-1}\overline{P}_{n,l}(\theta)\right.\\
&\left.\times\left[-W^{(\alpha)}(\theta)\left[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\right]^{-1}\overline{P}_{n,l}(\theta)\right]^n\right\|\end{aligned}$$

$$
\begin{aligned}
=&\Big\| \sum_{n=0}^{\infty} |B_\theta(\rho_0)|^{-1/2} \\
&\times |B_\theta(\rho_0)|^{+1/2} \big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2} \\
&\times \Big[-|B_\theta(\rho_0)^*|^{-1/2} W^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2} \\
&\times |B_\theta(\rho_0)|^{+1/2} \big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\Big]^n \\
&\times |B_\theta(\rho_0)^*|^{-1/2}\Big\| \\
\le& \Big\| |B_\theta(\rho_0)|^{-1/2}\Big\| \, \Big\| |B_\theta(\rho_0)^*|^{-1/2}\Big\| \\
&\times \Big\| |B_\theta(\rho_0)|^{+1/2} \big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\Big\| \\
&\times \sum_{n=0}^{\infty} \Big[\Big\| |B_\theta(\rho_0)^*|^{-1/2} W^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\Big\| \\
&\times \Big\| |B_\theta(\rho_0)|^{+1/2} \big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\Big\| \Big]^n \\
\le& \frac{C}{\sin^2\vartheta\sqrt{\rho_0}\sqrt{\rho_0}} \sum_{n=0}^{\infty} \Big[\frac{C}{\sin^2\vartheta}\sqrt{\alpha}\big(1+\alpha\big(1+\frac{1}{\sqrt{\rho_0}}\big)\big)\Big]^n \\
\le& \frac{C}{\sin^2\vartheta\rho_0} \sum_{n=0}^{\infty} \Big(\frac{C}{\sin^2\vartheta}\sqrt{\alpha}\Big)^n
\end{aligned}
$$

with a generic $C > 0$ independent of z and α. We used Corollary 2 b), Corollary 1 and Lemma 4. □

We turn now to the existence of the Feshbach operator and generalize [5, Lemma 3.15].

LEMMA 7. *Let $0 < \vartheta < \theta_0$ small enough. Then there is a $C > 0$ such that for all sufficiently small $\alpha > 0$ and all $z \in \mathcal{A}_{n,l}(\alpha, \sigma_0)$ the following estimates hold:*

a)

$$
\Big\| P_{n,l}(\theta) W^{(\alpha)}(\theta) [\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}\overline{P}_{n,l}(\theta)\Big\| \le g\frac{C}{\sin^2\vartheta\sqrt{\rho_0}}. \tag{48}
$$

$$
\Big\| [\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}\overline{P}_{n,l}(\theta) W^{(\alpha)}(\theta) P_{n,l}(\theta)\Big\| \le g\frac{C}{\sin^2\vartheta\sqrt{\rho_0}}. \tag{49}
$$

b) For all $1 \le l, l', l'' \le n$ we have

$$
\begin{aligned}
\Big\| P_{n,l'}(\theta) W^{(\alpha)}(\theta) \overline{P}_{n,l}(\theta) [\overline{P}_{n,l}(\theta)(H_\alpha(\theta)-z)\overline{P}_{n,l}(\theta)]^{-1}& \\
\times \overline{P}_{n,l}(\theta) W^{(\alpha)}(\theta) P_{n,l''}(\theta)\Big\| &\le \frac{C}{(\sin\vartheta)^2} g^2
\end{aligned} \tag{50}
$$

c) The Feshbach operator, defined in equation (17), *exists for all* $z \in \mathcal{A}_{n,l}(\alpha, \sigma_0)$ *and fulfills the equation*

$$
\begin{aligned}
&(H_\alpha(\theta) - z)^{-1} = \\
&= \Big[P_{n,l}(\theta) - \overline{P}_{n,l}(\theta)\big(\overline{P}_{n,l}(\theta)H_\alpha(\theta)\overline{P}_{n,l}(\theta) - z\big)^{-1}\overline{P}_{n,l}(\theta)W^{(\alpha)}(\theta)P_{n,l}(\theta)\Big] \\
&\qquad \times \big[\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - z)\big]^{-1} \\
&\times \Big[P_{n,l}(\theta) - P_{n,l}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l}(\theta)\big(\overline{P}_{n,l}(\theta)H_\alpha(\theta)\overline{P}_{n,l}(\theta) - z\big)^{-1}\overline{P}_{n,l}(\theta)\Big] \\
&\qquad + \overline{P}_{n,l}(\theta)\big(\overline{P}_{n,l}(\theta)H_\alpha(\theta)\overline{P}_{n,l}(\theta) - z\big)^{-1}\overline{P}_{n,l}(\theta),
\end{aligned} \tag{51}
$$

where the l.h.s. exists if and only if the r.h.s. exists.

Proof.
a) We obtain as in the proof of Lemma 6

$$
\begin{aligned}
&\Big\|P_{n,l}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l}(\theta)\big[\overline{P}_{n,l}(\theta)(H_\alpha(\theta) - z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)\Big\| \\
\le &\Big\|P_{n,l}(\theta)W^{(\alpha)}(\theta)|B_\theta(\rho)|^{-1/2}\Big\| \\
&\times \Big\||B_\theta(\rho)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta) - z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho)^*|^{1/2}\Big\| \\
&\times \sum_{n=0}^{\infty}\Big[\||B_\theta(\rho)^*|^{-1/2}W^{(\alpha)}(\theta)|B_\theta(\rho)|^{-1/2}\| \\
&\times \Big\||B_\theta(\rho)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta) - z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho)^*|^{1/2}\Big\|\Big]^n\| \\
&\times \||B_\theta(\rho)^*|^{-1/2}\| \\
\le &g\frac{C}{\sin^2\vartheta\sqrt{\rho_0}}\sum_{n=0}^{\infty}\left(\frac{C}{\sin^2\vartheta}\sqrt{\alpha}\right)^n,
\end{aligned}
$$

where we used additionally Lemma 5 a) and b). The other estimate follows analogously.
b) Follows similarly as in a).
c) This follows from Lemma 6 and Part a) of [4, Theorem IV.1]. ☐

Having shown the existence of the Feshbach operator, we can turn now to its approximation by suitable other operators. The aim is to control its numerical range and gain thus information about its invertability.
We define the operator

$$
\begin{aligned}
Q_{n,l}^{(\alpha)}(z;\theta) := \sum_{\mu=1,2}\int_{k\in\mathbb{R}^3} &\mathrm{d}k\, P_{n,l}(\theta)[w_{0,1}(k,\mu;\theta)\otimes \mathbf{1}_\mathrm{f}] \\
&\times \left[\frac{\overline{P}_{n,l}(\theta;|k|)}{H_\mathrm{el}^{(\alpha)}(\theta) + e^{-\mathrm{i}\,\vartheta}(H_\mathrm{f} + |k|) - z}\right][w_{1,0}(k,\mu;\theta)\otimes \mathbf{1}_\mathrm{f}]P_{n,l}(\theta)
\end{aligned}
$$

as operator on $\operatorname{Ran} P_{n,l}(\theta)$ for $z \in \mathcal{A}_{n,l}(\alpha, \sigma_0)$. Futhermore, we define θ-dependent versions of the operators $Z_{n,l,\pm}(\alpha)$ (cf. [19, Equation (8)]). We set for $\operatorname{Im}\theta \neq 0$

$$
\begin{aligned}
Z_{n,l}(\alpha;\theta) :=& \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} \mathrm{d}k\, P^{(\alpha)}_{\mathrm{el},n,l}(\theta) w_{0,1}(k,\mu;\theta) \underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) \\
&\times \left[\underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) H^{(\alpha)}_{\mathrm{el}}(\theta) - E_{n,l}(\alpha) + e^{-\theta}|k|\right]^{-1} \underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) w_{1,0}(k,\mu;\theta) P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \\
&+ \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} \frac{\mathrm{d}k}{e^{-\theta}|k|} P^{(\alpha)}_{\mathrm{el},n,l}(\theta) w_{0,1}(k,\mu;\theta) P^{(\alpha)}_{\mathrm{el},n,l}(\theta) w_{1,0}(k,\mu;\theta) P^{(\alpha)}_{\mathrm{el},n,l}(\theta).
\end{aligned}
$$

We have $Z_{n,l}(\alpha;\theta) = \mathcal{U}_{\mathrm{el}}(\theta) Z_{n,l,-}(\alpha) \mathcal{U}_{\mathrm{el}}(\theta)^{-1}$ for $\operatorname{Im}\theta > 0$ and $Z_{n,l}(\alpha;\theta) = \mathcal{U}_{\mathrm{el}}(\theta) Z_{n,l,+}(\alpha) \mathcal{U}_{\mathrm{el}}(\theta)^{-1}$ for $\operatorname{Im}\theta < 0$. Moreover, we define the following remainder terms:

$$
\begin{aligned}
&\mathrm{Rem}_0 := \\
&\quad P_{n,l}(\theta) W^{(\alpha)}(\theta) \overline{P}_{n,l}(\theta) [\overline{P}_{n,l}(\theta)(H_\alpha(\theta) - z)\overline{P}_{n,l}(\theta)]^{-1} \overline{P}_{n,l}(\theta) W^{(\alpha)}(\theta) P_{n,l}(\theta) \\
&- P_{n,l}(\theta) W^{(\alpha)}(\theta) \overline{P}_{n,l}(\theta) [\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta) - z)\overline{P}_{n,l}(\theta)]^{-1} \overline{P}_{n,l}(\theta) W^{(\alpha)}(\theta) P_{n,l}(\theta)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Rem}_1 := \\
&\quad P_{n,l}(\theta) W^{(\alpha)}(\theta) \overline{P}_{n,l}(\theta) [\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta) - z)\overline{P}_{n,l}(\theta)]^{-1} \overline{P}_{n,l}(\theta) W^{(\alpha)}(\theta) P_{n,l}(\theta) \\
&- P_{n,l}(\theta) W^{(\alpha)}_{0,1}(\theta) \overline{P}_{n,l}(\theta) [\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta) - z)\overline{P}_{n,l}(\theta)]^{-1} \overline{P}_{n,l}(\theta) W^{(\alpha)}_{1,0}(\theta) P_{n,l}(\theta)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Rem}_2 := \\
&= P_{n,l}(\theta) W^{(\alpha)}_{0,1}(\theta) \overline{P}_{n,l}(\theta) [\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta) - z)\overline{P}_{n,l}(\theta)]^{-1} \overline{P}_{n,l}(\theta) W^{(\alpha)}_{1,0}(\theta) P_{n,l}(\theta) \\
&\qquad - Q^{(\alpha)}_{n,l}(z;\theta)
\end{aligned}
$$

$$
\mathrm{Rem}_3 := P_{n,l}(\theta) W^{(\alpha)}(\theta) P_{n,l}(\theta)
$$

We generalize Lemma [5, Lemma 3.16] (see also [19, Lemma A.7]).

Lemma 8. *Let $0 < \vartheta < \theta_0$. Then there is a $C > 0$ such that for all sufficiently small $\alpha > 0$ and all $z \in \mathcal{A}_{n,l}(\alpha, \sigma_0)$ the estimate*

$$
\|[\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - z) - (H^{(\alpha)}_{\mathrm{el}}(\theta) - z + e^{-\theta} H_{\mathrm{f}} - Q^{(\alpha)}_{n,l}(z;\theta)) P_{n,l}(\theta)]\| \leq \frac{C}{\sin^4\vartheta} g^2 \sqrt{\alpha}
$$

holds.

Proof. We begin with the estimate on Rem_0:

$$
\|\mathrm{Rem}_0\| \leq \sum_{n=1}^{\infty} \left\| P_{n,l}(\theta) W^{(\alpha)}(\theta) |B_\theta(\rho_0)|^{-1/2} \right.
$$

$$
\begin{aligned}
&\times |B_\theta(\rho_0)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\\
&\times \Big[-|B_\theta(\rho_0)^*|^{-1/2}W^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\\
&\times |B_\theta(\rho_0)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\Big]^n\\
&\times |B_\theta(\rho_0)^*|^{-1/2}W^{(\alpha)}(\theta)P_{n,l}(\theta)\Big\|\\
\le&\big\|P_{n,l}(\theta)W^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\big\|\big\||B_\theta(\rho_0)^*|^{-1/2}W^{(\alpha)}(\theta)P_{n,l}(\theta)\big\|\\
&\times \Big\||B_\theta(\rho_0)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\Big\|\\
&\times \sum_{n=1}^{\infty}\big[\big\||B_\theta(\rho_0)^*|^{-1/2}W^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\big\|\\
&\times \big\||B_\theta(\rho_0)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\big\|\big]^n\\
\le&\frac{C}{\sin^2\vartheta}g^2\sum_{n=1}^{\infty}\Big(\frac{C}{\sin^2\vartheta}\sqrt{\alpha}\Big)^n\le\frac{C}{\sin^4\vartheta}g^2\sqrt{\alpha}
\end{aligned}
$$

We used here Lemma 5 a) and b), Lemma 4 and Corollary 1. For Rem_1 we find

$$
\begin{aligned}
\|\mathrm{Rem}_1\|\le&\Big\||B_\theta(\rho_0)|^{+1/2}\big[\overline{P}_{n,l}(\theta)(H_{\alpha,0}(\theta)-z)\overline{P}_{n,l}(\theta)\big]^{-1}\overline{P}_{n,l}(\theta)|B_\theta(\rho_0)^*|^{1/2}\Big\|\\
&\times\Big(\big\|P_{n,l}(\theta)W_{1,0}^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\big\|\big\||B_\theta(\rho_0)^*|^{-1/2}W_{1,0}^{(\alpha)}(\theta)P_{n,l}(\theta)\big\|\\
&+\big\|P_{n,l}(\theta)W_{1,0}^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\big\|\big\||B_\theta(\rho_0)^*|^{-1/2}W_{0,1}^{(\alpha)}(\theta)P_{n,l}(\theta)\big\|\\
&+\big\|P_{n,l}(\theta)W_{0,1}^{(\alpha)}(\theta)|B_\theta(\rho_0)|^{-1/2}\big\|\big\||B_\theta(\rho_0)^*|^{-1/2}W_{0,1}^{(\alpha)}(\theta)P_{n,l}(\theta)\big\|\Big)\\
\le&\frac{C}{\sin^2\vartheta}g^2\rho_0^{1/2}=\frac{C}{\sin^2\vartheta}g^2\alpha
\end{aligned}
$$

using Corollary 1 and Lemma 5 a) and b).
For Rem_2 we use the pull-through formula [4, Lemma IV.8]: We have

$$
\begin{aligned}
\mathrm{Rem}_2=&\alpha\sum_{\mu,\mu'=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\int_{k'\in\mathbb{R}^3}\mathrm{d}k'\,P_{n,l}(\theta)\boldsymbol{\alpha}\cdot G_{\alpha x}^{(\theta)}(k,\mu)a_{\mu'}^*(k')\\
&\times\frac{P_{\mathrm{el},n,l}^{(\alpha)}\otimes\mathbf{1}_{\mathrm{f}}+P_{\mathrm{el},n,l}^{(\alpha)}\otimes\chi_{H_{\mathrm{f}}+|k|+|k'|\ge\rho_0}}{H_{\mathrm{el}}^{(\alpha)}(\theta)+e^{-\theta}(H_{\mathrm{f}}+|k|+|k'|)-z}\boldsymbol{\alpha}\cdot G_{\alpha x}^{(\theta)}(k',\mu')a_\mu(k)P_{n,l}(\theta).
\end{aligned}
$$

Using Lemma 2 (for the resolvent) and [25, Theorem 11] (for the expectation values of the Dirac matrix) we obtain

$$
\begin{aligned}
&\big|\langle\psi,\mathrm{Rem}_2\psi'\rangle\big|\le C\alpha\sum_{\mu,\mu'=1}^{2}\int_{|k|\le\rho_0}\mathrm{d}k\int_{|k'|\le\rho_0}\mathrm{d}k'\,\frac{|\kappa(e^{-\theta}|k|)|}{\sqrt{|k|}}\frac{|\kappa(e^{-\theta}|k'|)|}{\sqrt{|k'|}}\\
&\times\big\|P_{\mathrm{el},n,l}^{(\alpha)}(\theta)\boldsymbol{\alpha}\cdot\epsilon_\mu(k)e^{\mathrm{i}\,\alpha x\cdot k}\Lambda_{\alpha^{-1},3}^{(+)}(\theta)\big\|\big\|\Lambda_{\alpha^{-1},3}^{(+)}(\theta)\boldsymbol{\alpha}\cdot\epsilon_\mu(k')e^{-\mathrm{i}\,\alpha x\cdot k'}P_{\mathrm{el},n,l}^{(\alpha)}(\theta)\big\|
\end{aligned}
$$

$$
\begin{aligned}
&\times \left\| \frac{\underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \mathbf{1}_{\mathrm{f}} + P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \chi_{H_{\mathrm{f}}+|k|+|k'|\geq \rho_0}}{H^{(\alpha)}_{\mathrm{el}}(\theta) + e^{-\theta}(H_{\mathrm{f}} + |k| + |k'|) - z} \right\| \\
&\times \left\| a_\mu(k)\chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\| \left\| a_{\mu'}(k')\chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\| \\
\leq & \frac{Cg^2}{\sin\vartheta\rho_0} \sum_{\mu,\mu'=1}^{2} \int_{|k|\leq\rho_0} \mathrm{d}k \, \frac{|\kappa(e^{-\theta}|k|)|(1+\alpha|k|)}{\sqrt{|k|}\sqrt{|k|}} \sqrt{|k|} \left\| a_\mu(k)\chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\| \\
&\times \int_{|k'|\leq\rho_0} \mathrm{d}k' \, \frac{|\kappa(e^{-\theta}|k'|)|(1+\alpha|k'|)}{\sqrt{|k'|}\sqrt{|k'|}} \sqrt{|k'|} \left\| a_{\mu'}(k')\chi_{H_{\mathrm{f}}\leq\rho_0}\psi' \right\| \\
\leq & \frac{Cg^2}{\sin\vartheta\rho_0} \Big(\int_{|k|\leq\rho_0} \mathrm{d}k \, \frac{1}{|k|^2} \Big) \left\| H_{\mathrm{f}}^{1/2}\chi_{H_{\mathrm{f}}\leq\rho_0}\psi' \right\| \left\| H_{\mathrm{f}}^{1/2}\chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\| \\
\leq & \frac{Cg^2}{\sin\vartheta\rho_0} \rho_0^2 \|\psi'\|\|\psi\| = \frac{C}{\sin\vartheta} g^2\alpha^2 \|\psi'\|\|\psi\|
\end{aligned}
$$

with a generic $C > 0$.

Finally, we consider $\mathrm{Rem}_3 := P_{n,l}(\theta)W^{(\alpha)}(\theta)P_{n,l}(\theta)$, where we show the estimate with $A^{(\theta)}_\kappa(\alpha x)_-$ only. The other estimate works analogously. We find using [25, Lemma 10]

$$
\begin{aligned}
&\sqrt{\alpha} \left| \left\langle \psi', P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \chi_{H_{\mathrm{f}}\leq\rho_0} \boldsymbol{\alpha} \cdot A^{(\theta)}_\kappa(\alpha x)_- P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\rangle \right| \\
\leq & \sqrt{\alpha} \sum_{\mu,\mu'=1}^{2} \int_{|k|\leq\rho_0} \mathrm{d}k \, \frac{|\kappa(e^{-\theta}|k|)|}{\sqrt{|k|}} \\
&\times \left\| P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\boldsymbol{\alpha}\cdot\epsilon_\mu(k) e^{\mathrm{i}\,\alpha x\cdot k} P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \right\| \|\psi'\| \left\| a_\mu(k)\chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\| \\
\leq & Cg \sqrt{\int_{|k|\leq\rho_0} \mathrm{d}k \, \frac{1}{|k|^2}} \|\psi'\| \left\| H_{\mathrm{f}}^{1/2}\chi_{H_{\mathrm{f}}\leq\rho_0}\psi \right\| \leq Cg\rho_0 \|\psi'\|\|\psi\| = C\sqrt{\alpha} g^2 \|\psi'\|\|\psi\|.
\end{aligned}
$$

□

Note that the following Lemma 9 holds only for $z \in \mathcal{A}^<_{n,l}(\alpha,\sigma_0)$, contrary to Lemma 8. It generalizes [5, Lemma 3.16] (see also [19, Lemma A.8]).

Lemma 9. *Let $0 < \vartheta < \theta_0$. Then there is a $C > 0$ such that for all $\alpha > 0$ sufficiently small and all $z \in \mathcal{A}^<_{n,l}(\alpha,\sigma_0)$ the estimate*

$$
\left\| Q^{(\alpha)}_{n,l}(z;\theta) - Z_{n,l}(\alpha;\theta) \right\| \leq \frac{C}{\sin^2\vartheta} g^2\alpha
$$

holds.

Proof. We split $Q_{n,l}^{(\alpha)}(z;\theta) - Z_{n,l}(\alpha;\theta) = \mathrm{Rem}_{4a} + \mathrm{Rem}_{4b}$ with

$$
\begin{aligned}
\mathrm{Rem}_{4a} := \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} & \mathrm{d}k \, P_{n,l}(\theta)[w_{0,1}(k,\mu;\theta) \otimes \mathbf{1}_{\mathrm{f}}] \\
& \times \left[\frac{P_{\mathrm{el},n,l}^{(\alpha)}(\theta) \otimes \chi_{H_{\mathrm{f}}+|k|\geq \rho_0}}{H_{\mathrm{el}}^{(\alpha)}(\theta) + e^{-\theta}(H_{\mathrm{f}} + |k|) - z} \right] [w_{1,0}(k,\mu;\theta) \otimes \mathbf{1}_{\mathrm{f}}] P_{n,l}(\theta) \\
- \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} & \frac{\mathrm{d}k}{e^{-\theta}|k|} P_{\mathrm{el},n,l}^{(\alpha)}(\theta) w_{0,1}(k,\mu;\theta) P_{\mathrm{el},n,l}^{(\alpha)}(\theta) w_{1,0}(k,\mu;\theta) P_{\mathrm{el},n,l}^{(\alpha)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Rem}_{4b} := \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} & \mathrm{d}k \, P_{n,l}(\theta)[w_{0,1}(k,\mu;\theta) \otimes \mathbf{1}_{\mathrm{f}}] \\
& \times \left[\frac{\underline{P}_{\mathrm{el},n,l}^{(\alpha)}(\theta) \otimes \mathbf{1}_{\mathrm{f}}}{H_{\mathrm{el}}^{(\alpha)}(\theta) + e^{-\theta}(H_{\mathrm{f}} + |k|) - z} \right] [w_{1,0}(k,\mu;\theta) \otimes \mathbf{1}_{\mathrm{f}}] P_{n,l}(\theta) \\
- \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} & \mathrm{d}k \, P_{\mathrm{el},n,l}^{(\alpha)} \otimes \chi_{H_{\mathrm{f}}\leq \rho_0} w_{0,1}(k,\mu;\theta) \\
& \times \left[\frac{\underline{P}_{\mathrm{el},n,l}^{(\alpha)}(\theta)}{H_{\mathrm{el}}^{(\alpha)}(\theta) - E_{n,l} + e^{-\theta}|k|} \right] w_{1,0}(k,\mu;\theta) P_{\mathrm{el},n,l}^{(\alpha)}(\theta) \otimes \chi_{H_{\mathrm{f}}\leq \rho_0}.
\end{aligned}
$$

We start with Rem_{4a}: As in the proof of Lemma 2 a) one shows for $\rho_0 \leq r + |k|$ the inequalities

$$
|E_{n,l}(\alpha) + e^{-\theta}(r + |k|) - z| \geq -\sigma_0 + \sin\vartheta (r + |k|) \geq \frac{|k| \sin\vartheta}{2} \tag{52}
$$

and

$$
|E_{n,l}(\alpha) + e^{-\theta}(r + |k|) - z| \geq -\sigma_0 + \sin\vartheta (r + |k|) \geq \frac{\rho_0 \sin\vartheta}{2}, \tag{53}
$$

since we have $\sigma_0 \leq \frac{\rho_0 \sin\vartheta}{2} \leq \frac{(r+|k|)\sin\vartheta}{2}$ for sufficiently small $\alpha > 0$.
As in the proof of Lemma 4 one obtains using [25, Lemma 10] the inequality

$$
\| P_{\mathrm{el},n,l}^{(\alpha)}(\theta) w_{0,1}(k,\mu;\theta) P_{\mathrm{el},n,l}^{(\alpha)}(\theta) \| \leq C g \frac{|\kappa(e^{-\theta}|k|)|}{\sqrt{|k|}}. \tag{54}
$$

We find after a little transformation of Rem_{4a}

$$
\begin{aligned}
\| \mathrm{Rem}_{4a} \| = \Big\| \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} & \mathrm{d}k \, P_{n,l}(\theta)[w_{0,1}(k,\mu;\theta) \otimes \mathbf{1}_{\mathrm{f}}] P_{\mathrm{el},n,l}^{(\alpha)}(\theta) \\
\times & \left[\frac{\big(e^{-\theta} H_{\mathrm{f}} + E_{n,l}(\alpha) - z\big) \chi_{H_{\mathrm{f}}+|k|\geq\rho_0} \chi_{H_{\mathrm{f}}\leq\rho_0}}{(E_{n,l}(\alpha) + e^{-\theta}(H_{\mathrm{f}} + |k|) - z)\, e^{-\theta}|k|} \right]
\end{aligned}
$$

$$
\begin{aligned}
&\times P^{(\alpha)}_{\mathrm{el},n,l}(\theta)[w_{1,0}(k,\mu;\theta)\otimes \mathbf{1}_{\mathrm{f}}]P_{n,l}(\theta)\\
&-\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,P^{(\alpha)}_{\mathrm{el},n,l}(\theta)w_{0,1}(k,\mu;\theta)P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\\
&\times\frac{\chi_{H_{\mathrm{f}}\le\rho_0}\chi_{H_{\mathrm{f}}+|k|\le\rho_0}}{e^{-\theta}|k|}w_{1,0}(k,\mu;\theta)P^{(\alpha)}_{\mathrm{el},n,l}\Big\|\\
\le&\frac{C}{\sin\vartheta}g^2(\alpha^2+\rho_0)\Big(\frac{1}{\rho_0}\int_{|k|\le\rho_0}\mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2}{|k|^2}+\int_{|k|\ge\rho_0}\mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2}{|k|^3}\Big)\\
&+g^2\int_{|k|\le\rho_0}\mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2}{|k|^2}\\
\le&\frac{C}{\sin\vartheta}g^2(\alpha^2\frac{\rho_0}{\rho_0}+\alpha^2\ln\rho_0^{-1}+\rho_0)\le\frac{C}{\sin\vartheta}g^2\alpha.
\end{aligned}
$$

Here, we split the integration in the first summand in the regions $|k|\le\rho_0$ and $|k|>\rho_0$. We use inequality (53) in the first region, and inequality (52) in the second region.

The estimate on Rem_{4b} is more difficult. We split the projection $\underline{P}^{(\alpha)}_{\mathrm{el},n,l}=\overline{P}^{(\alpha)}_{\mathrm{el},n}+\overline{P}^{(\alpha)}_{\mathrm{el},n,l}$ and obtain for $P=\overline{P}^{(\alpha)}_{\mathrm{el},n}$ as well as for $P=\overline{P}^{(\alpha)}_{\mathrm{el},n,l}$

$$
\begin{aligned}
&\Big\|\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_{\mathrm{f}}\le\rho_0}[w_{0,1}(k,\mu;\theta)\otimes\mathbf{1}_{\mathrm{f}}]\\
&\times\left[\frac{P\otimes\mathbf{1}_{\mathrm{f}}}{H^{(\alpha)}_{\mathrm{el}}(\theta)+e^{-\theta}(H_{\mathrm{f}}+|k|)-z}\right][w_{1,0}(k,\mu;\theta)\otimes\mathbf{1}_{\mathrm{f}}]P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_{\mathrm{f}}\le\rho_0}\\
&-\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_{\mathrm{f}}\le\rho_0}w_{0,1}(k,\mu;\theta)\\
&\times\left[\frac{P\otimes\mathbf{1}_{\mathrm{f}}}{H^{(\alpha)}_{\mathrm{el}}(\theta)-E_{n,l}+e^{-\theta}|k|}\right]w_{1,0}(k,\mu;\theta)P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes\chi_{H_{\mathrm{f}}\le\rho_0}\Big\|\\
\le&\alpha\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2}{|k|}\\
&\times\big\|P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\boldsymbol{\alpha}\cdot\epsilon_\mu(k)e^{\mathrm{i}\,\alpha x\cdot k}\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)\big\|\big\|\Lambda^{(+)}_{\alpha^{-1},\mathfrak{Z}}(\theta)\boldsymbol{\alpha}\cdot\epsilon_\mu(k')e^{-\mathrm{i}\,\alpha x\cdot k'}P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\big\|\\
&\times\Big[\Big\|\frac{P\otimes\mathbf{1}_{\mathrm{f}}}{H^{(\alpha)}_{\mathrm{el}}(\theta)+e^{-\theta}(H_{\mathrm{f}}+|k|)-z}\Big\|\Big\|\frac{P}{H^{(\alpha)}_{\mathrm{el}}(\theta)-E_{n,l}+e^{-\theta}|k|}\Big\|\\
&\times\big(|E_{n,l}-z|+\|H_{\mathrm{f}}\chi_{H_{\mathrm{f}}\le\rho_0}\|\big)\Big]\\
\le&Cg^2\alpha^2\int_{k\in\mathbb{R}^3}\mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|}\\
&\times\Big\|\frac{P\otimes\mathbf{1}_{\mathrm{f}}}{H^{(\alpha)}_{\mathrm{el}}(\theta)+e^{-\theta}(H_{\mathrm{f}}+|k|)-z}\Big\|\Big\|\frac{P}{H^{(\alpha)}_{\mathrm{el}}(\theta)-E_{n,l}+e^{-\theta}|k|}\Big\|.
\end{aligned}
$$

We used [25, Theorem 11]. Note that all estimates on $\| \frac{P\otimes \mathbf{1}_{\mathrm{f}}}{H_{\mathrm{el}}^{(\alpha)}(\theta)+e^{-\theta}(H_{\mathrm{f}}+|k|)-z}\|$ in Lemma 1 hold also for $\|[PH_{\mathrm{el}}^{(\alpha)}(\theta)-E_{n,l}+e^{-\theta}|k|]^{-1}P\|$, since the operator under the norm in the second expression is the projection of the operator in the first expression on the vacuum sector with $z=E_{n,l}$.

Case 1: $P=\overline{P}_{\mathrm{el},n,l}^{(\alpha)}$. We split the integration in the regions $B_1 := \{k\in\mathbb{R}^3 | |k|\le \rho_0\}$ and $B_2 := \{k\in\mathbb{R}^3 | |k|>\rho_0\}$. Using Formula (27) in Lemma 1 a), the integral over B_1 can be estimated by

$$\frac{C}{\sin^2\vartheta} g^2\alpha^2 \frac{1}{\delta_{n,l}(\alpha)^2}\int_{k\in B_1} \mathrm{d}k\,\frac{1}{|k|} \le \frac{C}{\sin^2\vartheta}g^2\alpha^{-2}\rho_0^2 = \frac{C}{\sin^2\vartheta}g^2\alpha^2.$$

With Formula (28) in Lemma 1 b) we estimate the integral over B_2 by

$$\frac{C}{\sin^2\vartheta} g^2\alpha^2 \int_{k\in B_2} \mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|^3} \le \frac{C}{\sin^2\vartheta}g^2\alpha^2\ln\rho_0^{-1}.$$

Case 2: $P=\overline{P}_{\mathrm{el},n}^{(\alpha)}$. We estimate the resolvents with Lemma 1 c) and obtain the estimate

$$\frac{C}{\delta_n^2\sin^2\vartheta} g^2\alpha^2 \int_{k\in \mathbb{R}^3} \mathrm{d}k\,\frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|} \le \frac{C}{\delta_n^2\sin^2\vartheta}g^2\alpha^2 = \frac{C}{\delta_n^2\sin^2\vartheta}g^2\alpha^2.$$

□

The following Lemma generalizes [19, Corollary A.9]. Note, however, that we do not remove the α-dependence of the real part.

Lemma 10. *There is a constant $C>0$ such that for all sufficiently small $\alpha>0$ the estimate*

$$\|\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1}Z_{n,l,\pm}(\alpha)\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)-Y_{n,l,\pm}(\alpha)\| \le Cg^2\alpha$$

holds.

Proof. We consider the case with the minus sign only. It suffices to show

$$\|\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1}\mathrm{Im}\,Z_{n,l,-}(\alpha)\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)-Z_{n,l,im}\| \le Cg^2\alpha.$$

Because of $[x,H_{\mathrm{el}}^{(\alpha)}]=\mathrm{i}\,\alpha^{-1}\boldsymbol{\alpha}$ and $|e^{\mathrm{i}\,\alpha k\cdot x}-1|\le\alpha|k||x|$ we obtain from [25, Lemma 10 and Lemma 12]

$$\begin{aligned}\|\mathrm{Im}\,Z_{n,l,-}(\alpha)-g^2\pi \sum_{\substack{n',l':\\ E_{n',l'}(\alpha)<E_{n,l}(\alpha)}}\sum_{\mu=1,2}\int_{|\omega|=1}\mathrm{d}\omega\,(E_{n',l'}(\alpha)-E_{n,l}(\alpha))\\ \times\frac{\kappa(|E_{n',l'}(\alpha)-E_{n,l}(\alpha)|)^2}{4\pi^2}P_{\mathrm{el},n,l}^{(\alpha)}\epsilon_\mu(\omega)\cdot xP_{\mathrm{el},n',l'}^{(\alpha)}\epsilon_\mu(\omega)\cdot xP_{\mathrm{el},n,l}^{(\alpha)}\| \le g^2\alpha.\end{aligned}$$

The integral over ω and the sum over the polarizations can be done in the same way as in the non-relativistic case (see [19, Remark 1]). If we take additionally into account that $|E_{n,l'}(\alpha) - E_{n,l}(\alpha)| \le C\alpha^2$, we obtain

$$\begin{aligned}\|\operatorname{Im} Z_{n,l,-}(\alpha) - g^2\frac{2}{3}\sum_{n',l':n'<n}&(E_{n',l'}(\alpha) - E_{n,l}(\alpha))\\ &\times\frac{\kappa(|E_{n',l'}(\alpha) - E_{n,l}(\alpha)|)^2}{4\pi^2}P^{(\alpha)}_{\mathrm{el},n,l}xP^{(\alpha)}_{\mathrm{el},n',l'}xP^{(\alpha)}_{\mathrm{el},n,l}\| \le g^2\alpha.\end{aligned}$$

[25, Lemma 8] implies $\mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0)^{-1}P^{(\alpha)}_{\mathrm{el},n,l} = P^{(0)}_{\mathrm{el},n,l}\mathcal{U}_{\mathrm{NR}}(\alpha^{-1}, \mathfrak{z}; 0)$. The claim follows together with [25, Lemma 7], [25, Equation (76) in Lemma 8] and [25, Lemma 11]. Note that κ admits an analytic continuation. □

6 Estimates on the Numerical Range

The estimates in Section 5 allow us to control the numerical range of the Feshbach operator. But since $\operatorname{Re} Z_{n,l,\pm}(\alpha)$ depends on α, we have to prove that $Z_{n,l,\pm}(\alpha)$ is of order g^2:

Lemma 11. *Let $0 < \vartheta < \theta_0$ and $n > 2$. Then the following holds:*

a) There is a $C > 0$ such that for all sufficiently small $\alpha > 0$ the estimate

$$\|Z_{n,l,\pm}(\alpha)\| \le Cg^2$$

holds.

b) There is a $c > 0$ such that for all sufficiently small $\alpha > 0$ the estimates

$$\begin{aligned}\operatorname{Im} Z_{n,l,-}(\alpha) &\ge cg^2 + \mathcal{O}(g^2\alpha)\\ \operatorname{Im} Z_{n,l,+}(\alpha) &\le -cg^2 + \mathcal{O}(g^2\alpha)\end{aligned}$$

hold.

Proof.
b) follows immediately from Lemma 10, since by [19, Theorem B.1] there is a $c > 0$ such that the estimates $\operatorname{Im} Y_{n,l,-}(\alpha) \ge cg^2$ and $\operatorname{Im} Y_{n,l,+}(\alpha) \le -cg^2$ hold (cf. the Definition (22) of $\operatorname{Im} Y_{n,l,\pm}(\alpha)$ as well as the remark before Theorem 1).
a) As in the estimates on Rem_{4a} in the proof of Lemma 9 we find

$$\|\sum_{\mu=1,2}\int_{k\in\mathbb{R}^3}\frac{\mathrm{d}k}{e^{-\theta}|k|}P^{(\alpha)}_{\mathrm{el},n,l}(\theta)w_{0,1}(k,\mu;\theta)P^{(\alpha)}_{\mathrm{el},n,l}(\theta)w_{1,0}(k,\mu;\theta)P^{(\alpha)}_{\mathrm{el},n,l}\| \le Cg^2.$$

Moreover, we obtain

$$\begin{aligned}&\| \sum_{\mu=1,2} \int_{k\in\mathbb{R}^3} \mathrm{d}k\, P^{(\alpha)}_{\mathrm{el},n,l}(\theta) \otimes \chi_{H_\mathrm{f}\le\rho_0} w_{0,1}(k,\mu;\theta)\\ &\times \underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta)[\underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) H^{(\alpha)}_{\mathrm{el}}(\theta) - E_{n,l} + e^{-\theta}|k|]^{-1}\\ &\times \underline{P}^{(\alpha)}_{\mathrm{el},n,l}(\theta) w_{1,0}(k,\mu;\theta) P^{(\alpha)}_{\mathrm{el},n,l}(\theta)\otimes \chi_{H_\mathrm{f}\le\rho_0}\| \le Cg^2.\end{aligned}$$

To see this, we proceed as in the estimate on Rem_{4b} in the proof of Lemma 9: In Case 1 we can estimate the integral over B_1 by

$$\frac{C}{\sin\vartheta} g^2 \frac{1}{\delta_{n,l}(\alpha)} \int_{k\in B_1} \mathrm{d}k\, \frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|} \le \frac{C}{\sin\vartheta} g^2\alpha^{-2}\rho_0^2 = \frac{C}{\sin^2\vartheta} g^2\alpha^2$$

and the integral over B_2 by

$$\frac{C}{\sin\vartheta} g^2 \int_{k\in B_2} \mathrm{d}k\, \frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|^2} \le \frac{C}{\sin\vartheta} g^2.$$

In Case 2 we obtain the estimate

$$\frac{C}{\delta_n \sin\vartheta} g^2 \int_{k\in \mathbb{R}^3} \mathrm{d}k\, \frac{|\kappa(e^{-\theta}|k|)|^2(1+\alpha|k|)^2}{|k|} \le \frac{C}{\delta_n\sin\vartheta} g^2.$$

[25, Lemma 9] yields the claim. □

This lemma implies in particular that the numerical range of $Z_{n,l,\pm}(\alpha)$ is contained in a ball around 0 with radius $\mathcal{O}(g^2)$. In particular, this holds for the real part $\mathrm{Re}\, Z_{n,l,\pm}(\alpha) = \mathrm{Re}\, Y_{n,l,\pm}(\alpha)$. As in [19], there are constants $a, b > 0$ such that $\mathrm{NumRan}\, Y_{n,l,\pm}(\alpha) \subset g^2 A(c,a,b)$ with $A(c,a,b) := ic + ([-a,a] + \mathrm{i}\,[0,b])$. As in the non-relativistic case, we set $\nu := \min\{\vartheta, \arctan(c/(2a))\}$. Since we are interested only in $n \le \tilde{n}$, we can choose the set $A(c,a,b)$ and the angle ν independent of n and l.
Thus, we can control the inverse of the Feshbach operator $\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - z)$ for $z \in \mathcal{A}^<_{n,l}(\alpha,\sigma_0)$ analogously to the non-relativsitic case (see [19, Lemma 6]) as follows (see Figure 1):

Lemma 12. *Let $0 < \vartheta < \theta_0$ and $0 < g \ll \vartheta$ small enough. Then the following estimates hold:*

*a) There are constants $C_1, C_2 > 0$ such that $\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - z)$ has a bounded inverse for all $z \in \mathcal{A}^<_{n,l}(\alpha,\sigma_0) \setminus D(\mathrm{NumRan}(E_{n,l}(\alpha) - Y_{n,l}(\alpha) \otimes \mathbf{1}_\mathrm{f} + e^{-\theta}\mathbf{1}_\mathrm{el} \otimes H_\mathrm{f})|_{\mathrm{Ran}\, P^{(0)}_{\mathrm{el},n,l}}, C_1 \cdot g^2\sqrt{\alpha})$, and for $\lambda \in [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha), E_{n,l}(\alpha)$
$+ \delta_{n,l,+}(\alpha)]$ the esimate*

$$\left\|\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - \lambda)^{-1}\right\| \le \frac{C_2}{\sin\nu\sqrt{(E_{n,l}(\alpha)-\lambda)^2 + cg^4}} \tag{55}$$

holds.

b) *There are constants* $C_1, C_2 > 0$ *such that for all* $z \in \mathbb{C} \setminus D(\mathrm{NumRan}(E_{n,l}(\alpha) - Y_{n,l}(\alpha))|_{\mathrm{Ran}\, P^{(0)}_{\mathrm{el},n,l}(0)}, C_1 \cdot g^2\alpha)$ *the operator* $(E_{n,l}(\alpha) - z - Z_{n,l}(\alpha;\theta))|_{\mathrm{Ran}\, P^{(\alpha)}_{\mathrm{el},n,l}(\theta)}$ *defined on* $\mathrm{Ran}\, P^{(\alpha)}_{\mathrm{el},n,l}(\theta)$ *has a bounded inverse which fulfills the estimate*

$$\begin{aligned} &\|[(E_{n,l}(\alpha) - z - Z_{n,l}(\alpha;\theta))|_{\mathrm{Ran}\, P^{(\alpha)}_{\mathrm{el},n,l}(\theta)}]^{-1}\| \\ &\qquad \leq \frac{C}{\mathrm{dist}(z, \mathrm{NumRan}(E_{n,l}(\alpha) - Y_{n,l}(\alpha))|_{\mathrm{Ran}\, P^{(\alpha)}_{\mathrm{el},n,l}(0)})}, \end{aligned} \tag{56}$$

and in particular (55).

Proof. This can be shown using Lemmas 8, 9 and 10 exactly as in the proof of [19, Lemma 6]. □

For $l = 1$ or $l = n$, the set $\mathcal{A}^{<}_{n,l}(\alpha, \sigma_0)$ is strictly interior of the set $\mathcal{A}_{n,l}(\alpha, \sigma_0)$, such that we need a relativistic analogon of [19, Lemma 7] in this case.

Lemma 13. *Let* $0 < \vartheta < \theta_0$ *and* $0 < g \ll \vartheta$ *small enough. Let moreover* $l = 1$ *or* $l = n$. *Then the following statements hold: The Feshbach operator* $\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - z)$ *is bounded invertible for all* $z \in \mathcal{A}_{n,l}(\alpha, \sigma_0) \setminus \mathcal{A}^{<}_{n,l}(\alpha, \sigma_0)$ *and there is a* $C > 0$ *such that for* $\lambda \in [E_n - \delta_{n,-}, E_{n,1}(\alpha) - \delta_{n,1,-}(\alpha)]$ *respectively* $\lambda \in [E_{n,n}(\alpha) + \delta_{n,n,+}(\alpha), E_n + \delta_{n,+}]$ *the estimate*

$$\|\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - \lambda)^{-1}\| \leq \frac{C}{\sin\vartheta |\lambda - E_{n,l}(\alpha)| - Cg^2}$$

holds with $l = 1$ *or* $l = n$, *respectively. The same estimate holds for* $[E_{n,l}(\alpha) - \lambda - Q^{(\alpha)}_{n,l}(\lambda;\theta)]^{-1}$.

Proof. This follows analogously to the non-relativistic case (see the proof of [19, Lemma 7]) from Lemma 7 b). For the claim on $[E_{n,l}(\alpha) - \lambda - Q^{(\alpha)}_{n,l}(\lambda;\theta)]^{-1}$, note additionally Lemma 8 and the proof thereof. □

Corollary 3. *Let* $0 < \vartheta < \theta_0$ *and* $0 < g \ll \vartheta$ *small enough. The for all* $1 \leq l \leq n$ *the following holds:*

$$\begin{aligned} &\sigma(H_\alpha(\theta)) \cap \mathcal{A}_{n,l}(\alpha, \sigma_0) \\ &\subset D(\mathrm{NumRan}(E_{n,l}(\alpha) - Y_{n,l}(\alpha) \otimes \mathbf{1}_\mathrm{f} + e^{-\theta}\mathbf{1}_\mathrm{el} \otimes H_\mathrm{f})|_{\mathrm{Ran}\, P^{(0)}_{\mathrm{el},n,l}}, C_1 \cdot g^2\sqrt{\alpha}), \end{aligned}$$

where C_1 *was defined in Lemma 12. In particular,* $[E_n - \delta_{n,-}, E_n + \delta_{n,+}] \subset \rho(H_\alpha(\theta))$.

Proof. This follows because of Lemma 7 c) immediately from Lemma 12 and Lemma 13. □

REMARK 3. *The estimates above hold as in the non-relativistic case (cf. [19, Remark 5]) also for* $-\theta_0 < \vartheta < 0$, *if one reflects the sets* $\mathcal{A}_{n,l}(\alpha, \sigma_0)$ *and* $\mathcal{A}^{<}_{n,l}(\alpha, \sigma_0)$ *about the real axis and replaces* $Y_{n,l}(\alpha) = Y_{n,l,-}(\alpha)$ *by* $Y_{n,l,+}(\alpha)$ *for the localization of the numerical range.*

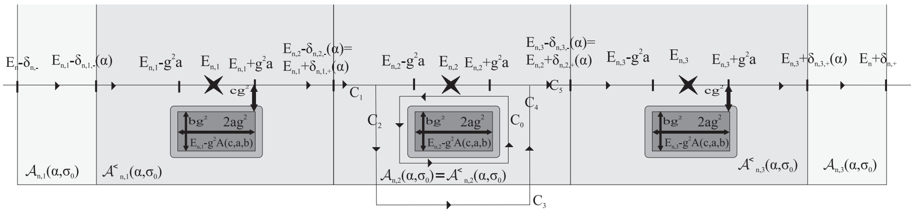


Figure 1: The integration contour in the relativistic model for the principal quantum number $n = 3$.

7 LIFETIME OF EXCITED STATES

We are now able to prove Theorem 1 similarly as in the non-relativistic case. The fine structure splitting induces some differences, however: Since a spectral cutoff around the fine structure component considered would converge to zero as α^2, we introduce a spectral cutoff around all the fine structure components of the corresponding principal quantum number so that additionally contributions of the other components have to be estimated.

Proof of Theorem 1.
Step 1: We pick a function $\tilde{F} \in C_0^\infty(\mathbb{R})$ with $\tilde{F}(x) = 0$ for $|x| \geq 1$ and $\tilde{F}(x) = 1$ for $|x| \leq 1/2$ and define a cutoff function $F(x) := \tilde{F}(\delta_n^{-1}(x - E_n))$. As in the non-relativistic case (see step 1 in the proof of [19, Theorem 1]) one shows $|\langle\psi(\alpha), e^{-isH_\alpha} F(H_\alpha)\psi(\alpha)\rangle - \langle\psi(\alpha), e^{-isH_\alpha}\psi(\alpha)\rangle| \leq C\sqrt{\alpha}$ uniformly in $s \geq 0$.
Step 2: We write

$$
\begin{aligned}
&\langle\psi(\alpha), e^{-isH_\alpha} F(H_\alpha)\psi(\alpha)\rangle \\
&\quad = -\frac{1}{2\pi \mathrm{i}} \lim_{\epsilon\downarrow 0} \int \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s} F(\lambda)[f(0, \lambda - \mathrm{i}\,\epsilon) - f(0, \lambda + \mathrm{i}\,\epsilon)] \\
&\quad = -\frac{1}{2\pi \mathrm{i}} \int \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s} F(\lambda)[f(\overline{\theta}, \lambda) - f(\theta, \lambda)],
\end{aligned}
$$

where $f(\theta, \lambda) := \langle\psi(\alpha; \overline{\theta}), \frac{1}{H_\alpha(\theta) - \lambda}\psi(\alpha; \theta)\rangle$ with $\psi(\alpha; \theta) := \phi(\alpha; \theta) \otimes \Omega$ and $\phi(\alpha; \theta) := \mathcal{U}_{\mathrm{el}}(\theta)\phi(\alpha)$. (We choose $\operatorname{Im}\theta > 0$.) In the first step, we used [36, Theorem VII.13]. In the second step, we used the dilation analyticity of $H_\alpha(\theta)$ (see Theorem 3) and the fact that $H_\alpha(\theta)$ has no spectrum in the interval

$[E_n - \delta_{n,-}, E_n + \delta_{n,+}/2]$ (see Corollary 3). We split the integration into several intervals:

$$
\begin{aligned}
&-\frac{1}{2\pi \mathrm{i}} \int \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s} F(\lambda)[f(\overline{\theta},\lambda) - f(\theta,\lambda)] \\
&= -\frac{1}{2\pi \mathrm{i}} \Bigg\{ \sum_{l'=1}^{n} \int_{E_{n,l'}(\alpha)-\delta_{n,l',-}(\alpha)}^{E_{n,l'}(\alpha)+\delta_{n,l',+}(\alpha)} \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s}[f(\overline{\theta},\lambda) - f(\theta,\lambda)] \\
&\quad + \int_{E_n-\delta_{n,-}}^{E_{n,1}(\alpha)-\delta_{n,1,-}(\alpha)} \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s} F(\lambda)[f(\overline{\theta},\lambda) - f(\theta,\lambda)] \\
&\quad + \int_{E_{n,n}(\alpha)+\delta_{n,n,+}(\alpha)}^{E_n+\delta_{n,+}} \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s} F(\lambda)[f(\overline{\theta},\lambda) - f(\theta,\lambda)] \Bigg\}
\end{aligned}
$$

We used here $F(\lambda) = 1$ for $\lambda \in [E_{n,1}(\alpha) - \delta_{n,1,-}(\alpha), E_{n,n}(\alpha) + \delta_{n,n,+}(\alpha)] \subset [E_n - \delta_n/2, E_n + \delta_n/2]$.

Step 3: For $\lambda \in [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha), E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)]$ we observe that Equation (51) in Lemma 7 implies

$$
\langle \psi(\alpha;\overline{\theta}), \frac{1}{H_g(\theta) - \lambda} \psi(\alpha;\theta) \rangle = \langle \psi(\alpha;\overline{\theta}), \mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - \lambda)^{-1} \psi(\alpha;\theta) \rangle
$$

and find

$$
\begin{aligned}
f(\theta,\lambda) &= \langle \psi(\alpha;\overline{\theta}), \mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - \lambda)^{-1} \psi(\alpha;\theta) \rangle \\
&= \langle \phi(\alpha;\overline{\theta}), [E_{n,l}(\alpha) - \lambda - Z_{n,l}(\alpha;\theta)]^{-1} \phi(\alpha;\theta) \rangle \\
&\quad - \langle \psi(\alpha;\overline{\theta}), [E_{n,l}(\alpha) - \lambda - Z_{n,l}(\alpha;\theta)]^{-1} \\
&\quad \times [\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - \lambda) - (E_{n,l}(\alpha) - \lambda + e^{-\theta}\mathbf{1}_{\mathrm{el}} \otimes H_{\mathrm{f}} - Z_{n,l}(\alpha;\theta))P_{n,l}(\theta)] \\
&\quad \times \mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta) - \lambda)^{-1} \psi(\alpha;\theta) \rangle =: \hat{f}(\theta,\lambda) + B_1(\theta,\lambda)
\end{aligned}
$$

using the second resolvent equation. Here $\hat{f}(\theta,\lambda)$ is the first term in the sum. Using the dilation analyticity and the resolvent identity once again, we obtain

$$
\begin{aligned}
\hat{f}(\theta,\lambda) &= \langle \phi(\alpha), [E_{n,l}(\alpha) - \lambda - Z_{n,l,-}(\alpha)]^{-1} \phi(\alpha) \rangle \\
&= \langle \phi(0), [E_{n,l}(\alpha) - \lambda - \mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1} Z_{n,l,-}(\alpha) \mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)]^{-1} \phi(0) \rangle \\
&= \langle \phi(0), [E_{n,l}(\alpha) - \lambda - Y_{n,l,-}(\alpha)]^{-1} \phi(0) \rangle \\
&\quad - \langle \phi(0), [E_{n,l}(\alpha) - \lambda - Y_{n,l,-}(\alpha)]^{-1} \\
&\quad \times [\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1} Z_{n,l,-}(\alpha) \mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0) - Y_{n,l,-}(\alpha)] \\
&\quad \times [E_{n,l}(\alpha) - \lambda - \mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1} Z_{n,l,-}(\alpha) \mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)]^{-1} \phi(0) \rangle \\
&=: \tilde{f}_-(\lambda) + B_{2,-}(\lambda),
\end{aligned}
$$

where $\tilde{f}_-(\lambda)$ is the first term in the sum. We set $B(\theta,\lambda) := B_1(\theta,\lambda) + B_{2,-}(\lambda)$.

Accordingly we obtain

$$\begin{aligned}
&\hat{f}(\bar{\theta},\lambda) = \langle \phi(\alpha), [E_{n,l}(\alpha) - \lambda - Z_{n,l,+}(\alpha)]^{-1}\phi(\alpha)\rangle \\
=&\langle \phi(0), [E_{n,l}(\alpha) - \lambda - Y_{n,l,+}(\alpha)]^{-1}\phi(0)\rangle \\
&- \langle \phi(0), [E_{n,l}(\alpha) - \lambda - Y_{n,l,+}(\alpha)]^{-1} \\
&\times [\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1} Z_{n,l,+}(\alpha)\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0) - Y_{+}(\alpha)] \\
&\times [E_{n,l}(\alpha) - \lambda - \mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)^{-1} Z_{n,l,+}(\alpha)\mathcal{U}_{\mathrm{NR}}(\alpha^{-1},\mathfrak{z};0)]^{-1}\phi(0)\rangle \\
=:& \tilde{f}_{+}(\lambda) + B_{2,+}(\lambda),
\end{aligned}$$

where $\tilde{f}_{+}(\lambda)$ is the first term in the sum. We set $B(\bar{\theta},\lambda) := B_1(\bar{\theta},\lambda) + B_{2,+}(\lambda)$. As in the non-relativistic case, we move the contour for $\tilde{f}_{\pm}(\lambda)$ and estimate the terms $B(\theta,\lambda)$ and $B(\bar{\theta},\lambda)$ on the real axis. We find

$$\begin{aligned}
&\int_{E_{n,l}(\alpha)-\delta_{n,l,-}(\alpha)}^{E_{n,l}(\alpha)+\delta_{n,l,+}(\alpha)} \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s}[f(\bar{\theta},\lambda) - f(\theta,\lambda)] \\
&= \int_{E_{n,l}(\alpha)-\delta_{n,l,-}(\alpha)}^{E_{n,l}(\alpha)+\delta_{n,l,+}(\alpha)} \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s}[B(\overline{\theta},\lambda) - B(\theta,\lambda)] \\
&+ \int_{C_1+C_5} \mathrm{d}z\, e^{-\mathrm{i}\,zs}[\tilde{f}_{+}(z) - \tilde{f}_{-}(z)] \\
&+ \int_{C_2+C_3+C_4} \mathrm{d}z\, e^{-\mathrm{i}\,zs}[\tilde{f}_{+}(z) - \tilde{f}_{-}(z)] - \int_{C_0} \mathrm{d}z\, e^{-\mathrm{i}\,zs}[\tilde{f}_{+}(z) - \tilde{f}_{-}(z)],
\end{aligned}$$

where $C := C_1 + C_2 + C_3 + C_4 + C_5$ with $C_1 := [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha), E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha)/2]$, $C_2 := [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha)/2, E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha)/2 - \mathrm{i}\,\delta_{n,l}(\alpha)]$, $C_3 := [E_{n,l}(\alpha) - \delta_{n,l,-}(\alpha)/2 - \mathrm{i}\,\delta_{n,l}(\alpha), E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)/2 - \mathrm{i}\,\delta_{n,l}(\alpha)]$, $C_4 := [E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)/2 - \mathrm{i}\,\delta_{n,l}(\alpha), E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)/2]$ and $C_5 := [E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)/2, E_{n,l}(\alpha) + \delta_{n,l,+}(\alpha)]$. Note that this contour lies partially *outside* $\mathcal{A}_{n,l}(\alpha,\sigma_0)$, which is possible since we do not consider any integrals which contain $Q_{n,l}^{(\alpha)}(z;\theta)$. C_0 is a suitable contour to pick a pole contribution of $\tilde{f}(\theta,z)$. We choose as in the non-relativistic case $C_0 = [E_{n,l}(\alpha) + g^2(-(a+c/2) - ic/2), E_{n,l}(\alpha) + g^2((a+c/2) - ic/2)] + [E_{n,l}(\alpha) + g^2((a+c/2) - ic/2), E_{n,l}(\alpha) + g^2((a+c/2) - \mathrm{i}\,(b+3c/2))] + [E_{n,l}(\alpha) + g^2((a+c/2) - \mathrm{i}\,(b+3c/2)), E_{n,l}(\alpha) + g^2(-(a+c/2) - \mathrm{i}\,(b+3c/2))] + [E_{n,l}(\alpha) + g^2(-(a+c/2) - \mathrm{i}\,(b+3c/2)), E_{n,l}(\alpha) + g^2(-(a+c/2) - \mathrm{i}\,(c/2))]$.

Estimates on the real axis: We show the estimate on $B_1(\theta,\lambda)$. Using Lemma 8, Lemma 9 and Lemma 12 we obtain $|B_1(\theta,\lambda)| \leq C\nu^{-2} \cdot \frac{g^2\sqrt{\alpha}}{(E_{n,l}(\alpha)-\lambda)^2 + c^2g^4}$. It is easy to see that $\int \mathrm{d}\lambda \frac{g^2\sqrt{\alpha}}{(E_{n,l}(\alpha)-\lambda)^2+c^2g^4}$ is $\mathcal{O}(\sqrt{\alpha})$. The same estimates hold for $B_1(\bar{\theta},\lambda)$. The estimates on $B_{2,\pm}(\lambda)$ work analogously using Lemma 10 and Lemma 12.

Estimates on the contour C: We estimate the integral $\int_C |e^{-isz}||\tilde{f}_{+}(z) -$

$\tilde{f}_-(z)||\mathrm{d}z|$: Note that

$$\begin{aligned}\tilde{f}_-(z) =& \frac{1}{E_{n,l}(\alpha)-z}\langle\phi(0),\phi(0)\rangle \\ &+\langle\phi(0),\frac{1}{E_{n,l}(\alpha)-z}Y_{n,l,-}(\alpha)\frac{1}{E_{n,l}(\alpha)-z-Y_{n,l,-}(\alpha)}\phi(0)\rangle\end{aligned}$$

holds. Accordingly, the leading terms of $\tilde{f}_-(z)$ and $\tilde{f}_+(z)$ cancel, and it suffices to show that the remaining terms are at least of order $\sqrt{\alpha}$. It follows from Equation (22) and Lemma 11 that $\|Y_{n,l,\pm}(\alpha)\|\le Cg^2$. Thus we can estimate

$$\begin{aligned}&|\langle\phi(0),\frac{1}{E_{n,l}(\alpha)-(\lambda-\mathrm{i}\,\delta_{n,l}(\alpha))}Y_{n,l,-}(\alpha) \\ &\times\frac{1}{E_{n,l}(\alpha)-(\lambda-\mathrm{i}\,\delta_{n,l}(\alpha))-Y_{n,l,-}(\alpha)}\phi(0)\rangle|\le C\cdot\frac{g^2}{(E_{n,l}(\alpha)-\lambda)^2+\delta_{n,l}(\alpha)^2}.\end{aligned}$$

Since the contour C_3 has length $\mathcal{O}(\alpha^2)$, we estimate the integral over the expression above by $C\alpha$. Similar estimates hold on C_1, C_2, C_4 and C_5. The integral over $\tilde{f}_+(z)$ can be estimated analogously.

Pole-Term: The integral along C_0 over $f_-(z)$ yields the claimed leading term, the integral over $f_+(z)$ is zero.

Step 4: For $\lambda\in[E_{n,l'}(\alpha)-\delta_{n,l',-}(\alpha),E_{n,l'}(\alpha)+\delta_{n,l',+}(\alpha)]$ with $l'\neq l$ we observe that $\phi(\alpha)\in\operatorname{Ran}P^{(\alpha)}_{\mathrm{el},n,l}$ implies

$$\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)=(P^{(\alpha)}_{\mathrm{el},n,l'}(\theta)\otimes\chi_{H_\mathrm{f}\ge\rho_0}+\underline{P}^{(\alpha)}_{\mathrm{el},n,l'}(\theta)\otimes\mathbf{1}_\mathrm{f})\psi(\alpha;\theta)=\psi(\alpha;\theta)$$

and $P_{n,l'}(\theta)\psi(\alpha;\theta)=0$, which in turn shows

$$\begin{aligned}&f(\theta,\lambda)= \\ &=\big\langle\psi(\alpha;\bar{\theta}),\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta) \\ &\times\big[\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)\big]^{-1} \\ &\times P_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\big\rangle \\ &+\big\langle\psi(\alpha;\bar{\theta}),\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\big\rangle \\ =&:f_1(\theta,\lambda)+f_2(\theta,\lambda)\end{aligned}$$

using (51) in Lemma 7, where $f_1(\theta,\lambda)$ is the first summand. Using the resolvent identity we find $f_1(\theta,\lambda)=f_{1,a}(\theta,\lambda)+f_{1,b}(\theta,\lambda)+f_{1,c}(\theta,\lambda)$ with

$$\begin{aligned}&f_{1,a}(\theta,\lambda):=\big\langle\psi(\alpha;\bar{\theta}),\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1} \\ &\times\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta)\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)^{-1}P_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta) \\ &\times\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\big\rangle,\end{aligned}$$

$$
\begin{aligned}
f_{1,b}(\theta,\lambda) := &-\big\langle \psi(\alpha;\bar{\theta}), \overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta)\big[\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)\big]^{-1}P_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\big\rangle\\
&-\big\langle \psi(\alpha;\bar{\theta}), \overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta)\big[\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)\big]^{-1}P_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_{\alpha}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\big\rangle
\end{aligned}
$$

and

$$
\begin{aligned}
f_{1,c}(\theta,\lambda) := &\big\langle \psi(\alpha;\bar{\theta}), \overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha}\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta)\big[\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)\big]^{-1}P_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_{\alpha}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\big\rangle.
\end{aligned}
$$

We obtain

$$
\begin{aligned}
f_{1,a}(\theta,\lambda) = &\frac{1}{(E_{n,l}(\alpha)-\lambda)^2}\langle\psi(\alpha;\bar{\theta}), W_{0,1}^{(\alpha)}(\theta)P_{n,l'}(\theta)\\
&\times\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)^{-1}P_{n,l'}(\theta)W_{1,0}^{(\alpha)}(\theta)\psi(\alpha;\theta)\rangle.
\end{aligned}
$$

Lemma 5 c) and Lemma 12 a) imply

$$
|f_{1,a}(\theta,\lambda)| \leq \frac{1}{|E_{n,l}(\alpha)-\lambda|^2}\frac{g^2\rho_0^2}{g^2},
$$

which shows

$$
\int_{E_{n,l'}(\alpha)-\delta_{n,l',-}(\alpha)}^{E_{n,l'}(\alpha)+\delta_{n,l',+}(\alpha)} \mathrm{d}\lambda\, |f_{1,a}(\theta,\lambda)| \leq \frac{\rho_0^2}{\alpha^2} = \mathcal{O}(\alpha^2).
$$

In order to estimate $f_{1,b}(\theta,\lambda)$ it suffices to consider the first summand, which can be estimated according to

$$
\begin{aligned}
&\frac{1}{|E_{n,l}(\alpha)-\lambda|^2}|\langle\psi(\alpha;\bar{\theta}),\\
&\times P_{n,l}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta)\\
&\times \big[\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)\big]^{-1}P_{n,l'}(\theta)W_{1,0}^{(\alpha)}(\theta)\psi(\alpha;\theta)\rangle|\\
&\leq \frac{C}{|E_{n,l}(\alpha)-\lambda|^2}\frac{g^2 g\rho_0}{g^2},
\end{aligned}
$$

where we used Lemma 5 c), Lemma 12 a) and Lemma 7 b) in the last step. It follows that

$$\int_{E_{n,l'}(\alpha)-\delta_{n,l',-}(\alpha)}^{E_{n,l'}(\alpha)+\delta_{n,l',+}(\alpha)} \mathrm{d}\lambda\, |f_{1,b}(\theta,\lambda)| \le C\frac{g\rho_0}{\alpha^2} = \mathcal{O}(g) = \mathcal{O}(\alpha^{3/2}).$$

Eventually, we obtain by Lemma 12 a) and Lemma 7 b)

$$\begin{aligned}
&|f_{1,c}(\theta,\lambda)| = \frac{1}{|E_{n,l}(\alpha)-\lambda|^2}|\langle\psi(\alpha;\bar\theta), P_{n,l}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l'}(\theta)\\
&\times \big[\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-\lambda)\big]^{-1}\\
&\times P_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_\alpha\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l}(\theta)\psi(\alpha;\theta)\rangle|\\
&\le C\frac{1}{|E_{n,l}(\alpha)-\lambda|^2}\frac{g^4}{g^2}.
\end{aligned}$$

Integration yields

$$\int_{E_{n,l'}(\alpha)-\delta_{n,l',-}(\alpha)}^{E_{n,l'}(\alpha)+\delta_{n,l',+}(\alpha)} \mathrm{d}\lambda\, |f_{1,c}(\theta,\lambda)| \le C\frac{g^2}{\alpha^2} = \mathcal{O}(\alpha).$$

Now, we have to treat the term $f_2(\theta,\lambda)$. Using the resolvent identity we find $f_2(\theta,\lambda) = f_{2,a}(\theta,\lambda) + f_{2,b}(\theta,\lambda) + f_{3,c}(\theta,\lambda)$, with

$$\begin{aligned}
f_{2,a}(\theta,\lambda) :=& \langle\psi(\alpha;\bar\theta), \overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\rangle,\\
f_{2,b}(\theta,\lambda) :=& -\langle\psi(\alpha;\bar\theta), \overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\\
&\times \big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\rangle
\end{aligned}$$

and

$$\begin{aligned}
f_{2,c}(\theta,\lambda) :=& \langle\psi(\alpha;\bar\theta), \overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\\
&\times \overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta)\big(\overline{P}_{n,l'}(\theta)H_{\alpha,0}(\theta)\overline{P}_{n,l'}(\theta)-\lambda\big)^{-1}\overline{P}_{n,l'}(\theta)\psi(\alpha;\theta)\rangle.
\end{aligned}$$

Using the dilation analyticity we obtain

$$f_{2,a}(\theta,\lambda) = \frac{1}{E_{n,l}(\alpha)-\lambda}\langle\psi(\alpha;0),\psi(\alpha;0)\rangle,$$

which implies $f_{2,a}(\bar\theta,\lambda) - f_{2,a}(\theta,\lambda) = 0$. Moreover, we have

$$f_{2,b}(\theta,\lambda) = -\frac{1}{(E_{n,l}(\alpha)-\lambda)^2}\langle\psi(\alpha;\bar\theta), W^{(\alpha)}(\theta)\psi(\alpha;\theta)\rangle = 0$$

and

$$|f_{2,c}(\theta,\lambda)| = \frac{1}{|E_{n,l}(\alpha)-\lambda|^2}|\langle\psi(\alpha;\bar{\theta}), P_{n,l}(\theta)W^{(\alpha)}(\theta)\overline{P}_{n,l'}(\theta) \\ \times (\overline{P}_{n,l'}(\theta)H_\alpha(\theta)\overline{P}_{n,l'}(\theta)-\lambda)^{-1}\overline{P}_{n,l'}(\theta)W^{(\alpha)}(\theta)P_{n,l}(\theta)\psi(\alpha;\theta)\rangle| \\ \leq C\frac{g^2}{|E_{n,l}(\alpha)-\lambda|^2},$$

where we used Lemma 7 b) in the last step. Integration yields

$$\int_{E_{n,l'}(\alpha)-\delta_{n,l',-}(\alpha)}^{E_{n,l'}(\alpha)+\delta_{n,l',+}(\alpha)} \mathrm{d}\lambda\, |f_{2,c}(\theta,\lambda)| \leq C\frac{g^2}{\alpha^2} = \mathcal{O}(\alpha).$$

Step 5: For $\lambda \in [E_{n,n}(\alpha)+\delta_{n,n,+}(\alpha), E_n+\delta_{n,+}]$ and also for $\lambda \in [E_n - \delta_{n,-}, E_{n,1}(\alpha)-\delta_{n,1,-}(\alpha)]$ we have to proceed somewhat differently: We consider the first case only and make a case distinction.
1st Case: $1 < l \leq n$. Lemma 13 with $l'=1$ implies $\|\mathcal{F}_{P_{n,l'}(\theta)}(H_\alpha(\theta)-z)^{-1}\| \leq \frac{C}{\sin\vartheta|\lambda-E_{n,l'}(\alpha)|-Cg^2} \leq \frac{C}{\alpha^2}$, which we use to estimate $f_1(\theta,\lambda)$. $f_2(\theta,\lambda)$ can be estimated as in Step 4. Note that for both the estimates on $f_1(\theta,\lambda)$ and on $f_2(\theta,\lambda)$ the integration limits have to be changed accordingly. Thus, we obtain as in Step 4

$$\Big|\int_{E_n-\delta_{n,-}}^{E_{n,1}(\alpha)-\delta_{n,1,-}(\alpha)} \mathrm{d}\lambda\, e^{-\mathrm{i}\,\lambda s}F(\lambda)[f(\overline{\theta},\lambda)-f(\theta,\lambda)]\Big| = \mathcal{O}(\alpha).$$

2nd case: $l=1$. Using the resolvent identity we find

$$\begin{aligned} f(\theta,\lambda) &= \langle\psi(\alpha;\overline{\theta}), \mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta)-\lambda)^{-1}\psi(\alpha;\theta)\rangle \\ =&\langle\psi(\alpha;\overline{\theta}), [E_{n,l}(\alpha)-\lambda-Q_{n,l}^{(\alpha)}(\lambda;\theta)]^{-1}\psi(\alpha;\theta)\rangle \\ &-\langle\psi(\alpha;\overline{\theta}), [E_{n,l}(\alpha)-\lambda-Q_{n,l}^{(\alpha)}(\lambda;\theta)]^{-1} \\ &\times [\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta)-\lambda)-(E_{n,l}(\alpha)-\lambda+e^{-\theta}\mathbf{1}_{\mathrm{el}}\otimes H_{\mathrm{f}}-Q_{n,l}^{(\alpha)}(\lambda;\theta))P_{n,l}(\theta)] \\ &\times \left[\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta)-\lambda)\right]^{-1}\psi(\alpha;\theta)\rangle =: \tilde{f}(\theta,\lambda)+B(\theta,\lambda), \end{aligned}$$

where $\tilde{f}(\theta,\lambda)$ is the first summand. Lemma 13 yields the estimate $\|\mathcal{F}_{P_{n,l}(\theta)}(H_\alpha(\theta)-\lambda)^{-1}\| \leq \frac{C}{\sin\vartheta|\lambda-E_{n,l}(\alpha)|-Cg^2}$ and the same estimate for $[E_{n,l}(\alpha)-\lambda-Q_{n,l}^{(\alpha)}(\lambda;\theta)]^{-1}$. Thus, Lemma 8 implies

$$|B(\theta,\lambda)| \leq \frac{Cg^2\sqrt{\alpha}}{(\sin\vartheta|\lambda-E_{n,l}(\alpha)|-Cg^2)^2}$$

and finally

$$\int_{E_n-\delta_{n,-}}^{E_{n,1}(\alpha)-\delta_{n,1,-}(\alpha)} \mathrm{d}\lambda\, F(\lambda)|B(\theta,\lambda)| = \mathcal{O}(g) = \mathcal{O}(\alpha^{3/2})$$

with the same reasoning as in the non-relativistic case (Proof of [19, Theorem 1], Step 2). The same holds for $B(\bar{\theta}, \lambda)$.
To estimate $\tilde{f}(\theta, \lambda)$,we use

$$\begin{aligned}\tilde{f}(\theta, \lambda) = \langle \psi(\alpha; \bar{\theta}), [E_{n,1}(\alpha) - \lambda]^{-1} \psi(\alpha; \theta) \rangle + \\ + \langle \psi(\alpha; \bar{\theta}), [E_{n,l}(\alpha) - \lambda]^{-1} Q_{n,1}^{(\alpha)}(\lambda; \theta) [E_{n,1}(\alpha) - \lambda - Q_{n,1}^{(\alpha)}(\lambda; \theta)]^{-1} \psi(\alpha; \theta) \rangle.\end{aligned}$$

The first summand cancels with the corresponding summand of $\tilde{f}(\bar{\theta}, \lambda)$. The second summand can be estimated by $g^2 \frac{C}{|E_{n,1}(\alpha) - \lambda| \, (\sin \vartheta |\lambda - E_{n,l}(\alpha)| - C g^2)}$, which implies

$$\int_{E_n - \delta_{n,-}}^{E_{n,1}(\alpha) - \delta_{n,1,-}(\alpha)} \mathrm{d}\lambda \, F(\lambda) |\tilde{f}(\theta, \lambda) - \tilde{f}(\bar{\theta}, \lambda)| = \mathcal{O}(g^{2/3}) = \mathcal{O}(\alpha)$$

as above. □

Acknowledgements. *The author would like to thank his thesis advisor Heinz Siedentop for suggesting the investigation of the lifetime of relativistic excited states and constantly surveying the progress of the work, which is part of his PhD thesis. He would like to thank Ira Herbst and David Hasler for interesting conversations and Christiane Tretter for pointing out references [34, 33]. He acknowledges support by the Deutsche Forschungsgemeinschaft (DFG), grant no. SI 348/12-2, as well as by a "Doktorandenstipendium" from the German Academic Exchange Service (DAAD).*

References

[1] W. K. Abou Salem, J. Faupin, J. Froehlich, and I. M. Sigal. On the Theory of Resonances in Non-Relativistic QED and Related models. *ArXiv e-prints*, 711, November 2007.

[2] Asao Arai. A particle-field Hamiltonian in relativistic quantum electrodynamics. *J. Math. Phys.*, 41(7):4271–4283, 2000.

[3] Asao Arai and Masao Hirokawa. Ground States of a General Class of Quantum Field Hamiltonians. *Rev. Math. Phys.*, 12(8):1085–1135, 2000.

[4] Volker Bach, Jürg Fröhlich, and Israel Michael Sigal. Quantum electrodynamics of confined nonrelativistic particles. *Adv. Math.*, 137(2):299–395, 1998.

[5] Volker Bach, Jürg Fröhlich, and Israel Michael Sigal. Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field. *Comm. Math. Phys.*, 207(2):249–290, 1999.

[6] H.A. Bethe. The electromagnetic shift of energy levels. *Phys. Rev., Minneapolis, II. Ser.*, 72:339–341, 1947.

[7] William S. Bickel. Mean lives of some excited states in multiply ionized oxygen and neon. *Phys. Rev.*, 162(1):7, Oct 1967.

[8] William S. Bickel and Allan S. Goodman. Mean lives of the 2p and 3p levels in atomic hydrogen. *Physical Review*, 148(2):1–4, August 1970.

[9] G. Breit and E. Teller. Metastability of Hydrogen and Helium Levels. *Astrophysical Journal*, 91:215–238, March 1940.

[10] E. L. Chupp, L. W. Dotchin, and D. J. Pegg. Radiative mean-life measurements of some atomic-hydrogen excited states using beam-foil excitation. *Physical Review*, 175(1):45–50, November 1968.

[11] Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg. *Photons & Atoms*. WILEY-VCH Verlag GmbH& Co. KGaA, 2004.

[12] Jan Dereziński and Christian Gérard. Scattering theory of infrared divergent Pauli-Fierz Hamiltonians. *Ann. Henri Poincaré*, 5(3):523–577, 2004.

[13] R. C. Etherton, L. M. Beyer, W. E. Maddox, and L. B. Bridwell. Lifetimes of 3p, 4p, and 5p states in atomic hydrogen. *Physical Review A*, 2(6):2177–2179, December 1970.

[14] Herman Feshbach. Unified Theory of Nuclear Reactions. *Ann. Phys.*, 5:357–390, 1958.

[15] J. Fröhlich, M. Griesemer, and B. Schlein. Rayleigh scattering at atoms with dynamical nuclei. *Comm. Math. Phys.*, 271(2):387–430, 2007.

[16] M. Griesemer, E. Lieb, and M. Loss. Ground states in non-relativistic quantum electrodynamics. *Inv. Math.*, 145:557–595, 2001.

[17] D. Hasler and I. Herbst. Absence of Ground States for a Class of Translation Invariant Models of Non-relativistic QED. *Comm. Math. Phys.*, 279(3):769–787, may 2008.

[18] D. Hasler and I. Herbst. On the self-adjointness and domain of Pauli-Fierz type Hamiltonians. *Rev. Math. Phys.*, 20(7):787–800, 2008.

[19] D. Hasler, I. Herbst, and M. Huber. On the Lifetime of Quasi-Stationary States in Non-Relativistic QED. *Ann. Henri Poincaré*, 9(5):1005 – 1028, 2008.

[20] Emilie V. Haynsworth. Determination of the inertia of a partitioned Hermitian matrix. *Linear Algebra and Appl.*, 1(1):73–81, 1968.

[21] F. Hiroshima and H. Spohn. Ground state degeneracy of the Pauli-Fierz Hamiltonian with spin. *Adv. Theor. Math. Phys.*, 5(6):1091–1104, 2001.

[22] Fumio Hiroshima. Ground states of a model in nonrelativistic quantum electrodynamics. II. *J. Math. Phys.*, 41(2):661–674, 2000.

[23] James S. Howland. The Livsic matrix in perturbation theory. *J. Math. Anal. Appl.*, 50:415–437, 1975.

[24] Matthias Huber. *Modelle relativistischer und nicht-relativistischer Coulomb-Systeme [Dissertation an der Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München]*. Logos-Verlag Berlin GmbH, 2008.

[25] Matthias Huber. Spectral analysis of relativistic atoms – Dirac operators with singular potentials. *Submitted to Documenta Mathematica*, 2009.

[26] Takashi Ichinose. Tensor products of linear operators and the method of separation of variables. *Hokkaido Math. J.*, 3:161–189, 1974.

[27] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1 edition, 1966.

[28] Willis E. Lamb and Robert C. Retherford. Fine structure of the hydrogen atom by a microwave method. *Physical Review*, 72(3):241–243, August 1947.

[29] L.D. Landau, E.M. Lifschitz, V.B. Berestetskij, and L.P. Pitaevskij. *Lehrbuch der theoretischen Physik. Band 4: Quantenelektrodynamik. 7., bearb. Aufl.* Frankfurt am Main: H. Deutsch., 1991.

[30] M. S. Livshits. The application of non-self-adjoint operators to scattering theory. *Soviet Physics. JETP*, 4:91–98, 1957.

[31] M. S. Livšic. The method of non-selfadjoint operators in scattering theory. *Uspehi Mat. Nauk (N.S.)*, 12(1(73)):212–218, 1957.

[32] Michael Loss, Tadahiro Miyao, and Herbert Spohn. Lowest energy states in nonrelativistic QED: atoms and ions in motion. *J. Funct. Anal.*, 243(2):353–393, 2007.

[33] Reinhard Mennicken and Alexander K. Motovilov. Operator interpretation of resonances arising in spectral problems for 2×2 operator matrices. *Math. Nachr.*, 201:117–181, 1999.

[34] R. Menniken and A. K. Motovilov. Operator interpretation of resonances generated by (2×2)-matrix Hamiltonians. *Teoret. Mat. Fiz.*, 116(2):163–181, 1998.

[35] Tadahiro Miyao and Herbert Spohn. Spectral analysis of the semi-relativistic pauli-fierz hamiltonian. *Journal of Functional Analysis*, In Press, Corrected Proof:–, 2008.

[36] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I. Functional analysis.* Academic Press, New York, 1972.

[37] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics*, volume 4: Analysis of Operators. Academic Press, New York, 1 edition, 1978.

[38] Edgardo Stockmeyer and Heribert Zenk. Dirac operators coupled to the quantized radiation field: Essential self-adjointness à la Chernoff. *Lett. Math. Phys.*, 83(1):59–68, 2008.

[39] Bernd Thaller. *The Dirac Equation*. Texts and Monographs in Physics. Springer-Verlag, Berlin, 1 edition, 1992.

[40] Fuzhen Zhang, editor. *The Schur Complement and its Applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, New York, 2005.

Matthias Huber
Fakultät für Mathematik
und Informatik
FernUniversität in Hagen
58095 Hagen
Germany
matthias.huber@gmx.de

Rationally Connected Foliations on Surfaces

Sebastian Neumann

Received: November 21, 2008

Communicated by Thomas Peternell

Abstract. In this short note we study foliations on surfaces with rationally connected leaves. Our main result is that on a surface there exists a polarisation such that the Harder-Narasimhan filtration of the tangent bundle with respect to this polarisation yields the maximal rationally connected quotient of the surface.

2000 Mathematics Subject Classification: 14J26, 37F75

1 Introduction

Let X be a smooth projective variety over the complex numbers. In this note we are interested in foliations with rationally connected leaves. In [KSCT07] it is shown how to construct such foliations from the *Harder-Narasimhan filtration* of the tangent bundle of the variety. This construction depends heavily on a chosen polarisation, and therefore the question arises how this foliation varies with the polarisation.

There is another way to construct a fibration with rationally connected fibers, the *maximal rationally connected quotient*. This is a rational map whose fibers are rationally connected. Almost every rational curve in X lies in a fiber of this map.

We can ask if the Harder-Narasimhan filtration of the tangent bundle always induces the maximal rationally connected quotient with respect to any polarisation. The answer is negative already on surfaces as shown by an example of Thomas Eckl [Eck08].

In this note we will prove that on surfaces there always exists a polarisation such that the Harder-Narasimhan filtration yields the maximal rationally connected quotient.

Acknowledgments

The author was supported in part by the Graduiertenkolleg "Globale Strukturen in Geometrie und Analysis" of the Deutsche Forschungsgemeinschaft. The result is part of the author's forthcoming Ph.D. thesis written under the supervision of Stefan Kebekus. He would like to thank Stefan Kebekus, Thomas Eckl and Sammy Barkowski for numerous discussions.
The results in this note were presented at a Workshop in Grenoble in April 2008. Similar results have been obtained independently in [SCT08].

2 Preliminary Results and Notation

Let X be an n-dimensional projective variety over the complex numbers with an ample line bundle H. Given a torsion-free coherent sheaf $\mathcal{F}$ on X, we define the *slope of $\mathcal{F}$ with respect to H* to be

$$\mu_H(\mathcal{F}) := \frac{c_1(\mathcal{F}) \cdot H^{n-1}}{\mathrm{rk}(\mathcal{F})}.$$

We call $\mathcal{F}$ *semistable with respect to H* if for any nonzero proper subsheaf $\mathcal{G}$ of $\mathcal{F}$ we have $\mu_H(\mathcal{G}) \leq \mu_H(\mathcal{F})$.
If there exists a nonzero subsheaf $\mathcal{G} \subset \mathcal{F}$ such that $\mu_H(\mathcal{G}) > \mu_H(\mathcal{F})$, we will call $\mathcal{G}$ a *destabilizing subsheaf* of $\mathcal{F}$.

Theorem 2.1 ([Mar80, Proposition 1.5.]). *Let $\mathcal{F}$ be a torsion-free coherent sheaf on a smooth projective variety and H be an ample line bundle on X. There exists a unique filtration*

$$0 = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_k = \mathcal{F}$$

of $\mathcal{F}$ depending on H, the Harder-Narasimhan filtration *or* HN-filtration*, with the following properties:*

(i) The quotients $\mathcal{G}_i := \mathcal{F}_i/\mathcal{F}_{i-1}$ are torsion-free and semistable.

(ii) The slopes of the quotients satisfy $\mu_H(\mathcal{G}_1) > \ldots > \mu_H(\mathcal{G}_k)$.

Definition 2.2. Let $\mathcal{F}$ be a torsion-free coherent sheaf on a smooth projective variety. The unique sheaf $\mathcal{F}_1$ appearing in the Harder-Narasimhan filtration of $\mathcal{F}$ is called *the maximal destabilizing subsheaf of $\mathcal{F}$*.

Definition 2.3. Let $\mathcal{F}$ be a coherent torsion-free sheaf on a smooth projective variety with Harder-Narasimhan filtration

$$0 = \mathcal{F}_0 \subset \ldots \subset \mathcal{F}_k = \mathcal{F}$$

with respect to an ample line bundle H. If the slope of the quotient $\mathcal{F}_i/\mathcal{F}_{i-1}$ is positive with respect to H, then $\mathcal{F}_i$ is called *positive with respect to H*.

Remark 2.4. Note that the construction of the Harder-Narasimhan filtration naturally extends to $\mathbb{Q}$- and $\mathbb{R}$-divisors, i.e. we do not need to assume that the chosen polarisation is integral.

Obviously, the Harder-Narasimhan filtration depends only on the numerical class of the chosen ample bundle. In particular it makes sense to ask how the filtration of a given sheaf depends on the ample bundle sitting in the finite dimensional vector space of all divisors modulo numerical equivalence.
We can now state an important result originally formulated by Miyaoka and explicitly shown in [KSCT07]. For a survey on these and related results we refer the reader to [KSC06].

Theorem 2.5 ([KSCT07, Theorem 1]). *Let X be a smooth projective variety and let*

$$0 = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_k = TX$$

be the Harder-Narasimhan filtration of the tangent bundle with respect to a polarisation H. Write $\mu_i := \mu_H(\mathcal{F}_i/\mathcal{F}_{i-1})$ for the slopes of the quotients. Assume $\mu_1 > 0$ and set $m := \max\{i \in \mathbb{N} | \mu_i > 0\}$. Then each $\mathcal{F}_i$ with $i \leq m$ is a foliation, i.e. a saturated subsheaf of the tangent bundle closed under Lie bracket. Furthermore the leaves of these foliations are algebraic and for general $x \in X$ the closure of the leaf through x is rationally connected.

Let X be a smooth projective variety and assume the conditions of Theorem (2.5) are fulfilled. Thus we obtain foliations $\mathcal{F}_1, \ldots, \mathcal{F}_k$ with algebraic and rationally connected leaves. By setting

$$\begin{array}{cccc} q_i: & X & \dashrightarrow & \mathrm{Im}(q_i) \subset \mathrm{Chow}(X) \\ & x & \mapsto & \mathcal{F}_i\text{-leaf through } x \end{array}$$

we obtain a rational map, such that the closure of the general fibre is rationally connected, see [KSCT07] Section 7.
There is another map with this property called the *maximal rationally connected quotient*, or *MRC-quotient*, for short based on a construction by Campana [Cam81] [Cam94] and Kollár-Miyaoka-Mori [KMM92], see also [Kol96, Chapter IV, Theorem 5.2].

Theorem 2.6 ([KMM92, Theorem 2.7.]). *Let X be a smooth projective variety. There exists a variety Z and a rational map $\phi : X \dashrightarrow Z$ with the following properties:*

- *the fibers of ϕ are rationally connected,*
- *a very general fiber of ϕ is an equivalence class with respect to rational connectivity and*
- *up to birational equivalence the map ϕ and the variety Z are unique.*

In this paper we ask if the Harder-Narasimhan filtration with respect to a certain polarisation yields the MRC-quotient. We will give a positive answer for surfaces in the next section.

3 Rationally Connected Foliations on Surfaces and the MRC-quotient

In this section X denotes a smooth projective surface over the the field of complex numbers.
We want to investigate the regions in the ample cone which induce the same HN-filtration. More precisely we divide the ample cone into parts, so that in each part we get the same HN-filtration of the tangent bundle. With this at hand we are able to show that the MRC-quotient comes from the Harder-Narasimhan filtration of the tangent bundle with respect to a certain polarisation.
In order to compute the HN-filtration of the tangent bundle on surfaces, we only have to search for a destabilizing subsheaf whose quotient is torsion-free. This is formulated in the next lemma.

Lemma 3.1. *Let X be a smooth projective surface. If $\mathcal{F} \subset TX$ is a destabilizing subsheaf with respect to a polarisation such that $TX/\mathcal{F}$ is torsion-free, then the Harder-Narasimhan filtration is given by $0 \subset \mathcal{F} \subset TX$.*

Proof. Let H be a polarisation and $\mathcal{F}$ a destabilizing subsheaf of TX with respect to H. Consider the exact sequence

$$0 \to \mathcal{F} \to TX \to TX/\mathcal{F} \to 0.$$

Using that the rank and the first Chern class are additive in short exact sequences, we obtain

$$\mu_H(TX) = \frac{1}{2}\mu_H(TX/\mathcal{F}) + \frac{1}{2}\mu_H(\mathcal{F}).$$

Since $\mu_H(\mathcal{F}) > \mu_H(TX)$, we therefore have $\mu_H(\mathcal{F}) > \mu_H(TX/\mathcal{F})$. That is,

$$0 \subset \mathcal{F} \subset TX$$

satisfies the properties of the Harder-Narasimhan filtration and by the uniqueness of the HN-filtration we are done. □

Notation 3.2. We write $N^1(X)$ for the Néron-Severi group and $N^1_{\mathbb{Q}}(X)$ (resp. $N^1_{\mathbb{R}}(X)$) for the vector space of $\mathbb{Q}$–divisors (resp. $\mathbb{R}$–divisors) modulo numerical equivalence on X. The convex cone of all ample $\mathbb{R}$–divisors in $N^1_{\mathbb{R}}(X)$ is denoted by $\mathrm{Amp}_{\mathbb{R}}(X)$.

Now we define the regions in $\mathrm{Amp}_{\mathbb{R}}(X)$ we are interested in. Let $H \in N^1_{\mathbb{R}}(X)$ be an ample bundle. If TX is not semistable with respect to H, let $\mathcal{F}$ be the maximal destabilizing subsheaf of TX with respect to H, i.e. the Harder-Narasimhan filtration of TX with respect to H is given by $0 \subset \mathcal{F} \subset TX$. We call

$$\Delta_H := \left\{ \tilde{H} \in \mathrm{Amp}_{\mathbb{R}}(X) \mid \big(c_1(\mathcal{F}) - \frac{1}{2}c_1(TX)\big) \cdot \tilde{H} > 0 \right\}$$

the *destabilizing chamber with respect to H*.

Remark 3.3. By Lemma (3.1) the condition $(c_1(\mathcal{F}) - \frac{1}{2}c_1(TX)) \cdot \tilde{H} > 0$ ensures that for all polarisations in Δ_H we get the same HN-filtration, namely $0 \subset \mathcal{F} \subset TX$. So we have indeed defined the regions in the ample cone, in which the Harder-Narasimhan filtration of the tangent bundle remains constant.

Note that if the tangent bundle is semistable with respect to a certain polarisation, then we get a chamber such that for all polarisations in this chamber TX is semistable. This region is called the *semistable chamber*.
Concerning the structure of these chambers we prove the following lemma.

Lemma 3.4. *Let X be a smooth projective surface. We have:*

(i) The destabilizing chambers and the semistable chamber are convex cones in $\mathrm{Amp}_{\mathbb{R}}(X)$.

(ii) The semistable chamber is closed in $\mathrm{Amp}_{\mathbb{R}}(X)$.

(iii) The destabilizing chambers are open in $\mathrm{Amp}_{\mathbb{R}}(X)$.

(iv) The destabilizing chambers and the semistable chamber give a decomposition of the ample cone, i.e. the union of all chambers is the ample cone and the chambers are pairwise disjoint.

Proof. The convexity property of both the semistable chamber and the destabilizing chamber follows directly from the linearity of the intersection product. Statement (iii) is a direct consequence of the continuity of the intersection product, since for a maximal destabilizing subsheaf $\mathcal{F} \subset TX$ the condition

$$\big(c_1(\mathcal{F}) - \frac{1}{2}c_1(TX)\big) \cdot H > 0$$

is an open condition.
To prove (iv) note that by definition of the chambers, each polarisation appears in at least one chamber. Since for a given polarisation the associated maximal destabilizing subsheaf of TX is unique, the polarisation appears in exactly one chamber.
Statement (ii) is a direct consequence of (iii) and (iv). □

In the proof of our main result, we will use the following corollary.

Corollary 3.5. *Let X be a smooth projective surface. Let ℓ be a line segment in $\mathrm{Amp}_{\mathbb{R}}(X)$, such that ℓ does not intersect the semistable chamber. Then ℓ is contained in a single destabilizing chamber.*

Proof. Assume ℓ intersects at least two destabilizing chambers. By Lemma (3.4) we get a partition of ℓ into disjoint open sets. This is impossible because ℓ is connected. □

To prove semistability of the tangent bundle on certain surfaces having many automorphisms, we will give a useful lemma. Let $\sigma \in \mathrm{Aut}(X)$ and $\mathcal{F} \subset TX$. By means of the differential of σ, we can identify TX and σ^*TX. Thus we can interpret $\sigma^*(\mathcal{F})$ as a subsheaf of TX. For instance, if $p \in X$ and $\mathcal{F} := TX \otimes \mathcal{I}_p$, then $\sigma^*(\mathcal{F})$ is identified with $TX \otimes \mathcal{I}_{\sigma^{-1}(p)} \subset TX$.

LEMMA 3.6. *Let X be a smooth projective surface and let $\sigma \in \mathrm{Aut}^0(X)$. Let $\mathcal{F}$ be the maximal destabilizing subsheaf of TX with respect to some polarisation. We then have $\sigma^*\mathcal{F} = \mathcal{F}$. In particular: If $\mathcal{F}$ is a foliation then the automorphism σ maps each leaf of $\mathcal{F}$ to another leaf of $\mathcal{F}$.*

Proof. Let $H \in \mathrm{Amp}_{\mathbb{R}}(X)$ and let $\mathcal{F}$ be the maximal destabilizing subsheaf of TX with respect to H. We compute the slope of $\sigma^*(\mathcal{F}) \subset TX$:

$$\begin{array}{rcl} \mu_H\big(\sigma^*(\mathcal{F})\big) & = & H \cdot \big(c_1(\sigma^*(\mathcal{F}))\big) \\ & = & H \cdot \sigma^*(c_1(\mathcal{F})) \\ & = & H \cdot c_1(\mathcal{F}) \\ & > & \frac{1}{2} c_1(TX) \cdot H. \end{array}$$

We give an explanation of the third equality. Recall that the group of automorphisms acts on the Néron-Severi group. Since $N^1(X)$ is discrete, $\mathrm{Aut}^0(X)$ acts trivially on $N^1(X)$, i.e. $\sigma^*(c_1(\mathcal{F})) = c_1(\mathcal{F})$.
We therefore have shown that $\sigma^*(\mathcal{F})$ is a destabilizing subsheaf of TX. By Lemma (3.1) and the uniqueness of the maximal destabilizing subsheaf of TX, we conclude that $\sigma^*\mathcal{F} = \mathcal{F}$. □

EXAMPLE 3.7. *Hirzebruch Surfaces*
Let Σ_n be the n-th Hirzebruch surface and let $\pi : \Sigma_n \longrightarrow \mathbb{P}^1$ be the projection onto the projective line. We denote the fiber under the projection by f and the distinguished section with selfintersection $-n$ by C_0. Recall (see [Har77], chapter V.2) that $N^1_{\mathbb{R}}(\Sigma_n) =< C_0, f >$ and a divisor $D \equiv_{num} aC_0 + bf$ is ample if and only if $a > 0$ and $b > an$. The canonical bundle is given by $-K_{\Sigma_n} = 2C_0 + (2+n)f$. The relative tangent bundle of π is a natural candidate for a destabilizing subbundle. We have the sequence

$$0 \to T_{\Sigma_n/\mathbb{P}^1} \to T\Sigma_n \to \pi^*T\mathbb{P}^1 \to 0$$

Let $H := xC_0 + yf$ be a polarisation. Then one can compute that $T_{\Sigma_n/\mathbb{P}^1}$ is destabilizing if and only if $-2x - nx + 2y > 0$. In particular we compute for $n \geq 2$:

$$-2x - nx + 2y > -2x - nx + 2nx = -2x + nx \geq 0.$$

Therefore, for $n \geq 2$ the HN-filtration is given by

$$0 \subset T_{X/\mathbb{P}^1} \subset TX$$

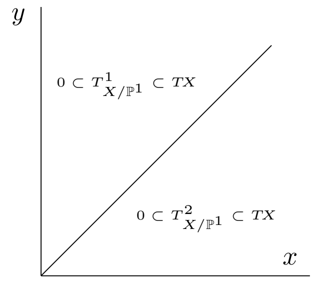


Figure 1: The ample cone of $X = \Sigma_0$ and the chamber structure. Here $T^1_{X/\mathbb{P}^1}$ and $T^2_{X/\mathbb{P}^1}$ denote the relative tangent bundle of the first and second projection.

for all polarisations. In other words we obtain only one destabilizing chamber. For $n = 0$ we have $\Sigma_0 = \mathbb{P}^1 \times \mathbb{P}^1$ and we get three chambers. The two destabilizing chambers correspond to the two relative tangent bundles of the projections. They are cut out by the inequalities $x > y$ and $x < y$. There is a chamber of semistability, which is determined by the equation $x = y$.
For $n = 1$ we see that for $x > \frac{3}{2}y$ the relative tangent bundle is destabilizing. Since Σ_1 is the projective plane blown up at a point p, the group of automorphisms is the automorphism group of the projective plane leaving p fixed. The destabilizing foliation corresponds to the radial foliation through p in the plane. So if there were another foliation $\mathcal{F}$ coming from the Harder-Narasimhan filtration of $T\Sigma_1$, we could deform the leaves with these automorphisms. Then we would again obtain leaves of this foliation by Lemma (3.6). So unless $\mathcal{F}$ is the foliation given by the relative tangent bundle of the projection morphism, we could deform each leaf of $\mathcal{F}$ while leaving a point on the leaf not lying on C_0 fixed. Thus the foliation induced by $\mathcal{F}$ would have singularities on a dense open subset of Σ_1 which is absurd. So the tangent bundle is semistable for $x \leq \frac{3}{2}y$.

Now we want to answer the question if there always exists a polarisation, such that the Harder-Narasimhan filtration gives rise to the MRC-quotient.

Theorem 3.8. *Let X be a uniruled projective surface. Then there exists a polarisation, such that the maximal rationally connected quotient of X is given by the foliation associated to highest positive term in the Harder-Narasimhan filtration with respect to this polarisation.*

Proof. To start, observe that there is always a polarisation A such that $c_1(TX) \cdot A > 0$. Indeed, there exists a free rational curve $f : \mathbb{P}^1 \to X$. See [Deb01, Corollary 4.11] for a proof of the existence of such a curve. Writing

$$f^*(TX) = \mathcal{O}(a_1) \oplus \mathcal{O}(a_2)$$

with $a_1 + a_2 \geq 2$, we compute

$$-K_X \cdot f_*\mathbb{P}^1 = a_1 + a_2 \geq 2.$$

Write $\ell := f_*\mathbb{P}^1$ for this curve. Since ℓ is movable, it is in particular nef. So for an ample class H, the class $\ell + \epsilon H$ will be ample. Thus for sufficiently small ϵ the class $\ell + \epsilon H$ will intersect $-K_X$ positively.

First let us assume that X is not rationally connected. As we have just seen, we can find a polarisation H with $c_1(TX) \cdot H > 0$. There exists a destabilizing subsheaf $\mathcal{F}$ of TX, since otherwise X would be rationally connected by Theorem (2.5). Furthermore the slope of $\mathcal{F}$ has to be bigger than $c_1(TX) \cdot H$ and therefore positive. So this sheaf will give a foliation with rationally connected leaves and hence the maximal rationally connected quotient.

Now we consider the case where X is rationally connected. We then fix a very free rational curve ℓ on X. For a proof of the existence of a very free rational curve see [Deb01, Corollary 4.17]. This means that $TX|_\ell$ is ample. So we know that each quotient of $TX|_\ell$ has strictly positive degree.

Since ℓ is movable, it is in particular nef. Let H be an ample class. Because ℓ is nef, we know that $H_\epsilon := \ell + \epsilon H$ is ample in $N^1_{\mathbb{Q}}(X)$ for any $\epsilon > 0$. Observe that $c_1(TX) \cdot H_\epsilon > 0$ for sufficiently small ϵ, say for $0 < \epsilon < \epsilon_0$. If TX is semistable with respect to a certain polarisation H_ϵ with $0 < \epsilon < \epsilon_0$, the claim follows since TX has positive slope and induces a trivial foliation which gives the rationally connected quotient. If TX is not semistable for all polarisations H_ϵ with $0 < \epsilon < \epsilon_0$, let $\mathcal{F}_\epsilon$ be the maximal destabilizing subsheaf of TX with respect to H_ϵ. Because of Corollary (3.5) the ray H_ϵ stays in one destabilizing chamber and Remark (3.3) ensures that $\mathcal{F} := \mathcal{F}_\epsilon$ remains constant.

Now it is clear that for sufficiently small ϵ both the slope of $\mathcal{F}$ and the slope of $TX/\mathcal{F}$ will be positive with respect to H_ϵ. Therefore the HN-filtration of TX with respect to H_ϵ yields the maximal rationally connected quotient. □

REFERENCES

[Cam81] Frédéric Campana. Coréduction algébrique d'un espace analytique faiblement kählérien compact. *Invent. Math.*, 63(2):187–223, 1981.

[Cam94] Frédéric Campana. Remarques sur le revêtement universel des variétés kählériennes compactes. *Bull. Soc. Math. France*, 122(2):255–284, 1994.

[Deb01] Olivier Debarre. *Higher-dimensional algebraic geometry*. Universitext. Springer-Verlag, New York, 2001.

[Eck08] Thomas Eckl. Lower bounds for Seshadri constants. *Math. Nachr.*, 281(8):1119–1128, 2008.

[Har77] Robin Hartshorne. *Algebraic geometry*. Springer-Verlag, New York, 1977. Graduate Texts in Mathematics, No. 52.

[KMM92] János Kollár, Yoichi Miyaoka, and Shigefumi Mori. Rationally connected varieties. *J. Algebraic Geom.*, 1(3):429–448, 1992.

[Kol96] János Kollár. *Rational curves on algebraic varieties*, volume 32 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 1996.

[KSC06] Stefan Kebekus and Luis Solá Conde. Existence of rational curves on algebraic varieties, minimal rational tangents, and applications. In *Global aspects of complex geometry*, pages 359–416. Springer, Berlin, 2006.

[KSCT07] Stefan Kebekus, Luis Solá Conde, and Matei Toma. Rationally connected foliations after Bogomolov and McQuillan. *J. Algebraic Geom.*, 16(1):65–81, 2007.

[Mar80] Masaki Maruyama. Boundedness of semistable sheaves of small ranks. *Nagoya Math. J.*, 78:65–94, 1980.

[SCT08] Luis Solá Conde and Matei Toma. Maximally rationally connected fibrations and movable curves. *preprint, arXiv:0811.2141v2 [math.AG]*, 2008.

Sebastian Neumann
Albert-Ludwigs-Universität
Freiburg
Mathematisches Institut
Eckerstrasse 1
D-79104 Freiburg

The Allegretto-Piepenbrink Theorem for Strongly Local Dirichlet Forms

Dedicated to Jürgen Voigt in celebration of his 65th birthday

Daniel Lenz, Peter Stollmann, Ivan Veselić

Received: November 18, 2008
Revised: March 11, 2009

Communicated by Heinz Siedentop

Abstract. The existence of positive weak solutions is related to spectral information on the corresponding partial differential operator.

2000 Mathematics Subject Classification: 35P05, 81Q10

Introduction

The Allegretto-Piepenbrink theorem relates solutions and spectra of 2nd order partial differential operators H and has quite some history, cf. [1, 5, 6, 7, 34, 35, 36, 43, 37, 38].
One way to phrase it is that the supremum of those real E for which a nontrivial positive solution of $H\Phi = E\Phi$ exists coincides with the infimum of the spectrum of H. In noncompact cases this can be sharpened in the sense that nontrivial positive solutions of the above equation exist for all $E \leq \inf \sigma(H)$.
In the present paper we consider the Allegretto-Piepenbrink theorem in a general setting in the sense that the coefficients that are allowed may be very singular. In fact, we regard $H = H_0 + \nu$, where H_0 is the generator of a strongly local Dirichlet form and ν is a suitable measure perturbation. Let us stress, however, that one main motivation for the present work is the conceptual simplicity that goes along with the generalisation.
The Allegretto-Piepenbrink theorem as stated above consists of two statements: the first one is the fact that positive solutions can only exist for E below the spectrum. Turned around this means that the existence of a nontrivial positive solution of $H\Phi = E\Phi$ implies that $H \geq E$. For a strong enough notion of positivity, this comes from a "ground state transformation". We present this

simple extension of known classical results in Section 2, after introducing the necessary set-up in Section 1. For the ground state transformation not much structure is needed.
For the converse statement, the existence of positive solutions below $\sigma(H)$, we need more properties of H and the underlying space: noncompactness, irreducibility and what we call a Harnack principle. All these analytic properties are well established in the classical case. Given these tools, we prove this part of the Allegretto-Piepenbrink theorem in Section 3 with arguments reminiscent of the corresponding discussion in [18]. For somewhat complementary results we refer to [14] where it is shown that existence of a nontrivial subexponentially bounded solution of $H\Phi = E\Phi$ yields that $E \in \sigma(H)$. This implies, in particular, that the positive solutions we construct for energies below the spectrum cannot behave to well near infinity. We dedicate this paper to Jürgen Voigt - teacher, collaborator and friend - in deep gratitude and wish him many more years of fun in analysis.

1. Basics and notation concerning strongly local Dirichlet forms and measure perturbations

Dirichlet forms. We will now describe the set-up; we refer to [22] as the classical standard reference as well as [13, 19, 23, 31] for literature on Dirichlet forms. Let us emphasize that in contrast to most of the work done on Dirichlet forms, we treat real and complex function spaces at the same time and write $\mathbb{K}$ to denote either $\mathbb{R}$ or $\mathbb{C}$.
Throughout we will work with a locally compact, separable metric space X endowed with a positive Radon measure m with $\operatorname{supp} m = X$.
The central object of our studies is a regular Dirichlet form $\mathcal{E}$ with domain $\mathcal{D}$ in $L^2(X)$ and the selfadjoint operator H_0 associated with $\mathcal{E}$. Let us recall the basic terminology of Dirichlet forms: Consider a dense subspace $\mathcal{D} \subset L^2(X, m)$ and a sesquilinear and non-negative map $\mathcal{E} \colon \mathcal{D} \times \mathcal{D} \to \mathbb{K}$ such that $\mathcal{D}$ is closed with respect to the energy norm $\| \cdot \|_{\mathcal{E}}$, given by

$$\|u\|_{\mathcal{E}}^2 = \mathcal{E}[u,u] + \|u\|_{L^2(X,m)}^2,$$

in which case one speaks of a *closed form* in $L^2(X, m)$. In the sequel we will write

$$\mathcal{E}[u] := \mathcal{E}[u,u].$$

The selfadjoint operator H_0 associated with $\mathcal{E}$ is then characterized by

$$D(H_0) \subset \mathcal{D} \text{ and } \mathcal{E}[f,v] = (H_0 f \mid v) \quad (f \in D(H_0), v \in \mathcal{D}).$$

Such a closed form is said to be a *Dirichlet form* if $\mathcal{D}$ is stable under certain pointwise operations; more precisely, $T : \mathbb{K} \to \mathbb{K}$ is called a *normal contraction* if $T(0) = 0$ and $|T(\xi) - T(\zeta)| \leq |\xi - \zeta|$ for any $\xi, \zeta \in \mathbb{K}$ and we require that for any $u \in \mathcal{D}$ also

$$T \circ u \in \mathcal{D} \text{ and } \mathcal{E}[T \circ u] \leq \mathcal{E}[u].$$

Here we used the original condition from [9] that applies in the real and the complex case at the same time. Today, particularly in the real case, it is mostly

expressed in an equivalent but formally weaker statement involving $u \vee 0$ and $u \wedge 1$, see [22], Thm. 1.4.1 and [31], Section I.4.
A Dirichlet form is called *regular* if $\mathcal{D} \cap C_c(X)$ is large enough so that it is dense both in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$ and $(C_c(X), \|\cdot\|_\infty)$, where $C_c(X)$ denotes the space of continuous functions with compact support.

Capacity. Due to regularity, we find a set function, the *capacity* that allows to measure the size of sets in a way that is adapted to the form $\mathcal{E}$: For $U \subset X$, U open,

$$\operatorname{cap}(U) := \inf\{\|v\|_{\mathcal{E}}^2 \mid v \in \mathcal{D}, \chi_U \leq v\}, (\inf \emptyset = \infty),$$

and

$$\operatorname{cap}(A) := \inf\{\operatorname{cap}(U) \mid A \subset U\}$$

(see [22], p. 61f.). We say that a property holds *quasi-everywhere*, short *q.e.*, if it holds outside a set of capacity 0. A function $f : X \to \mathbb{K}$ is said to be *quasi-continuous*, *q.c.* for short, if, for any $\varepsilon > 0$ there is an open set $U \subset X$ with $\operatorname{cap}(U) \leq \varepsilon$ so that the restriction of f to $X \setminus U$ is continuous.
A fundamental result in the theory of Dirichlet forms says that every $u \in \mathcal{D}$ admits a q.c. representative $\tilde{u} \in u$ (recall that $u \in L^2(X, m)$ is an equivalence class of functions) and that two such q.c. representatives agree q.e. Moreover, for every Cauchy sequence (u_n) in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$ there is a subsequence (u_{n_k}) such that the $(\tilde{u}_{n_k})$ converge q.e. (see [22], p.64f).

Measure perturbations. We will be dealing with Schrödinger type operators, i.e., perturbations $H = H_0 + V$ for suitable potentials V. In fact, we can even include measures as potentials. Here, we follow the approach from [45, 46]. Measure perturbations have been regarded by a number of authors in different contexts, see e.g. [4, 24, 47] and the references there.
We denote by $\mathcal{M}_R(U)$ the signed Radon measures on the open subset U of X and by $\mathcal{M}_{R,0}(U)$ the subset of measures ν that do not charge sets of capacity 0, i.e., those measures with $\nu(B) = 0$ for every Borel set B with $\operatorname{cap}(B) = 0$. In case that $\nu = \nu_+ - \nu_- \in \mathcal{M}_{R,0}(X)$ we can define

$$\nu[u, v] = \int_X \tilde{u}\overline{\tilde{v}} d\nu \text{ for } u, v \in \mathcal{D} \text{ with } \tilde{u}, \tilde{v} \in L^2(X, \nu_+ + \nu_-).$$

We have to rely upon more restrictive assumptions concerning the negative part ν_- of our measure perturbation. We write $\mathcal{M}_{R,1}$ for those measures $\nu \in \mathcal{M}_R(X)$ that are $\mathcal{E}$-bounded with bound less than one; i.e. measures ν for which there is a $\kappa < 1$ and a c_κ such that

$$\nu[u, u] \leq \kappa \mathcal{E}[u] + c_\kappa \|u\|^2.$$

The set $\mathcal{M}_{R,1}$ can easily be seen to be a subset of $\mathcal{M}_{R,0}$. We write $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$ if the positive part ν_+ of the measure is in $\mathcal{M}_{R,0}$ and the negative ν_- is in $\mathcal{M}_{R,1}$.
By the KLMN theorem (see [39], p. 167), the sum $\mathcal{E} + \nu$ given by $D(\mathcal{E} + \nu) = \{u \in \mathcal{D} \mid \tilde{u} \in L^2(X, \nu_+)\}$ is closed and densely defined (in fact $\mathcal{D} \cap C_c(X) \subset D(\mathcal{E} + \nu)$) for $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. We denote the associated selfadjoint operator

by $H_0 + \nu$. An important special case is given by $\nu = V dm$ with $V \in L^1_{\rm loc}(X)$. As done in various papers, one can allow for more singular measures, a direction we are not going to explore here due to the technicalities involved.

Approximation and Regularity. By assumption the Dirichlet form $(\mathcal{E}, \mathcal{D})$ is regular. We show now that this property carries over to the perturbed form $(\mathcal{E} + \nu, D(\mathcal{E} + \nu))$. Along the way we prove an approximation result which will be useful in the context of Theorem 2.3. It will be convenient to introduce a notation for the natural norm in $D(\mathcal{E} + \nu)$. For all $\psi \in D(\mathcal{E} + \nu)$ we define

$$\|\psi\|^2_{\mathcal{E}+\nu} := \|\psi\|^2_{\mathcal{E}} + \nu_+(\psi, \psi)\,.$$

Lemma 1.1. *Let $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$, and $\mathcal{E}$ and $\mathcal{E} + \nu$ be as above. Then*

(a) For each $u \in D(\mathcal{E} + \nu)$ there exists a sequence (u_n) in $\mathcal{D} \cap L^\infty_c(X)$ such that $|u_n| \leq |u|$ for all $n \in \mathbb{N}$ and $\|u - u_n\|_{\mathcal{E}+\nu} \to 0$ for $n \to \infty$.

(b) For any $v \in \mathcal{D} \cap L^\infty_c(X)$ with $v \geq 0$ and any $\eta \in \mathcal{D} \cap C_c(X)$ with $\eta \equiv 1$ on the support of v there exists a sequence (ϕ_n) in $\mathcal{D} \cap C_c(X)$ with $\phi_n \to v$ in $(D(\mathcal{E} + \nu), \|\cdot\|_{\mathcal{E}+\nu})$ and $0 \leq v, \phi_n \leq \eta$ for all $n \in \mathbb{N}$.

In particular, $\mathcal{D} \cap C_c(X)$ is dense in $(D(\mathcal{E} + \nu), \|\cdot\|_{\mathcal{E}+\nu})$ and the form $(\mathcal{E} + \nu, D(\mathcal{E} + \nu))$ is regular.

Note that $\mathcal{D} \cap L^\infty_c(X) \subset D(\mathcal{E} + \nu)$.

Proof. By splitting u into its real and imaginary and then positive and negative part we can assume afterwards that $u \geq 0$.

We now prove the first statement. Since $\mathcal{E}$ is regular there exists a sequence (ϕ_n) in $\mathcal{D} \cap C_c(X)$ such that $\|u - \phi_n\|_{\mathcal{E}} \to 0$. By the contraction property of Dirichlet forms we can suppose that $\phi_n \geq 0$ and deduce that $u_n := \phi_n \wedge u \to u$ in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$ as well. (Note that $u_n = T(\phi_n - u)$ with the normal contraction $T : \mathbb{R} \longrightarrow \mathbb{R}$, $T(y) = y$ for $y \leq 0$ and $T(y) = 0$ for $y \geq 0$.) Choosing a subsequence, if necessary, we can make sure that $\tilde{u}_n \to \tilde{u}$ q.e. Therefore $\tilde{u}_n \to \tilde{u}$ a.e. with respect to ν_+ and ν_-. Now $(\mathcal{E} + \nu)$-convergence follows by Lebesgue's dominated convergence theorem.

Now we turn to the proof of the second statement. Without loss of generality we may chose $0 \leq v \leq 1$. Consider the convex set

$$C := \{\phi \in \mathcal{D} \cap C_c(X) \mid 0 \leq \phi \leq \eta\}$$

Since C is convex, its weak and norm closure in $(D(\mathcal{E} + \nu), \|\cdot\|_{\mathcal{E}+\nu})$ coincide. Therefore it suffices to construct a sequence $(\phi_n) \subset C$ that is bounded w.r.t $\|\cdot\|_{\mathcal{E}+\nu}$ and converges to $\tilde{v}$ q.e. By regularity we can start with a sequence $(\psi_n) \subset \mathcal{D} \cap C_c(X)$ such that $\psi_n \to v$ w.r.t $\|\cdot\|_{\mathcal{E}}$ and $\tilde{\psi}_n \to \tilde{v}$ q.e. By the contraction property of Dirichlet forms the sequence $\phi_n := 0 \vee \psi_n \wedge \eta$ is bounded in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$. Since $0 \leq \phi_n \leq \eta$, (ϕ_n) is also bounded in $L^2(\nu_+ + \nu_-)$. We finally prove the 'in particular' statemtent. Since $\mathcal{E}$ is regular, we can find an $\eta \in \mathcal{D} \cap C_c(X)$, $0 \leq \eta \leq 1$ with $\eta \equiv 1$ on $\operatorname{supp} v$. Now, the proof follows from the previous parts. $\square$

Strong locality and the energy measure. $\mathcal{E}$ is called *strongly local* if

$$\mathcal{E}[u,v]=0$$

whenever u is constant a.s. on the support of v.
The typical example one should keep in mind is the Laplacian

$$H_0=-\Delta \text{ on } L^2(\Omega), \quad \Omega\subset\mathbb{R}^d \text{ open},$$

in which case

$$\mathcal{D}=W_0^{1,2}(\Omega) \text{ and } \mathcal{E}[u,v]=\int_\Omega(\nabla u|\nabla v)dx.$$

Now we turn to an important notion generalizing the measure $(\nabla u|\nabla v)dx$ appearing above.
In fact, every strongly local, regular Dirichlet form $\mathcal{E}$ can be represented in the form

$$\mathcal{E}[u,v]=\int_X d\Gamma(u,v)$$

where Γ is a nonnegative sesquilinear mapping from $\mathcal{D}\times\mathcal{D}$ to the set of $\mathbb{K}$-valued Radon measures on X. It is determined by

$$\int_X \phi\, d\Gamma(u,u)=\mathcal{E}[u,\phi u]-\frac{1}{2}\mathcal{E}[u^2,\phi]$$

for realvalued $u\in\mathcal{D}$, $\phi\in\mathcal{D}\cap C_c(X)$ and called *energy measure*; see also [13]. We discuss properties of the energy measure next (see e.g. [13, 22, 47]). The energy measure satisfies the Leibniz rule,

$$d\Gamma(u\cdot v,w)=ud\Gamma(v,w)+vd\Gamma(u,w),$$

as well as the chain rule

$$d\Gamma(\eta(u),w)=\eta'(u)d\Gamma(u,w).$$

One can even insert functions from $\mathcal{D}_{\mathrm{loc}}$ into $d\Gamma$, where $\mathcal{D}_{\mathrm{loc}}$ is the set

$$\{u\in L^2_{\mathrm{loc}} \mid \text{for all compact } K\subset X \text{ there is } \phi\in\mathcal{D} \text{ s.t. } \phi=u \text{ } m\text{-a.e. on } K\},$$

as is readily seen from the following important property of the energy measure, strong locality:
Let U be an open set in X on which the function $\eta\in\mathcal{D}_{\mathrm{loc}}$ is constant, then

$$\chi_U d\Gamma(\eta,u)=0,$$

for any $u\in\mathcal{D}$. This, in turn, is a consequence of the strong locality of $\mathcal{E}$ and in fact equivalent to the validity of the Leibniz rule.
We write $d\Gamma(u):=d\Gamma(u,u)$ and note that the energy measure satisfies the Cauchy-Schwarz inequality:

$$\begin{aligned}\int_X |fg|d|\Gamma(u,v)| &\le \left(\int_X |f|^2 d\Gamma(u)\right)^{\frac{1}{2}}\left(\int_X |g|^2 d\Gamma(v)\right)^{\frac{1}{2}}\\ &\le \frac{1}{2}\int_X |f|^2 d\Gamma(u)+\frac{1}{2}\int_X |g|^2 d\Gamma(v).\end{aligned}$$

In order to introduce weak solutions on open subsets of X, we extend $\mathcal{E}$ and $\nu[\cdot,\cdot]$ to $\mathcal{D}_{\rm loc}(U) \times \mathcal{D}_c(U)$: where,

$$\mathcal{D}_{\rm loc}(U) := \{u \in L^2_{\rm loc}(U) \mid \forall \text{compact } K \subset U \exists\, \phi \in \mathcal{D} \text{ s.t. } \phi = u \ m\text{-a.e. on } K\}$$

$$\mathcal{D}_c(U) := \{\varphi \in \mathcal{D} | \operatorname{supp} \varphi \text{ compact in } U\}.$$

For $u \in \mathcal{D}_{\rm loc}(U), \varphi \in \mathcal{D}_c(U)$ we define

$$\mathcal{E}[u,\varphi] := \mathcal{E}[\eta u, \varphi],$$

where $\eta \in \mathcal{D} \cap C_c(U)$ is arbitrary with constant value 1 on the support of φ. This makes sense as the RHS does not depend on the particular choice of η by strong locality. In the same way, we can extend $\nu[\cdot,\cdot]$, using that every $u \in \mathcal{D}_{\rm loc}(U)$ admits a quasi continuous version $\tilde{u}$. Moreover, also Γ extends to a mapping $\Gamma : \mathcal{D}_{\rm loc}(U) \times \mathcal{D}_{\rm loc}(U) \to \mathcal{M}_R(U)$.

For completeness reasons we explicitly state the following lemma.

Lemma 1.2. (a) *Let $\Psi \in \mathcal{D}_{loc} \cap L^\infty_{loc}(X)$ and $\varphi \in \mathcal{D} \cap L^\infty_c(X)$ be given. Then, $\varphi\Psi$ belongs to $\mathcal{D}$.*

(b) *Let $\Psi \in \mathcal{D}_{loc}$ and $\varphi \in \mathcal{D} \cap L^\infty_c(X)$ be such that $d\Gamma(\varphi) \le C \cdot dm$. Then, $\varphi\Psi$ belongs to $\mathcal{D}$.*

Proof. Let K be the support of φ and V an open neighborhood of K.
(a) Locality and the Leibniz rule give

$$\int d\Gamma(\varphi\Psi) = \int_K |\varphi|^2 d\Gamma(\Psi) + 2\int_K \varphi\Psi d\Gamma(\varphi,\Psi) + \int_K |\Psi|^2 d\Gamma(\varphi).$$

Obviously, the first and the last term are finite and the middle one can be estimated by Cauchy Schwarz inequality. Putting this together, we infer $\int d\Gamma(\varphi\Psi) < \infty$.
(b) Clearly, it suffices to treat the case $\Psi \ge 0$. Since $\Psi_n := \Psi \wedge n$ is a normal contraction of Ψ for every $n \in \mathbb{N}$ it follows that $d\Gamma(\Psi_n) \le d\Gamma(\Psi)$. By part (a) we know that $\varphi\Psi_n \in \mathcal{D}$ and an estimate as above gives that

$$\begin{aligned}
\mathcal{E}(\varphi\Psi_n) &= \int_X d\Gamma(\varphi\Psi_n) \\
&\le 2\left(\int_X \varphi^2 d\Gamma(\Psi_n) + \int_X \Psi_n^2 d\Gamma(\varphi)\right) \\
&\le 2\left(\int_X \varphi^2 d\Gamma(\Psi) + C\int_X \chi_V \Psi^2 dm\right),
\end{aligned}$$

is bounded independently of $n \in \mathbb{N}$. As $\varphi\Psi_n$ converge to $\varphi\Psi$ in $L^2(X,m)$, an appeal to the Fatou type lemma for closed forms, [31], Lemma 2.12., p. 21 gives the assertion. □

We close this section by noting that both $\mathcal{D} \cap C_c(X)$ and $\mathcal{D} \cap L^\infty_c(X)$ are closed under multiplication (due to Leibniz rule).

The intrinsic metric. Using the energy measure one can define the *intrinsic metric* ρ by

$$\rho(x,y) = \sup\{|u(x) - u(y)| \; |u \in \mathcal{D}_{\mathrm{loc}} \cap C(X) \text{ and } d\Gamma(u) \leq dm\}$$

where the latter condition signifies that $\Gamma(u)$ is absolutely continuous with respect to m and the Radon-Nikodym derivative is bounded by 1 on X. Note that, in general, ρ need not be a metric. We say that $\mathcal{E}$ is *strictly local* if ρ is a metric that induces the original topology on X. Note that this implies that X is connected, since otherwise points in x, y in different connected components would give $\rho(x,y) = \infty$, as characteristic functions of connected components are continuous and have vanishing energy measure. We denote the intrinsic balls by

$$B(x,r) := \{y \in X | \rho(x,y) \leq r\}.$$

An important consequence of the latter assumption is that the distance function $\rho_x(\cdot) := \rho(x,\cdot)$ itself is a function in $\mathcal{D}_{\mathrm{loc}}$ with $d\Gamma(\rho_x) \leq dm$, see [47]. This easily extends to the fact that for every closed $E \subset X$ the function $\rho_E(x) := \inf\{\rho(x,y) | y \in E\}$ enjoys the same properties (see the Appendix of [14]). This has a very important consequence. Whenever $\zeta : \mathbb{R} \longrightarrow \mathbb{R}$ is continuously differentiable, and $\eta := \zeta \circ \rho_E$, then η belongs to $\mathcal{D}_{\mathrm{loc}}$ and satisfies

$$d\Gamma(\eta) = (\zeta' \circ \rho_E)^2 d\Gamma(\rho_E) \leq (\zeta' \circ \rho_E)^2 dm. \tag{1}$$

For this reason a lot of good cut-off functions are around in our context. More explicitly we note the following lemma (see [14] as well).

Lemma 1.3. *For any compact K in X there exists a $\varphi \in C_c(X) \cap \mathcal{D}$ with $\varphi \equiv 1$ on K, $\varphi \geq 0$ and $d\Gamma(\varphi) \leq C\, dm$ for some $C > 0$. If L is another compact set containing K in its interior, then φ can be chosen to have support in L.*

Proof. Let $r > 0$ be the positive distance of K to the complement of L. Choose a two times differentiable $\zeta : \mathbb{R} \to [0, \infty)$ with $\zeta(0) = 1$ and support contained in $(-\infty, r)$. Then, $\zeta \circ \rho_K$ does the job by (1). $\square$

Irreducibility. We will now discuss a notion that will be crucial in the proof of the existence of positive weak solutions below the spectrum. In what follows, $\mathfrak{h}$ will denote a densely defined, closed semibounded form in $L^2(X)$ with domain $D(\mathfrak{h})$ and positivity preserving semigroup $(T_t; t \geq 0)$. We denote by H the associated operator. Actually, the cases of interest in this paper are $\mathfrak{h} = \mathcal{E}$ or $\mathfrak{h} = \mathcal{E} + \nu$ with $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. We refer to [40], XIII.12 and a forthcoming paper [30] for details. We say that $\mathfrak{h}$ is *reducible*, if there is a measurable set $M \subset X$ such that M and its complement M^c are nontrivial (have positive measure) and $L^2(M)$ is a reducing subspace for M, i.e., $\mathbb{1}_M D(\mathfrak{h}) \subset D(\mathfrak{h})$, $\mathfrak{h}$ restricted to $\mathbb{1}_M D(\mathfrak{h})$ is a closed form and $\mathcal{E}(u,v) = \mathcal{E}(u\mathbb{1}_M, v\mathbb{1}_M) + \mathcal{E}(u\mathbb{1}_{M^c}, v\mathbb{1}_{M^c})$ for all u, v. If there is no such decomposition of $\mathfrak{h}$, the latter form is called *irreducible*. Note that reducibility can be rephrased in terms of the semigroup and the resolvent:

Theorem 1.4. *Let $\mathfrak{h}$ be as above. Then the following conditions are equivalent:*

- $\mathfrak{h}$ *is irreducible.*
- T_t *is positivity improving, for every* $t > 0$*, i.e.* $f \geq 0$ *and* $f \neq 0$ *implies that* $T_t f > 0$ *a.e.*
- $(H + E)^{-1}$ *is positivity improving for every* $E < \inf \sigma(H)$*.*

In [30] we will show that for a strictly local Dirichlet form $\mathcal{E}$ as above and a measure perturbation $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$, irreducibility of $\mathcal{E}$ implies irreducibility of $\mathcal{E} + \nu$.

2. Positive weak solutions and the associated transformation

Throughout this section we consider a strongly local, regular Dirichlet form, $(\mathcal{E}, \mathcal{D})$ on X and denote by $\Gamma : \mathcal{D}_{\text{loc}} \times \mathcal{D}_{\text{loc}} \to \mathcal{M}(X)$ the associated energy measure. We will be concerned with weak solutions Φ of the equation

$$(H_0 + V)\Phi = E \cdot \Phi, \tag{2}$$

where H_0 is the operator associated with $\mathcal{E}$ and V is a realvalued, locally integrable potential. In fact, we will consider a somewhat more general framework, allowing for measures instead of functions, as presented in the previous section. Moreover, we stress the fact that (2) is formal in the sense that Φ is not assumed to be in the operator domain of neither H_0 nor V. Here are the details.

Definition 2.1. *Let* $U \subset X$ *be open and* $\nu \in \mathcal{M}_{R,0}(U)$ *be a signed Radon measure on* U *that charges no set of capacity zero. Let* $E \in \mathbb{R}$ *and* $\Phi \in L^2_{loc}(U)$*. We say that* Φ *is a* weak supersolution *of* $(H_0 + \nu)\Phi = E \cdot \Phi$ *in* U *if:*

(i) $\Phi \in \mathcal{D}_{loc}(U)$,
(ii) $\tilde{\Phi} d\nu \in \mathcal{M}_R(U)$,
(iii) $\forall \varphi \in \mathcal{D} \cap C_c(U), \varphi \geq 0 :$

$$\mathcal{E}[\Phi, \varphi] + \int_U \varphi \tilde{\Phi} d\nu \geq E \cdot (\Phi|\varphi).$$

We call Φ *a* weak solution *of* $(H_0 + \nu)\Phi = E \cdot \Phi$ *in* U *if equality holds in (iii) above (which extends to all* $\varphi \in \mathcal{D} \cap C_c(U)$*). If* $V \in L^1_{loc}(U)$ *we say that* Φ *is a* weak (super-)solution *of* $(H_0 + V)\Phi = E \cdot \Phi$ *in* U *if it is a weak (super-) solution of* $(H_0 + \nu)\Phi = E \cdot \Phi$ *for* $\nu = V dm$*.*

Remark 2.2. (1) *If* $\nu = V dm$ *and* $V \in L^2_{loc}(U)$*, then property (ii) of the Definition above is satisfied.*
(2) *If* $\Phi \in L^\infty_{loc}(U)$ *and* $\nu \in \mathcal{M}_R(U)$ *then (ii) of the Definition above is satisfied.*
(3) *If* $\nu \in \mathcal{M}_R(U)$ *satisfies (ii) above then* $\nu - E dm \in \mathcal{M}_R(U)$ *satisfies (ii) as well and any weak solution of* $(H_0 + \nu)\Phi = E \cdot \Phi$ *in* U *is a weak solution of* $(H_0 + \nu - E dm)\Phi = 0$ *in* U*. Thus it suffices to consider the case* $E = 0$*.*
(4) *If* Φ *is a weak solution on* U*, then*

$$\mathcal{E}[\Phi, \varphi] + \int_U \varphi \tilde{\Phi} d\nu = E \cdot (\Phi|\varphi).$$

for all $\varphi \in \mathcal{D} \cap L^\infty_c(U)$. This follows easily from (b) of the approximation Lemma 1.1. (Note that we can indeed approximate within U by first choosing an appropriate η with compact support in U according to Lemma 1.3.)

We will deal with function $\Phi \in \mathcal{D}_{\rm loc}$ with $\Phi > 0$. If Φ is such a function and $\Phi^{-1} \in L^\infty_{\rm loc}$, we can use the chain rule and suitable smoothed version of the function $x \mapsto 1/x$ to conclude that Φ^{-1} must belong to $\mathcal{D}_{\rm loc}$ as well. This will be used various times in the sequel.

Here comes the first half of the Allegretto-Piepenbrink Theorem in a general form.

Theorem 2.3. *Let $(\mathcal{E}, \mathcal{D})$ be a regular, strictly local Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0}(U)$. Suppose that Φ is a weak solution of $(H_0 + \nu)\Phi = E \cdot \Phi$ in U with $\Phi > 0$ m-a.e. and $\Phi, \Phi^{-1} \in L^\infty_{loc}(U)$. Then, for all $\varphi, \psi \in \mathcal{D} \cap L^\infty_c(U)$:*

$$\mathcal{E}[\varphi, \psi] + \nu[\varphi, \psi] = \int_U \Phi^2 d\Gamma(\varphi\Phi^{-1}, \psi\Phi^{-1}) + E \cdot (\varphi|\psi).$$

In particular, $\mathcal{E} + \nu \geq E$ if furthermore $U = X$.

Proof. The "in particular" is clear as the desired inequality holds on $\mathcal{D} \cap C_c(X)$ and the form is regular by Lemma 1.1.
For the rest of the proof we may assume $E = 0$ without restriction, in view of the preceding remark. Without loss of generality we may also assume that φ and ψ are real valued functions. We now evaluate the RHS of the above equation, using the following identity. The Leibniz rule implies that for arbitrary $w \in \mathcal{D}_{\rm loc}(U)$:

$$0 = d\Gamma(w, 1) = d\Gamma(w, \Phi\Phi^{-1}) = \Phi^{-1} d\Gamma(w, \Phi) + \Phi d\Gamma(w, \Phi^{-1}) \qquad (\bigstar)$$

Therefore, for $\varphi, \psi \in \mathcal{D} \cap C_c(X)$:

$$\begin{aligned}
\int_X \Phi^2 d\Gamma(\varphi\Phi^{-1}, \psi\Phi^{-1}) &= \int_X \Phi d\Gamma(\varphi, \psi\Phi^{-1}) + \int_X \Phi^2 \varphi d\Gamma(\Phi^{-1}, \psi\Phi^{-1}) \\
\text{(by symmetry)} \quad &= \int_X d\Gamma(\varphi, \psi) + \int_X \Phi\psi d\Gamma(\varphi, \Phi^{-1}) \\
&\quad + \int_X \Phi^2 \varphi d\Gamma(\psi\Phi^{-1}, \Phi^{-1}) \\
&= \mathcal{E}[\varphi, \psi] + \int_X \Phi^2 d\Gamma(\varphi\psi\Phi^{-1}, \Phi^{-1}) \\
\text{(by } (\bigstar)\text{)} \quad &= \mathcal{E}[\varphi, \psi] - \int_X d\Gamma(\varphi\psi\Phi^{-1}, \Phi) \\
&= \mathcal{E}[\varphi, \psi] - \mathcal{E}[\varphi\psi\Phi^{-1}, \Phi].
\end{aligned}$$

As Φ is a weak solution we can now use part (4) of the previous remark to continue the computation by

$$\begin{aligned} \ldots &= \mathcal{E}[\varphi,\psi] - \big(-\nu[\varphi\psi\Phi^{-1},\Phi]\big) \\ &= \mathcal{E}[\varphi,\psi] + \nu[\varphi,\psi]. \end{aligned}$$

This finishes the proof. □

We note a number of consequences of the preceding theorem. The first is rather a consequence of the proof, however:

Corollary 2.4. *Assume that there is a weak supersolution Φ of $(H_0+\nu)\Phi = E\cdot\Phi$ on X with $\Phi > 0$ m-a.e. and $\Phi, \Phi^{-1} \in L^\infty_{loc}(X)$. Then $\mathcal{E}+\nu \geq E$.*

For the *Proof* we can use the same calculation as in the proof of the Theorem with $\varphi = \psi$ and use the inequality instead of the equality at the end.

Remark 2.5.
(1) *We can allow for complex measures ν without problems. In the context of PT–symmetric operators there is recent interest in this type of Schrödinger operators, see* [8]
(2) *Instead of measures also certain distributions could be included. Cf* [25] *for such singular perturbations.*

We will extend Theorem 2.3 to all of $\varphi,\psi\in\mathcal{D}$. This is somewhat technical. The main part is done in the next three propositions. We will assume the situation (S):

(S) Let $(\mathcal{E},\mathcal{D})$ be a regular, strictly local Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0}-\mathcal{M}_{R,1}$. Suppose that Φ is a weak solution of $(H_0+\nu)\Phi = E\cdot\Phi$ in X with $\Phi>0$ m-a.e. and $\Phi,\Phi^{-1}\in L^\infty_{\mathrm{loc}}(X)$.

Proposition 2.6. *Assume (S). Let $u \in \mathcal{D}(\mathcal{E}+\nu)$ be given. Let (u_n) be a sequence in $\mathcal{D}(\mathcal{E}+\nu)\cap L^\infty_c(X)$ which converges to u with respect to $\|\cdot\|_{\mathcal{E}+\nu}$. Then, $\varphi u_n\Phi^{-1}$ and $\varphi u\Phi^{-1}$ belong to $\mathcal{D}(\mathcal{E}+\nu)$ and*

$$\|\varphi u_n\Phi^{-1} - \varphi u\Phi^{-1}\|_{\mathcal{E}+\nu} \to 0, n\to\infty$$

for any $\varphi\in\mathcal{D}\cap C_c(X)$ with $d\Gamma(\varphi)\leq C dm$ for some $C>0$. In particular, $u\Phi^{-1}$ belongs to $\mathcal{D}_{loc}$.

Proof. Without loss of generality we assume $E=0$.
As shown above $\varphi\Phi^{-1}$ belongs to $\mathcal{D}\cap L^\infty_c$. Hence, $\varphi u_n\Phi^{-1} = u_n(\varphi\Phi^{-1})$ is a product of functions in $\mathcal{D}\cap L^\infty_c$ and therefore belongs to $\mathcal{D}\cap L^\infty_c$ as well.
As $\varphi\Phi^{-1}$ belongs to L^∞, the sequence $\varphi u_n\Phi^{-1}$ converges to $\varphi u\Phi^{-1}$ in $L^2(X,m)$. It therefore suffices to show that $\varphi u_n\Phi^{-1}$ is a Cauchy sequence with respect to $\|\cdot\|_{\mathcal{E}+\nu}$.
As (u_n) is Cauchy with respect to $\|\cdot\|_{\mathcal{E}+\nu}$ and $\varphi\Phi^{-1}$ is bounded, convergence of the ν part is taken care of and it suffices to show that

$$\mathcal{E}(\varphi(u_n-u_m)\Phi^{-1}) \to 0, n,m\to\infty.$$

Let K be the compact support of φ. Let $c > 0$ be an upper bound for Φ^{-2} on K. Choose $n, m \in \mathbb{N}$ and set $v := u_n - u_m$. Then, we can calculate

$$\begin{aligned}
\mathcal{E}(\varphi v \Phi^{-1}) &= \int_K d\Gamma(\varphi v \Phi^{-1}) \\
&= \int_K \frac{1}{\Phi^2} \Phi^2 d\Gamma(\varphi v \Phi^{-1}) \\
&\leq c \int_K \Phi^2 d\Gamma(\varphi v \Phi^{-1}) \\
\text{(Previous theorem)} \qquad &= c(\mathcal{E}(\varphi v) + \nu(\varphi v)) \\
&= c(\mathcal{E}(\varphi(u_n - u_m)) + \nu(\varphi(u_n - u_m))).
\end{aligned}$$

Now, convergence of $\nu(\varphi(u_n - u_m))$ to 0 for $n, m \to \infty$ can easily be seen (with arguments as at the beginning of the proof). As for $\mathcal{E}(\varphi(u_n - u_m))$ we can use Leibniz rule and Cauchy-Schwarz and $d\Gamma(\varphi) \leq C\, dm$ to compute

$$\begin{aligned}
\mathcal{E}(\varphi(u_n - u_m)) &= \int_K d\Gamma(\varphi(u_n - u_m)) \\
&= \int \varphi^2 d\Gamma(u_n - u_m) + 2 \int \varphi(u_n - u_m) d\Gamma(\varphi, u_n - u_m) \\
&\quad + \int |u_n - u_m|^2 d\Gamma(\varphi) \\
&\leq 2\Big(\int \varphi^2 d\Gamma(u_n - u_m) + \int |u_n - u_m|^2 d\Gamma(\varphi)\Big) \\
&\leq 2\|\varphi\|^2 \mathcal{E}(u_n - u_m) + 2C \int |u_n - u_m|^2 dm.
\end{aligned}$$

This gives easily the desired convergence to zero and $(\varphi u_n \Phi^{-1})$ is a Cauchy sequence with respect to $\|\cdot\|_{\mathcal{E}+\nu}$.
We now turn to a proof of the last statement: By Lemma 1.3, for any compact K we can find a φ satisfying the assumptions of the proposition with $\varphi \equiv 1$ on K. Then, $\varphi u \Phi^{-1}$ belongs to $\mathcal{D}$ by the above argument and agrees with $u\Phi^{-1}$ on K be construction. □

Proposition 2.7. *Assume (S). Let $u \in \mathcal{D}(\mathcal{E} + \nu)$ be given. Let (u_n) be a sequence in $\mathcal{D}(\mathcal{E} + \nu) \cap L_c^\infty(X)$ which converges to u with respect to $\|\cdot\|_{\mathcal{E}+\nu}$. Then,*

$$\int \psi d\Gamma(u_n \Phi^{-1}) \to \int \psi d\Gamma(u \Phi^{-1})$$

for any $\psi \in L_c^\infty(X)$.

Proof. We start with an intermediate claim.

Claim. For any $\psi \in L^\infty(X)$ and $\varphi \in \mathcal{D} \cap C_c(X)$ with $d\Gamma(\varphi) \leq C\, dm$ for some $C > 0$, we have $\int \psi d\Gamma(\varphi u_n \Phi^{-1}) \to \int \psi d\Gamma(\varphi u \Phi^{-1})$.

Proof of the claim. By triangle inequality, the difference between the terms in question can be estimated by

$$|\int \psi d\Gamma(\varphi(u-u_n)\Phi^{-1}, \varphi u_n \Phi^{-1})| + |\int \psi d\Gamma(\varphi u \Phi^{-1}, \varphi(u-u_n)\Phi^{-1})|.$$

By Cauchy Schwarz inequality these terms can be estimated by

$$\|\psi\|_\infty \mathcal{E}((\varphi(u_n-u)\Phi^{-1})^{1/2}\mathcal{E}(\varphi u_n \Phi^{-1})^{1/2}$$

and

$$\|\psi\|_\infty \mathcal{E}((\varphi(u_n-u)\Phi^{-1})^{1/2}\mathcal{E}(\varphi u \Phi^{-1})^{1/2}.$$

The previous proposition gives that $\mathcal{E}(\varphi(u_n-u)\Phi^{-1}) \to 0$, $n \to \infty$ and the claim follows.

Let now $\psi \in L_c^\infty(X)$ be given. Let K be the compact support of ψ. We use Lemma 1.3 to find $\varphi \in C_c(X) \cap \mathcal{D}$ with $\varphi \equiv 1$ on K and $d\Gamma(\varphi) \le C\, dm.$ for some $C > 0$. Locality gives

$$\int \psi d\Gamma(u_n\Phi^{-1}) = \int \psi d\Gamma(\varphi u_n \Phi^{-1})$$

and

$$\int \psi d\Gamma(u\Phi^{-1}) = \int \psi d\Gamma(\varphi u \Phi^{-1})$$

and the proposition follows from the claim. □

Proposition 2.8. *Assume (S). Let $u \in \mathcal{D}(\mathcal{E}+\nu)$ be given. Let (u_n) be a sequence in $\mathcal{D}(\mathcal{E}+\nu) \cap L_c^\infty(X)$ which converges to u with respect to $\|\cdot\|_{\mathcal{E}+\nu}$. Then,*

$$\int \Phi^2 d\Gamma(u_n\Phi^{-1}) \to \int \Phi^2 d\Gamma(u\Phi^{-1})$$

for any $\psi \in L_c^\infty(X)$.

Proof. Without loss of generality we assume $E = 0$. We start with the following claim.

Claim. $\mathcal{E}(u) + \nu(u) \ge \int \Phi^2 d\Gamma(u\Phi^{-1})$.

Proof of claim. By convergence of u_n to u w.r.t. $\|\cdot\|_{\mathcal{E}+\nu}$ and the last theorem, we have

$$\mathcal{E}(u) + \nu(u) = \lim_{n\to\infty} \mathcal{E}(u_n) + \nu(u_n) = \lim_{n\to\infty} \int \Phi^2 d\Gamma(u_n\Phi^{-1}).$$

Now, the claim follows easily from the preceeding proposition.

We now note that for fixed $n \in \mathbb{N}$, the sequence $(u_m - u_n)_m$ converges to $u - u_n$ w.r.t. $\|\cdot\|_{\mathcal{E}+\nu}$. We can therefore apply the claim to $u - u_n$ instead of u. This gives

$$\mathcal{E}(u-u_n) + \nu(u-u_n) \ge \int \Phi^2 d\Gamma((u-u_n)\Phi^{-1}) \ge 0$$

for any $n \in \mathbb{N}$. As the left hand side converges to zero for $n \to \infty$, so does the right hand side.

Mimicking the argument given in the proof of the Claim of the previous proposition, we can now conclude the desired statement. □

Corollary 2.9. *Let $(\mathcal{E},\mathcal{D})$ be a regular, strictly local Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. Suppose that Φ is a weak solution of $(H_0+\nu)\Phi = E\cdot\Phi$ in X with $\Phi > 0$ m-a.e. and $\Phi, \Phi^{-1} \in L^\infty_{loc}(X)$. Then, for all $\varphi,\psi \in D(\mathcal{E}+\nu)$, the products $\varphi\Phi^{-1}, \psi\Phi^{-1}$ belong to $\mathcal{D}_{loc}$ and the formula*

$$\mathcal{E}[\varphi,\psi] + \nu[\varphi,\psi] = \int_X \Phi^2 d\Gamma(\varphi\Phi^{-1},\psi\Phi^{-1}) + E\cdot(\varphi|\psi) \tag{3}$$

holds.

Proof. Without loss of generality we assume $E = 0$. It suffices to consider $\varphi = \psi$. By Proposition 2.6, $\varphi\Phi^{-1}$ belongs to $\mathcal{D}_{\rm loc}$. According to Lemma 1.1, we can choose a sequence (φ_n) in $\mathcal{D}\cap L^\infty_c(X)$ converging to φ w.r.t. $\|\cdot\|_{\mathcal{E}+\nu}$. This convergence and the last theorem then give

$$\mathcal{E}(\varphi)+\nu(\varphi) = \lim_{n\to\infty}\mathcal{E}(\varphi_n)+\nu(\varphi_n) = \lim_{n\to\infty}\int \Phi^2 d\Gamma(\varphi_n\Phi^{-1}).$$

The previous proposition then yields the desired formula. □

3. The existence of positive weak solutions below the spectrum

As noted in the preceding section, we find that $H_0+\nu \geq E$ whenever $\mathcal{E}+\nu$ is closable and admits a positive weak solution of $(H_0+\nu)\Phi = E\Phi$. In this section we prove the converse under suitable conditions. We use an idea from [18] where the corresponding statement for ordinary Schrödinger operators on $\mathbb{R}^d$ can be found. A key property is related to the celebrated *Harnack inequality*.

Definition 3.1.

(1) We say that $H_0+\nu$ *satisfies a Harnack inequality* for $E\in\mathbb{R}$ if, for every relatively compact, connected open $X_0 \subset X$ there is a constant C such that all positive weak solutions Φ of $(H_0+\nu)\Phi = E\Phi$ on X_0 are locally bounded and satisfy

$$\mathrm{esssup}_{B(x,r)} u \leq C \mathrm{essinf}_{B(x,r)} u,$$

for every $B(x,r) \subset X_0$ where esssup and essinf denote the essential supremum and infimum.

(2) We say that $H_0+\nu$ satisfies the *Harnack principle* for $E\in\mathbb{R}$ if for every relatively compact, connected open subset U of X and every sequence $(\Phi_n)_{n\in\mathbb{N}}$ of nonnegative solutions of $(H_0+\nu)\Phi = E\cdot\Phi$ in U the following implication holds: If, for some measurable subset $A\subset U$ of positive measure

$$\sup_{n\in\mathbb{N}} \|\Phi_n 1_A\|_2 < \infty$$

then, for all compact $K \subset U$ also

$$\sup_{n\in\mathbb{N}} \|\Phi_n 1_K\|_2 < \infty.$$

(3) We say that $H_0+\nu$ satisfies the *uniform Harnack principle* if for every bounded intervall $I \subset \mathbb{R}$, every relatively compact, connected open subset U of X and every sequence $(\Phi_n)_{n\in\mathbb{N}}$ of nonnegative solutions of $(H_0+\nu)\Phi = E_n \cdot \Phi$ in U with $E_n \in I$ the following implication holds: If, for some measurable subset $A \subset U$ of positive measure

$$\sup_{n\in\mathbb{N}} \|\Phi_n 1\!\!1_A\|_2 < \infty$$

then, for all compact $K \subset U$ also

$$\sup_{n\in\mathbb{N}} \|\Phi_n 1\!\!1_K\|_2 < \infty.$$

Note that validity of a Harnack principle implies that a nonnegative weak solution Φ must vanish identically if it vanishes on a set of positive measure (as $\Phi_n := n\Phi$ has vanishing L^2 norm on the set of positive measure in question). Note also that validity of an Harnack inequality extends from balls to compact sets by a standard chain of balls argument. This easily shows that $H_0+\nu$ satisfies the Harnack principle for $E \in \mathbb{R}$ if it obeys a Harnack inequality for $E \in \mathbb{R}$. Therefore, many situations are known in which the Harnack principle is satisfied:

Remark 3.2. (1) *For $\nu \equiv 0$ and $E = 0$ a Harnack inequality holds, whenever $\mathcal{E}$ satisfies a Poincaré and a volume doubling property; cf* [12] *and the discussion there.*

(2) *The most general results for $H_0 = -\Delta$ in terms of the measures ν that are allowed seem to be found in* [24]*, which also contains a thorough discussion of the literature prior to 1999. A crucial condition concerning the measures involved is the Kato condition and the uniformity of the estimates from* [24] *immediately gives that the uniform Harnack principle is satisfied in that context. Of the enormous list of papers on Harnack's inequality, let us mention* [2, 10, 11, 17, 24, 26, 27, 33, 41, 42, 49, 50]

Apart from the Harnack principle there is a second property that will be important in the proof of existence of positive general eigensolutions at energies below the spectrum: We say that $\mathcal{E}$ satisfies the *local compactness property* if $D_0(U) := \overline{D \cap C_c(U)}^{\|\cdot\|_{\mathcal{E}}}$ is compactly embedded in $L^2(X)$ for every relatively compact open $U \subset X$. (In case of the classical Dirichlet form this follows from Rellich's Theorem on compactness of the embedding of Sobolev spaces in L^2.)

Theorem 3.3. *Let $(\mathcal{E}, \mathcal{D})$ be a regular, strictly local, irreducible Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. Suppose that $\mathcal{E}$ satisfies the local compactness property and X is noncompact. Then, if $E < \inf \sigma(H_0 + \nu)$ and $H_0+\nu$ satisfies the Harnack principle for E, there is an a.e. positive solution of $(H_0+\nu)\Phi = E\Phi$.*

Proof. Let $E < \inf \sigma(H_0 + \nu)$. Since X is noncompact, locally compact and σ-compact, it can be written as a countable union

$$X = \bigcup_{R\in\mathbb{N}} U_R,\ U_R \text{ open, relatively compact },\overline{U_R} \subset U_{R+1};$$

where the U_R can be chosen connected, as X is connected, see [30] for details. For $n \in \mathbb{N}$ let $g_n \in L^2(X)$ with $\operatorname{supp} g_n \subset X \setminus U_{n+2}$, $g_n \geq 0$ and $g_n \neq 0$. It follows that

$$\Phi_n := (H_0 + \nu + E)^{-1} g_n \geq 0$$

is nonzero and is a weak solution of $(H_0+\nu)\Phi = E\Phi$ on $X\backslash\operatorname{supp} g_n$, in particular on the connected open subset U_{n+2}. Since $(H_0+\nu+E)^{-1}$ is positivity improving, it follows that $\|\Phi_n 1\!\!1_{U_1}\|_2 > 0$. By multiplying with a positive constant we may and will assume that $\|\Phi_n 1\!\!1_{U_1}\|_2 = 1$ for all $n \in \mathbb{N}$. We want to construct a suitably convergent subsequence of $(\Phi_n)_{n\in\mathbb{N}}$ so that the corresponding limit Φ is a positive weak solution.

First note that by the Harnack principle, for fixed $R \in \mathbb{N}$ and $n \geq R$ we know that

$$\sup_{n\in\mathbb{N}} \|\Phi_n 1\!\!1_{U_R}\|_2 < \infty,$$

since all the corresponding Φ_n are nonnegative solutions on U_{R+2}. In particular, $(\Phi_n 1\!\!1_{U_R})_{n\in\mathbb{N}}$ is bounded in $L^2(X)$ and so has a weakly convergent subsequence. By a standard diagonal argument, we find a subsequence, again denoted by $(\Phi_n)_{n\in\mathbb{N}}$, so that $\Phi_n 1\!\!1_{U_R} \to \Psi_R$ weakly in $L^2(X)$ for all $R \in \mathbb{N}$ and suitable Ψ_R. As multiplication with $1\!\!1_{U_R}$ is continuous and hence also weak-weak continuous, there is $\Phi \in L^2_{\mathrm{loc}}(X)$ such that $\Psi_R = \Phi 1\!\!1_{U_R}$. We will now perform some bootstrapping to show that the convergence is, in fact, much better than just local weak convergence in L^2 which will imply that Φ is the desired weak solution.

Since for fixed $R > 0$ and $n \geq R$ the Φ_n are nonnegative solutions on U_{R+2} the Caccioppoli inequality, cf [14] implies that

$$\int_{U_R} d\Gamma(\Phi_n) \leq C \int_{U_{R+1}} \Phi_n^2 dm$$

is uniformly bounded w.r.t. $n \in \mathbb{N}$. Combined with Leibniz rule and Cauchy Schwarz inequality this directly gives that $\int_{U_R} d\Gamma(\psi\Phi_n)$ is uniformly bounded w.r.t. $n \in \mathbb{N}$ for every $\psi \in \mathcal{D}$ with $d\Gamma(\psi) \leq dm$ (see [14] as well). Therefore, by Lemma 1.3, we can find for suitable cut-off functions $\eta_R \in \mathcal{D} \cap C_c(X)$ with $1\!\!1_{U_R} \leq \eta_R \leq 1\!\!1_{U_{R+1}}$ such that the sequence $(\eta_R \Phi_n)$ is bounded in $(D, \|\cdot\|_{\mathcal{E}})$.

The local compactness property implies that $(\eta_R\Phi_n)$ has an L^2-convergent subsequence. Using a diagonal argument again, we see that there is a common subsequence, again denoted by $(\Phi_n)_{n\in\mathbb{N}}$, such that

$$\Phi_n 1\!\!1_{U_R} \to \Phi 1\!\!1_{U_R} \text{ in } L^2(X) \text{ as } n \to \infty$$

for all $R \in \mathbb{N}$.

As a first important consequence we note that $\Phi \neq 0$, since $\|\Phi 1\!\!1_{U_1}\|_2 = \lim_n \|\Phi_n 1\!\!1_{U_1}\|_2 = 1$.

Another appeal to the Caccioppoli inequality gives that

$$\int_{U_R} d\Gamma(\Phi_n - \Phi_k) \leq C \int_{U_{R+1}} (\Phi_n - \Phi_k)^2 dm \to 0 \text{ as } n, k \to \infty.$$

Therefore, by the same reasoning as above, for every $R \in \mathbb{N}$ the sequence $(\eta_R \Phi_n)$ converges in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$. Since this convergence is stronger than weak convergence in $L^2(X)$, its limit must be $\eta_R\Phi$, so that the latter is in $\mathcal{D}$. We have thus proven that $\Phi \in \mathcal{D}_{\text{loc}}(X)$. Moreover, we also find that

$$\mathcal{E}[\Phi_n, \varphi] \to \mathcal{E}[\Phi, \varphi] \text{ for all } \varphi \in \mathcal{D} \cap C_c(X),$$

(since, by strong locality, for every cut-off function $\eta \in \mathcal{D} \cap C_c(X)$ that is 1 on $\operatorname{supp}\varphi$, we get

$$\mathcal{E}[\Phi_n, \varphi] = \mathcal{E}[\eta\Phi_n, \varphi] \to \mathcal{E}[\eta\Phi, \varphi] = \mathcal{E}[\Phi, \varphi].)$$

We will now deduce convergence of the potential term. This will be done in two steps. In the first step we infer convergence of the ν_- part from convergence w.r.t. $\|\cdot\|_{\mathcal{E}}$ and the relative boundedness of ν_-. In the second step, we use the fact that Φ is a weak solution to reduce convergence of the ν_+ part to convergence w.r.t. $\|\cdot\|_{\mathcal{E}}$ and convergence of the ν_- part. Here are the details: Consider cut-off functions η_R for $R \in \mathbb{N}$ as above. Due to convergence in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$, we know that there is a subsequence of $(\eta_R\Phi_n)_{n\in\mathbb{N}}$ that converges q.e., see [22] and the discussion in Section 1. One diagonal argument more will give a subsequence, again denoted by $(\Phi_n)_{n\in\mathbb{N}}$, such that the $\tilde{\Phi}_n$ converge to $\tilde{\Phi}$ q.e., where $\tilde{}$ denotes the quasi-continuous representatives. Since ν is absolutely continuous w.r.t capacity we now know that the $\tilde{\Phi}_n$ converge to $\tilde{\Phi}$ ν-a.e. Moreover, again due to convergence in $(\mathcal{D}, \|\cdot\|_{\mathcal{E}})$, we know that $(\eta_R\tilde{\Phi}_n)_{n\in\mathbb{N}}$ is convergent in $L^2(\nu_-)$ as $\nu_- \in \mathcal{M}_{R,1}$. Its limit must coincide with $\eta_R\tilde{\Phi}$, showing that $\tilde{\Phi}d\nu_- \in \mathcal{M}_R$.

We now want to show the analogous convergence for ν_+; we do so by approximation and omit the $\tilde{}$ for notational simplicity. By simple cut-off procedures, every $\varphi \in \mathcal{D}_c(X) \cap L^\infty(X)$ can be approximated w.r.t. $\|\cdot\|_{\mathcal{E}}$ by a uniformly bounded sequence of continuous functions in $\mathcal{D}$ with common compact support. Thus, the equation

$$\mathcal{E}[\Phi, \varphi] + \nu[\Phi, \varphi] = E \cdot (\Phi|\varphi),$$

initially valid for $\varphi \in \mathcal{D} \cap C_c(X)$ extends to $\varphi \in \mathcal{D}_c(X) \cap L^\infty(X)$ by continuity. Therefore, for arbitrary $k \in \mathbb{N}$, and $R < \min(n-2, m-2)$

$$\begin{aligned}
\int_{|\Phi_n-\Phi_m|\leq k} (\Phi_n - \Phi_m)^2 \eta_R d\nu_+ \quad &\leq \quad \int (\Phi_n - \Phi_m)\{(-k) \vee (\Phi_n - \Phi_m) \wedge k\}\eta_R d\nu_+ \\
&= \quad \nu_+[(\Phi_n - \Phi_m), \{(-k) \vee (\Phi_n - \Phi_m) \wedge k\}\eta_R] \\
&= \quad E((\Phi_n - \Phi_m)|\{\ldots\}\eta_R) + \nu_-[(\Phi_n - \Phi_m), \{\ldots\}\eta_R)] \\
&\quad\quad -\mathcal{E}[(\Phi_n - \Phi_m), \{(-k) \vee (\Phi_n - \Phi_m) \wedge k\}\eta_R]
\end{aligned}$$

By what we already know about convergence in $\mathcal{D}$, L^2 and $L^2(\nu_-)$, the RHS goes to zero as $n, m \to \infty$, independently of k. This gives the desired convergence of $\eta_R\tilde{\Phi}_n$, the limit being $\eta_R\tilde{\Phi}$ since this is the limit pointwise.

Finally, an appeal to the Harnack principle gives that Φ is positive a.e. on every U_R and, therefore, a.e. on X. □

Remark 3.4. *That we have to assume that X is noncompact can easily be seen by looking at the Laplacian on a compact manifold. In that situation any positive weak solution must in fact be in L^2 due to the Harnack principle. Thus the corresponding energy must lie in the spectrum. In fact, the corresponding energy must be the infimum of the spectrum as we will show in the next theorem. The theorem is standard. We include a proof for completeness reasons.*

Theorem 3.5. *Let $(\mathcal{E}, \mathcal{D})$ be a regular, strictly local, irreducible Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. Suppose that X is compact and $\mathcal{E}$ satisfies the local compactness property. Then, $H_0 + \nu$ has compact resolvent. In particular, there exists a positive weak solution to $(H_0 + \nu)\Phi = E_0\Phi$ for $E_0 := \inf \sigma(H_0 + \nu)$. This solution is unique (up to a factor) and belongs to $L^2(X)$. If $H_0 + \nu$ satisfies a Harnack principle, then E_0 is the only value in $\mathbb{R}$ allowing for a positive weak solution.*

Proof. As X is compact, the local compactness property gives that the operator associated to $\mathcal{E}$ has compact resolvent. In particular, the sequence of eigenvalues of H_0 is given by the minmax principle and tends to ∞. As ν_+ is a nonnegative operator and ν_- is form bounded with bound less than one, we can apply the minmax principle to $H_0 + \nu$ as well to obtain empty essential spectrum.
In particular, the infimum of the spectrum is an eigenvalue. By irreducibility and abstract principles, see e.g. [40], XIII.12, the corresponding eigenvector must have constant sign and if a Harnack principle holds then any other energy allowing for a positive weak solution must be an eigenvalue as well (as discussed in the previous remark). As there can not be two different eigenvalues with positive solutions, there can not be another energy with a positive weak solution. □

Combining the results for the compact and noncompact case we get:

Corollary 3.6. *Let $(\mathcal{E}, \mathcal{D})$ be a regular, strictly local, irreducible Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. Suppose that $\mathcal{E}$ satisfies the local compactness property and $H_0 + \nu$ satisfies the Harnack principle for all $E \in \mathbb{R}$. Then,*

$$\inf \sigma(H_0 + \nu) \le \sup\{E \in \mathbb{R} | \exists \text{ a.e. positive weak solution } (H_0 + \nu)\Phi = E\Phi\}.$$

This doesn't settle the existence of a positive weak solution for the groundstate energy $\inf \sigma(H_0 + \nu)$ in the noncompact case. The uniform Harnack principle settles this question:

Theorem 3.7. *Let $(\mathcal{E}, \mathcal{D})$ be a regular, strictly local, irreducible Dirichlet form, H_0 be the associated operator, $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. Suppose that $\mathcal{E}$ satisfies the local compactness property and $H_0 + \nu$ satisfies the uniform Harnack principle. Then there is an a.e. positive weak solution of $(H_0 + \nu)\Phi = E\Phi$ for $E = \inf \sigma(H_0 + \nu)$.*

Proof. It suffices to consider the case of noncompact X. Take a sequence (E_n) increasing to $E = \inf \sigma(H_0 + \nu)$. From Theorem 3.3 we know that there is an a.e. positive solution Ψ_n of $(H_0+\nu)\Phi = E_n\Phi$. We use the exhaustion $(U_R)_{R\in\mathbb{N}}$ from the proof of Theorem 3.3 and assume that

$$\|\Psi_n \mathbb{1}_{U_1}\|_2 = 1 \text{ for all } n \in \mathbb{N}.$$

As in the proof of Theorem 3.3 we can now show that we can pass to a subsequence such that $(\eta_R \Psi_n)$ converges in $\mathcal{D}$, $L^2(m)$ and $L^2(\nu_+ + \nu_-)$ for every $R \in \mathbb{N}$. The crucial point is that the uniform Harnack principle gives us a control on $\|\eta_R\Psi_n\|_2$, uniformly in n, due to the norming condition above. With aruments analogous to those in the proof of Theorem 3.3, the assertion follows. $\square$

Note that Corollaries 2.4 and 3.6 together almost give

$$\inf \sigma(H_0 + \nu) = \sup\{E \in \mathbb{R} | \exists \text{ a.e. positive weak solution } (H_0 + \nu)\Phi = E\Phi\}.$$

The only problem is that for the "$\geq$" from Corollary 2.4 we would have to replace a.e. positive by a.e. positive and $\Phi, \Phi^{-1} \in L^\infty_{\mathrm{loc}}$. This, however, is fulfilled whenever a Harnack inequality holds.

Corollary 3.8. *Let $(\mathcal{E}, \mathcal{D})$ be a regular, strictly local, irreducible Dirichlet form, H_0 be the associated operator and $\nu \in \mathcal{M}_{R,0} - \mathcal{M}_{R,1}$. Suppose that $\mathcal{E}$ satisfies the local compactness property and $H_0+\nu$ satisfies a Harnack inequality for all $E \in \mathbb{R}$. Then,*

$$\inf \sigma(H_0 + \nu) = \sup\{E \in \mathbb{R} | \exists \textit{ a.e. positive weak solution } (H_0 + \nu)\Phi = E\Phi\}.$$

4. Examples and applications

We discuss several different types of operators to which our results can be applied. Parts of the implications have been known before. However, previous proofs dealt with each of the mentioned operators separately, while we have a uniform argument of proof.

Examples. Classical examples of operators for which our results have been known before can be found in [5, 6, 7, 34, 35, 36, 18]. They concern Schrödinger operators and, more generally, symmetric elliptic second order differential operators on unbounded domains in $\mathbb{R}^d$, whose coefficients satisfy certain regularity conditions. For Laplace-Beltrami operators on Riemannian manifolds the Allegretto-Piepenbrink theorem has been stablished in [51].

Here we want to concentrate on two classes of examples which have attracted attention more recently: Hamiltonians with singular interactions and quantum graphs.

Hamiltonians with singular interactions. These are operators acting on $\mathbb{R}^d$ which may be formally written as $H = -\Delta - \alpha\delta(\cdot - M)$ where α is a positive real and $M \subset \mathbb{R}^d$ is a manifold of codimension one satisfying certain regularity conditions, see e.g. [15] or Appendix K of [3]. In fact, the delta interaction can be given a rigorous interpretation as a measure ν_M concentrated on the manifold M. More precisely, for any Borel set $B \subset \mathbb{R}^d$, one sets $\nu_M(B) := \mathrm{vol}_{d-1}(B \cap M)$ where vol_{d-1} denotes the $(d-1)$-dimensional Hausdorff measure on M. In [15], page 132, one can find suitable regularity conditions on M under which the measure ν_γ belongs to the class $\mathcal{M}_{R,1}$. Thus the singular interaction operator H falls into our general framework, cf. Remark 3.2.

If M is a C^2-regular, compact curve in $\mathbb{R}^2$ the essential spectrum of H equals $\sigma_{ess}(-\Delta) = [0, \infty)$, cf. [15]. On the other hand, the bottom of the spectrum of H is negative and consists consequently of an eigenvalue. This can be seen using the proof of Corollary 11 in [16]. In Section 3 of [21] it has been established that the ground state is nondegenerate and the corresponding eigenfunction strictly positive. This corresponds to part of our Theorem 3.3.

Quantum graphs. Quantum graphs are given in terms of a metric graph X and a Laplace (or more generally) Schrödinger operator H defined on the edges of X together with a set of (generalised) boundary conditions at the vertices which make H selfadjoint. To make sure that we are dealing with a strongly-local Dirichlet form we restrict ourselves here to the case of so called free or Kirchoff boundary conditions. A function in the domain of the corresponding quantum graph Laplacian H_0 is continuous at each vertex and the boundary values of the derivatives obtained by approaching the vertex along incident edges sum up to zero. Note that any non-negative Borel measure on X belongs to the class $\mathcal{M}_{R,0}(X)$. For $\nu_+ \in \mathcal{M}_{R,0}(X)$ and $\nu_- \in \mathcal{M}_{R,1}(X)$ the quantum graph operator $H = H_0 + \nu_+ - \nu_-$ falls into our framework.

See Section 5 of [14] for a more detailed discussion of the relation between Dirichlet forms and quantum graphs.

Applications. The ground state transformation which featured in Theorem 2.3 and Corollary 2.9 can be used to obtain a formula for the lowest spectral gap. To be more precise let us assume that $\mathcal{E}$, ν and Φ satisfy the conditions of Theorem 2.3 with $U = X$. Assume in addition that Φ is in $\mathcal{D}(\mathcal{E} + \nu)$. Then Φ is an eigenfunction of H corresponding to the eigenvalue $E = \min \sigma(H)$. We denote by

$$E' := \inf\{\mathcal{E}[u,u] + \nu[u,u] \mid u \in \mathcal{D}, \|u\| = 1, u \perp \Phi\}$$

the second lowest eigenvalue below the essential spectrum of H, or, if it does not exist, the bottom of $\sigma_{ess}(H)$. Then we obtain the following formula

$$E' - E = \inf_{\{u \in \mathcal{D}(\mathcal{E}+\nu), \|u\|=1, u \perp \Phi\}} \int_X \Phi^2 d\Gamma(u\Phi^{-1}, u\Phi^{-1}) \tag{4}$$

which determines the lowest spectral gap. It has been used in [28, 29, 52] to derive lower bounds on the distance between the two lowest eigenvalues of different classes of Schrödinger operators (see [44] for a related approach). In [28] bounded potentials are considered, in [29] singular interactions along curves in $\mathbb{R}^2$ are studied, and [52] generalises these results using a unified approach based on Kato-class measures.
If for a subset $U \subset X$ of positive measure and a function $u \in \{u \in \mathcal{D}, \|u\| = 1, u \perp \Phi\}$ the non-negative measure $\Gamma(u\Phi^{-1}, u\Phi^{-1})$ is absolutely continuous with respect to m, one can exploit formula (4) to derive the following estimate (cf. Section 3 in [52], and [28, 29] for similar bounds). Denote by $\gamma(u\Phi^{-1}) = \frac{d\Gamma(u\Phi^{-1}, u\Phi^{-1})}{dm}$ the Radon-Nykodim derivative. Then

$$\int_U \Phi^2 d\Gamma(u\Phi^{-1}, u\Phi^{-1}) \geq \frac{1}{m(U)} \inf_U \Phi^2 \left(\int_U \sqrt{\gamma(u\Phi^{-1})} dm \right)^2$$

In specific situations one can chose u to be an eigenfunction associated to the second eigenvalue E' and use geometric properties of Φ and u to derive explicit lower bounds on the spectral gap.

Other uses of the ground state transformation include the study of L^p-L^q mapping properties of the semigroup associated to $\mathcal{E}$ [20] and the proof of Lifschitz tails [32].

Acknowledgements. We would like to thank J. Brasche and N. Peyerimhoff for helpful discussions concerning the previous literature.

References

[1] S. Agmon. On positivity and decay of solutions of second order elliptic equations on Riemannian manifolds. In *Methods of functional analysis and theory of elliptic equations (Naples, 1982)*, pages 19–52. Liguori, Naples, 1983.

[2] M. Aizenman and B. Simon. Brownian motion and Harnack inequality for Schrödinger operators. *Commun. Pure Appl. Math.*, 35:209–273, 1982.

[3] S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and H. Holden. *Solvable models in quantum mechanics*. AMS Chelsea Publishing, Providence, RI, second edition, 2005. With an appendix by P. Exner.

[4] S. Albeverio and Z. M. Ma. Perturbation of Dirichlet forms—lower semiboundedness, closability, and form cores. *J. Funct. Anal.*, 99(2):332–356, 1991.

[5] W. Allegretto. On the equivalence of two types of oscillation for elliptic operators. *Pac. J. Math.*, 55:319–328, 1974.

[6] W. Allegretto. Spectral estimates and oscillation of singular differential operators. *Proc. Am. Math. Soc.*, 73:51, 1979.

[7] W. Allegretto. Positive solutions and spectral properties of second order elliptic operators. *Pac. J. Math.*, 92:15–25, 1981.

[8] C.M. Bender, S. Boettcher and P. Meisinger. PT symmetric quantum mechanics. *J.Math.Phys.*, 40:2201-2229, 1999.

[9] A. Beurling and J. Deny. Dirichlet spaces. *Proc. Nat. Acad. Sci. U.S.A.*, 45:208–215, 1959.

[10] M. Biroli. Schrödinger type and relaxed Dirichlet problems for the subelliptic p-Laplacian. *Potential Analysis*, 15, 1–16, 2001.

[11] M. Biroli and S. Marchi. Harnack inequality for the Schrödinger problem relative to strongly local Riemannian p-homogeneous forms with a potential in the Kato class. *Bound. Value Probl. 2007, Art. ID 24806, 19 pp.*

[12] M. Biroli and U. Mosco. A Saint-Venant type principle for Dirichlet forms on discontinuous media. *Ann. Mat. Pura Appl., IV. Ser.*, 169:125–181, 1995.

[13] N. Bouleau and F. Hirsch. *Dirichlet forms and analysis on Wiener space*, volume 14 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1991.

[14] A. Boutet de Monvel, D. Lenz, and P. Stollmann. Schnol's theorem for strongly local forms. *Israel J. Math.*, to appear, 2008.

[15] J. F. Brasche, P. Exner, Y. A. Kuperin, and P. Šeba. Schrödinger operators with singular interactions. *J. Math. Anal. Appl.*, 184(1):112–139, 1994.

[16] J. F. Brasche. On eigenvalues and eigensolutions of the Schrödinger equation on the complement of a set with classical capacity zero *Methods Funct. Anal. Topology*, 9(3), 189–206, 2003. http://www.math.chalmers.se/ brasche/ett.pdf

[17] F. Chiarenza, E. Fabes, and N. Garofalo. Harnack's inequality for Schrödinger operators and the continuity of solutions. *Proc. Amer. Math. Soc.*, 98(3):415–425, 1986.

[18] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon. *Schrödinger Operators with Application to Quantum Mechanics and Global Geometry.* Text and Monographs in Physics. Springer, Berlin, 1987.

[19] E. B. Davies: *Spectral theory and differential operators.* Cambridge University Press, Cambridge, 1995.

[20] E. B. Davies and B. Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.*, 59(2):335–395, 1984.

[21] P. Exner. An isoperimetric problem for leaky loops and related mean-chord inequalities. *J. Math. Phys.*, 46(6):062105, 2005. http://arxiv.org/abs/math-ph/0501066.

[22] M. Fukushima. *Dirichlet forms and Markov processes.* North-Holland Mathematical Library, Vol. 23. Amsterdam - Oxford -New York: North-Holland Publishing Company. Tokyo: Kodansha Ltd. X, 196 p., 1980.

[23] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet forms and symmetric Markov processes.* de Gruyter Studies in Mathematics. 19. Berlin: Walter de Gruyter. viii, 392 p., 1994.

[24] W. Hansen. Harnack inequalities for Schrödinger operators. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 28(3):413–470, 1999.

[25] I. W. Herbst and A. D. Sloan. Perturbation of translation invariant positivity preserving semigroups on $L^2(\mathbf{R}^N)$. *Trans. Amer. Math. Soc.*, 236:325–360, 1978.

[26] M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and N. Nadirashvili. Interior Hölder estimates for solutions of Schrödinger equations and the regularity of nodal sets. *Commun. Part. Diff. Eqns*, 20:1241–1273, 1995.

[27] M Kassmann. Harnack inequalities: an introduction. *Bound. Value Probl.* 2007, Art. ID 81415, 21 pp. 35-01

[28] W. Kirsch and B. Simon. Comparison theorems for the gap of Schrödinger operators. *J. Funct. Anal.*, 75:396–410, 1987.

[29] S. Kondej and I. Veselić. Lower bounds on the lowest spectral gap of singular potential Hamiltonians. *Ann. Henri Poincaré*, 8(1), 2006. http://arxiv.org/abs/math-ph/0605080.

[30] D. Lenz, P. Stollmann and I. Veselić. Irreducibility and connectedness for Dirichlet forms. *In preparation*, 2008

[31] Z.-M. Ma and M. Röckner. *Introduction to the theory of (non-symmetric) Dirichlet forms.* Universitext. Berlin: Springer-Verlag, 1992.

[32] G. A. Mezincescu. Lifschitz singularities for periodic operators plus random potentials. *J. Statist. Phys.*, 49(5-6):1181–1190, 1987.

[33] J. Moser. On harnacks theorem for elliptic differential equations. *Comm. Pure Appl. Math.*, 14:577–591, 1961.

[34] W.F. Moss and J. Piepenbrink. Positive solutions of elliptic equations. *Pacific J. Math.*, 75(1):219–226, 1978.

[35] J. Piepenbrink. Nonoscillatory elliptic equations. *J. Differential Equations*, 15:541–550, 1974.

[36] J. Piepenbrink. A conjecture of Glazman. *J. Differential Equations*, 24(2):173–177, 1977.

[37] Y. Pinchover. Topics in the theory of positive solutions of second-order elliptic and parabolic partial differential equations. In *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday*, volume 76 of *Proc. Sympos. Pure Math.*, pages 329–355. Amer. Math. Soc., Providence, RI, 2007.

[38] R.G. Pinsky. *Positive harmonic functions and diffusion*, volume 45 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1995.

[39] M. Reed and B. Simon. *Methods of Modern Mathematical Physics II, Fourier Analysis, Self-Adjointness.* Academic Press, San Diego, 1975.

[40] M. Reed and B. Simon. *Methods of Modern Mathematical Physics IV, Analysis of Operators.* Academic Press, San Diego, 1978.

[41] L. Saloff-Coste. Parabolic Harnack inequality for divergence-form second-order differential operators. *Potential Anal.*, 4(4):429–467, 1995. Potential theory and degenerate partial differential operators (Parma).

[42] J. Serrin. Local behavior of solutions of quasi-linear equations. *Acta Math.*, 111:247–302, 1964.

[43] B. Simon. Schrödinger semigroups. *Bull. Amer. Math. Soc.* 7 (3): 447–526 (1982).

[44] I. M. Singer, B. Wong, S.-T. Yau, and S. S.-T. Yau. An estimate of the gap of the first two eigenvalues in the Schrödinger operator. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 12(2):319–333, 1985.

[45] P. Stollmann. Smooth perturbations of regular Dirichlet forms. *Proc. Am. Math. Soc.*, 116(3):747–752, 1992.

[46] P. Stollmann and J. Voigt. Perturbation of Dirichlet forms by measures. *Potential Anal.*, 5(2):109–138, 1996.

[47] K.-T. Sturm. Analysis on local Dirichlet spaces. I: Recurrence, conservativeness and L^p-Liouville properties.

[48] K.-T. Sturm. Measures charging no polar sets and additive functionals of Brownian motion. *Forum Math.*, 4(3):257–297, 1992.

[49] K.-T. Sturm. Harnack's inequality for parabolic operators with singular low order terms. *Math. Z.*, 216(4):593–611, 1994.

[50] K.T. Sturm. Analysis on local Dirichlet spaces. III: The parabolic Harnack inequality. *J. Math. Pures Appl.* (9) 75 (1996), 273–297.

[51] D. Sullivan. Related aspects of positivity in Riemannian geometry. *J. Differential Geom.*, 25(3):327–351, 1987.

[52] H. Vogt. A lower bound on the first spectral gap of Schrödinger operators with Kato class measures. to appear: Ann. Henri Poincaré.

Daniel Lenz
Mathematisches Institut
Friedrich-Schiller
Universität Jena
Ernst-Abbé Platz 2
07743 Jena
Germany
daniel.lenz@uni-jena.de
http://www.analysis-lenz.uni-jena.de/

Peter Stollmann,
Fakultät für Mathematik
TU Chemnitz
09107 Chemnitz
Germany
stollman@mathematik.tu-chemnitz.de

Ivan Veselić
Emmy-Noether-Programme
of the DFG &
Fakultät für Mathematik
TU Chemnitz
09107 Chemnitz
Germany
http://www.tu-chemnitz.de/mathematik/enp/

Erratum for "On the Parity of Ranks of Selmer Groups III" cf. Documenta Math. 12 (2007), 243–274

Jan Nekovář

Received: March 31, 2008

Communicated by Ulf Rehmann

Abstract. In the manuscript "On the Parity of Ranks of Selmer Groups III" Documenta Math. 12 (2007), 243–274, [1], Remark 4.1.2(4) and the treatment of archimedean ε-factors in 4.1.3 are incorrect. Contrary to what is stated in 0.3, the individual archimedean ε-factors $\varepsilon_u(M)$ $(u \mid \infty)$ cannot be expressed, in general, in terms of $M_{\mathfrak{p}}$, but their product can.

2000 Mathematics Subject Classification: 11G40 11R23

To motivate the corrections below, consider a motive M (pure of weight w) over F with coefficients in L. Set $\widetilde{S}_\infty = \{\tau : F \hookrightarrow \mathbf{C}\}$, $\widetilde{S}_p = \{\sigma : F \hookrightarrow \overline{\mathbf{Q}}_p\}$ and denote by $r_\infty : \widetilde{S}_\infty \longrightarrow S_\infty$, $r_p : \widetilde{S}_p \longrightarrow S_p$ the canonical surjections. Fix an embedding $\iota : L \hookrightarrow \mathbf{C}$ and an isomorphism $\lambda : \overline{\mathbf{Q}}_p \xrightarrow{\sim} \mathbf{C}$ such that $\mathfrak{p}$ is induced by $\iota_p = \lambda^{-1} \circ \iota : L \hookrightarrow \overline{\mathbf{Q}}_p$. To each $v \in S_p$ then corresponds a subset

$$S_\infty(v) = \{r_\infty(\lambda \circ \sigma) \mid r_p(\sigma) = v\} \subset S_\infty$$

such that

$$\sum_{w \in S_\infty(v)} [L_w : \mathbf{R}] = [F_v : \mathbf{Q}_p].$$

For each $\tau \in \widetilde{S}_\infty$, the Betti realization $M_{B,\tau}$ is an L-vector space and there is a Hodge decomposition

$$M_{B,\tau} \otimes_{L,\iota} \mathbf{C} = \bigoplus_{i \in \mathbf{Z}} (\iota M_\tau)^{i,w-i}.$$

The corresponding Hodge numbers

$$h^{i,w-i}(\iota M_u) := h^{i,w-i}(\iota M_\tau) = \dim_{\mathbf{C}} (\iota M_\tau)^{i,w-i}$$

depend only on $u = r_\infty(\tau) \in S_\infty$. The de Rham realization M_{dR} is a free $L \otimes_{\mathbf{Q}} F$-module; its Hodge filtration is given by submodules $F^r M_{dR}$ (not necessarily free) which correspond, under the de Rham comparison isomorphism

$$M_{dR} \otimes_{L\otimes_{\mathbf{Q}} F, \iota\otimes\tau} \mathbf{C} \xrightarrow{\sim} M_{B,\tau} \otimes_{L,\iota} \mathbf{C},$$

to

$$(F^r M_{dR}) \otimes_{L\otimes_{\mathbf{Q}} F, \iota\otimes\tau} \mathbf{C} \xrightarrow{\sim} \bigoplus_{i\geq r} (\iota M_\tau)^{i,w-i},$$

hence

$$\dim_{\mathbf{C}} \left((gr_F^i M_{dR}) \otimes_{L\otimes_{\mathbf{Q}} F, \iota\otimes\tau} \mathbf{C}\right) = h^{i,w-i}(\iota M_\tau).$$

The $\mathfrak{p}$-adic realization $M_{\mathfrak{p}}$ of M is isomorphic, as an $L_{\mathfrak{p}}$-vector space, to $M_{B,\tau} \otimes_L L_{\mathfrak{p}}$ (for any $\tau \in \widetilde{S}_\infty$). For each $v \in S_p$, $D_{dR}(M_{\mathfrak{p},v})$ is a free $L_{\mathfrak{p}} \otimes_{\mathbf{Q}_p} F_v$-module equipped with a filtration satisfying

$$D^r_{dR}(M_{\mathfrak{p},v}) \xrightarrow{\sim} F^r M_{dR} \otimes_{L\otimes_{\mathbf{Q}} F} (L_{\mathfrak{p}} \otimes_{\mathbf{Q}_p} F_v).$$

This implies that, for each $i \in \mathbf{Z}$, the dimension

$$d_v^i(M_{\mathfrak{p}}) := \dim_{L_{\mathfrak{p}}} \left(D^i_{dR}(M_{\mathfrak{p},v}) / D^{i+1}_{dR}(M_{\mathfrak{p},v})\right)$$

is equal to

$$\begin{aligned}
&\dim_{L_{\mathfrak{p}}} \left(gr_F^i M_{dR}\right) \otimes_{L\otimes_{\mathbf{Q}} F} (L_{\mathfrak{p}} \otimes_{\mathbf{Q}_p} F_v) \\
&\qquad = \dim_{\overline{\mathbf{Q}}_p} \left(gr_F^i M_{dR}\right) \otimes_{L\otimes_{\mathbf{Q}} F, \iota_p\otimes \mathrm{incl}} (\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_v) \\
&\qquad = \sum_{\sigma : F_v \hookrightarrow \overline{\mathbf{Q}}_p} \dim_{\overline{\mathbf{Q}}_p} \left(gr_F^i M_{dR}\right) \otimes_{L\otimes_{\mathbf{Q}} F, \iota_p\otimes\sigma} \overline{\mathbf{Q}}_p \\
&\qquad = \sum_{u\in S_\infty(v)} [F_u : \mathbf{R}]\, h^{i,w-i}(\iota M_u),
\end{aligned}$$

hence

$$d_v^-(M_{\mathfrak{p}}) := \sum_{i<0} i\, d_v^i(M_{\mathfrak{p}}) = \sum_{u\in S_\infty(v)} [F_u : \mathbf{R}]\, d^-(\iota M_u),$$

$$d^-(\iota M_u) := \sum_{i<0} i\, h^{i,w-i}(\iota M_u). \qquad (\star)$$

Corrections to §4.1 and §5.1: firstly, 4.1.2(4) and 5.1.2(9) should be deleted. Secondly, §4.1.3 should be reformulated as follows: we assume that V satisfies 4.1.2(1)-(3). For each $v \in S_p$ we define

$$d_v^-(V) := \sum_{i<0} i\, d_v^i(V), \qquad d_v^i(V) = \dim_{L_\mathfrak{p}} \left(D_{dR}^i(V_v)/D_{dR}^{i+1}(V_v)\right) \tag{4.1.3.1'}$$

and

$$\prod_{u \in S_\infty(v)} \varepsilon(V_u) := (-1)^{d_v^-(V)} \prod_{u \in S_\infty(v), F_u = \mathbf{C}} (-1)^{\dim_{L_\mathfrak{p}}(V)/2} \tag{4.1.3.2'}$$

(even though we are unable to define the individual $\varepsilon(V_u)$). If $V = M_\mathfrak{p}$, where $M \xrightarrow{\sim} M^*(1)$ is pure (of weight -1), it follows from $(\star)$ and (2.3.1) that this definition gives the correct product of archimedean ε-factors.
The formula (4.1.3.6) should be replaced by

$$\forall v \in S_p \qquad \widetilde{\varepsilon}(V_v) = (-1)^{d_v^-(V)}(\det V_v^+)(-1) = \varepsilon(WD(V_v)^{N-ss}), \tag{4.1.3.6'}$$

which implies that

$$\widetilde{\varepsilon}(V_v) \prod_{u \in S_\infty(v)} \varepsilon(V_u) = (\det V_v^+)(-1) \prod_{u \in S_\infty(v), F_u = \mathbf{C}} (-1)^{\dim_{L_\mathfrak{p}}(V)/2},$$

hence

$$\prod_{v \in S_p \cup S_\infty} \widetilde{\varepsilon}(V_v) = (-1)^{r_2(F)\ \dim_{L_\mathfrak{p}}(V)/2} \prod_{v \in S_p} (\det V_v^+)(-1), \tag{4.1.3.7'}$$

where $r_2(F)$ denotes the number of complex places of F.
Corrections to Theorem 5.3.1 and its proof: the statement should say that, under the assumptions 5.1.2(1)-(8), the quantity

$$\begin{gathered}(-1)^{h_f^1(F,V)}/\varepsilon(V) = (-1)^{\widetilde{h}_f^1(F,V)}/\widetilde{\varepsilon}(V) = \\ = (-1)^{\widetilde{h}_f^1(F,V)}(-1)^{r_2(F)\ \dim_{\mathscr{L}}(\mathcal{V})/2} \prod_{v \in S_p} (\det \mathcal{V}_v^+)(-1) \prod_{v \notin S_p \cup S_\infty} \varepsilon(\mathcal{V}_v)\end{gathered}$$

depends only on $\mathcal{V}$ and $\mathcal{V}_v^+$ $(v \in S_p)$.
In the proof, a reference to (4.1.3.7) should be replaced by that to (4.1.3.7'), which yields

$$\widetilde{\varepsilon}(V) = \prod_{v \in S_p \cup S_\infty} \widetilde{\varepsilon}(V_v) \prod_{v \notin S_p \cup S_\infty} \varepsilon(V_v) = (-1)^{r_2(F)\ \dim_{L_\mathfrak{p}}(V)/2} \prod_{v \in S_p} (\det V_v^+)(-1) \prod_{v \notin S_p \cup S_\infty} \varepsilon(V_v).$$

CORRECTIONS TO §5.3.3: the first question should ask whether

$$(-1)^{d_v^-(V)}\varepsilon(WD(V_v)^{N-ss}) \qquad\qquad (v \in S_p)$$

depends only on $\mathcal{V}_v$?

CORRECTIONS TO §5.3.4-5: it is often useful to use a slightly more general version of Example 5.3.4 with $\Gamma = \Gamma_0 \times \Delta$, where Γ_0 is isomorphic to $\mathbf{Z}_p$ and Δ is finite (abelian). Given a character $\alpha : \Delta \longrightarrow \mathcal{O}_{\mathfrak{p}}^*$, set

$$R = \mathcal{O}_{\mathfrak{p}}[[\Gamma_0]], \qquad \mathcal{T} = \left(T \otimes_{\mathcal{O}_{\mathfrak{p}}} \mathcal{O}_{\mathfrak{p}}[[\Gamma^+]]\right) \otimes_{\mathcal{O}_{\mathfrak{p}}[\Delta],\alpha} \mathcal{O}_{\mathfrak{p}},$$

$$\mathcal{T}_v^+ = \left(T_v^+ \otimes_{\mathcal{O}_{\mathfrak{p}}} \mathcal{O}_{\mathfrak{p}}[[\Gamma^+]]\right) \otimes_{\mathcal{O}_{\mathfrak{p}}[\Delta],\alpha} \mathcal{O}_{\mathfrak{p}} \quad (v \in S_p).$$

As in 5.3.4(2)-(3), $\mathcal{T}$ is an $R[G_{F,S}]$-module equipped with a skew-symmetric R-bilinear pairing $(\ ,\ ) : \mathcal{T} \times \mathcal{T} \longrightarrow R(1)$ inducing an isomorphism

$$\mathcal{T} \otimes \mathbf{Q} \xrightarrow{\sim} \mathrm{Hom}_R(\mathcal{T}, R(1)) \otimes \mathbf{Q}.$$

In 5.3.4(5) we have to replace $\beta : \Gamma \longrightarrow L_{\mathfrak{p}}(\beta)$ by $\beta : \Gamma_0 \longrightarrow L_{\mathfrak{p}}(\beta)$; then

$$\mathcal{T}_P/\varpi_P\mathcal{T}_P = \mathrm{Ind}_{G_{F_0,S}}^{G_{F,S}}(V \otimes (\beta \times \alpha)).$$

In 5.3.5, we set, for any $L_{\mathfrak{p}}[\Gamma]$-module M,

$$M^{(\beta\times\alpha)} = \{x \in M \otimes_{L_{\mathfrak{p}}} L_{\mathfrak{p}}(\beta) \mid \forall\sigma \in \Gamma \quad \sigma(x) = (\beta \times \alpha)(x)\};$$

then

$$\begin{aligned} H_f^1(F, \mathcal{T}_P/\varpi_P\mathcal{T}_P) &= H_f^1(F_0, V \otimes (\beta \times \alpha)) \\ &= \left(H_f^1(F_\beta, V) \otimes (\beta \times \alpha)\right)^{\mathrm{Gal}(F_\beta/F_0)} = H_f^1(F_\beta, V)^{(\beta^{-1}\times\alpha^{-1})} \end{aligned}$$

and

$$\tau : H_f^1(F_\beta, V)^{(\beta^{-1}\times\alpha^{-1})} \xrightarrow{\sim} H_f^1(F_\beta, V)^{(\beta\times\alpha)}.$$

Applying Corollary 5.3.2, we obtain, for any pair of characters of finite order $\beta, \beta' : \Gamma_0 \longrightarrow \overline{L}_{\mathfrak{p}}^*$, that

$$\begin{aligned} &(-1)^{h_f^1(F_0, V\otimes(\beta\times\alpha))}/\,\varepsilon(F_0, V \otimes (\beta \times \alpha)) \\ &\qquad = (-1)^{h_f^1(F_0, V\otimes(\beta'\times\alpha))}/\,\varepsilon(F_0, V \otimes (\beta' \times \alpha)). \qquad (5.3.5.1') \end{aligned}$$

REFERENCES

[1] Jan Nekovář, On the Parity of Ranks of Selmer Groups III, Documenta Math. 12 (2007) 243–274

The Max-Plus Martin Boundary

Marianne Akian, Stéphane Gaubert, and Cormac Walsh

Received: June 3, 2005
Revised: May 17, 2009

Communicated by Alfred K. Louis

Abstract. We develop an idempotent version of probabilistic potential theory. The goal is to describe the set of max-plus harmonic functions, which give the stationary solutions of deterministic optimal control problems with additive reward. The analogue of the Martin compactification is seen to be a generalisation of the compactification of metric spaces using (generalised) Busemann functions. We define an analogue of the minimal Martin boundary and show that it can be identified with the set of limits of "almost-geodesics", and also the set of (normalised) harmonic functions that are extremal in the max-plus sense. Our main result is a max-plus analogue of the Martin representation theorem, which represents harmonic functions by measures supported on the minimal Martin boundary. We illustrate it by computing the eigenvectors of a class of Lax-Oleinik semigroups with nondifferentiable Lagrangian: we relate extremal eigenvector to Busemann points of normed spaces.

2000 Mathematics Subject Classification: Primary 31C35; Secondary 49L20, 47175.
Keywords and Phrases: Martin boundary, metric boundary, potential theory, Lax-Oleinik semigroup, weak KAM solutions, max-plus algebra, dynamic programming, deterministic optimal control, Markov decision process, eigenvalues, eigenvectors, Busemann functions, extremal generators.

This work was started during a post-doctoral stay of the third author at INRIA, supported by an ERCIM-INRIA fellowship

1 Introduction

There exists a correspondence between classical and idempotent analysis, which was brought to light by Maslov and his collaborators [Mas87, MS92, KM97, LMS01]. This correspondence transforms the heat equation to an Hamilton-Jacobi equation, and Markov operators to dynamic programming operators. So, it is natural to consider the analogues in idempotent analysis of harmonic functions, which are the solutions of the following equation

$$u_i = \sup_{j\in S}(A_{ij} + u_j) \qquad \text{for all } i \in S. \tag{1}$$

The set S and the map $A : S \times S \to \mathbb{R} \cup \{-\infty\},\ (i,j) \mapsto A_{ij}$, which plays the role of the Markov kernel, are given, and one looks for solutions $u : S \to \mathbb{R} \cup \{-\infty\},\ i \mapsto u_i$. This equation is the dynamic programming equation of a deterministic optimal control problem with infinite horizon. In this context, S is the set of states, the map A gives the weights or rewards obtained on passing from one state to another, and one is interested in finding infinite paths that maximise the sum of the rewards. Equation (1) is linear in the max-plus algebra, which is the set $\mathbb{R} \cup \{-\infty\}$ equipped with the operations of maximum and addition. The term idempotent analysis refers to the study of structures such as this, in which the first operation is idempotent.

In potential theory, one uses the Martin boundary to describe the set of harmonic and super-harmonic functions of a Markov process, and the final behaviour of its paths. Our goal here is to obtain analogous results for Equation (1).

The original setting for the Martin boundary was classical potential theory [Mar41], where it was used to describe the set of positive solutions of Laplace's equation. Doob [Doo59] gave a probabilistic interpretation in terms of Wiener processes and also an extension to the case when time is discrete. His method was to first establish an integral representation for super-harmonic functions and then to derive information about final behaviour of paths. Hunt [Hun60] showed that one could also take the opposite approach: establish the results concerning paths probabilistically and then deduce the integral representation. The approach taken in the present paper is closest to that of Dynkin [Dyn69], which contains a simplified version of Hunt's method.

There is a third approach to this subject, using Choquet theory. However, at present, the tools in the max-plus setting, are not yet sufficiently developed to allow us to take this route.

Our starting point is the max-plus analogue of the *Green kernel*,

$$A^*_{ij} := \sup\{A_{i_0 i_1} + \cdots + A_{i_{n-1} i_n} \mid n \in \mathbb{N},\ i_0, \ldots, i_n \in S,\ i_0 = i,\ i_n = j\} \ .$$

Thus, A^*_{ij} is the maximal weight of a path from i to j. We fix a map $i \mapsto \sigma_i$, from S to $\mathbb{R} \cup \{-\infty\}$, which will play the role of the *reference measure*. We set $\pi_j := \sup_{k\in S} \sigma_k + A^*_{kj}$. We define the *max-plus Martin space* $\mathscr{M}$ to be the closure of the set of maps $\mathscr{K} := \{A^*_{\cdot j} - \pi_j \mid j \in S\}$ in the product topology, and the *Martin boundary* to be $\mathscr{M} \setminus \mathscr{K}$. This term must be used with caution however, since $\mathscr{K}$ may not be open in $\mathscr{M}$ (see Example 10.6). The reference measure is often chosen to be a max-plus Dirac function, taking the value 0 at some *basepoint* $b \in S$ and the value $-\infty$ elsewhere. In this case, $\pi_j = A^*_{bj}$.

One may consider the analogue of an "almost sure" event to be a set of outcomes (in our case paths) for which the maximum reward over the complement is $-\infty$. So we are lead to the notion of an "almost-geodesic", a path of finite total reward, see Section 7. The almost sure convergence of paths in the probabilistic case then translates into the convergence of every almost-geodesic to a point on the boundary.

The spectral measure of probabilistic potential theory also has a natural analogue, and we use it to give a representation of the analogues of harmonic functions, the solutions of (1). Just as in probabilistic potential theory, one does not need the entire Martin boundary for this representation, a particular subset, called the *minimal Martin space*, will do. The probabilistic version is defined in [Dyn69] to be the set of boundary points for which the spectral measure is a Dirac measure located at the point itself. Our definition (see Section 4) is closer to an equivalent definition given in the same paper in which the spectral measure is required only to have a unit of mass at the point in question. The two definitions are not equivalent in the max-plus setting and this is related to the main difference between the two theories: the representing max-plus measure may not be unique.

Our main theorem (Theorem 8.1) is that every (max-plus) harmonic vector u that is integrable with respect to π, meaning that $\sup_{j\in S} \pi_j + u_j < \infty$, can be represented as

$$u = \sup_{w \in \mathscr{M}^m} \nu(w) + w, \tag{2}$$

where ν is an upper semicontinuous map from the minimal Martin space $\mathscr{M}^m$ to $\mathbb{R} \cup \{-\infty\}$, bounded above. The map ν is the analogue of the density of the spectral measure.

We also show that the (max-plus) minimal Martin space is exactly the set of (normalised) harmonic functions that are *extremal* in the max-plus sense, see Theorem 8.3. We show that each element of the minimal Martin space is either recurrent, or a boundary point which is the limit of an almost-geodesic (see Corollary 7.7 and Proposition 7.8).

To give a first application of our results, we obtain in Corollary 11.3 an existence theorem for non-zero harmonic functions of max-plus linear kernels satisfying a tightness condition, from which we derive a characterisation of the spectrum of some of these kernels (Corollary 11.4).

To give a second application, we obtain in Section 12 a representation of the eigenvectors of the Lax-Oleinik semigroup [Eva98, §3.3]:

$$T^t u(x) = \sup_{y \in \mathbb{R}^n} -tL\Big(\frac{y-x}{t}\Big) + u(y) \ ,$$

where L is a convex Lagrangian. This is the evolution semigroup of the Hamilton-Jacobi equation

$$\frac{\partial u}{\partial t} = L^\star(\nabla u) \ ,$$

where $L^\star$ denotes the Legendre-Fenchel transform of L. An eigenvector with eigenvalue $\lambda \in \mathbb{R}$ is a function u such that $T^t u = \lambda t + u$ holds for all $t > 0$. We compute the eigenvectors for a subclass of possibly nondifferentiable Lagrangians (Corollary 12.3 and Theorem 12.5).

Results and ideas related to the ones of present paper have appeared in several works: we now discuss them.

Max-plus harmonic functions have been much studied in the finite-dimensional setting. The representation formula (2) extends the representation of harmonic vectors given in the case when S is finite in terms of the *critical* and *saturation* graphs. This was obtained by several authors, including Romanovski [Rom67], Gondran and Minoux [GM77] and Cuninghame-Green [CG79, Th. 24.9]. The reader may consult [MS92, BCOQ92, Bap98, GM02, AG03, AGW05] for more background on max-plus spectral theory. Relations between max-plus spectral theory and infinite horizon optimisation are discussed by Yakovenko and Kontorer [YK92] and Kolokoltsov and Maslov [KM97, § 2.4]. The idea of "almost-geodesic" appears there in relation with "Turnpike" theorems.

The max-plus Martin boundary generalises to some extent the boundary of a metric space defined in terms of (generalised) Busemann functions by Gromov in [Gro81] in the following way (see also [BGS85] and [Bal95, Ch. II]). (Note that this is not the same as the Gromov boundary of hyperbolic spaces.) If (S, d) is a complete metric space, one considers, for all $y, x \in S$, the function $b_{y,x}$ given by

$$b_{y,x}(z) = d(x, z) - d(x, y) \quad \text{for } z \in S \ .$$

One can fix the *basepoint* y in an arbitrary way. The space $\mathscr{C}(S)$ can be equipped with the topology of uniform convergence on bounded sets, as in [Gro81, Bal95], or with the topology of uniform convergence on compact sets, as in [BGS85]. The limits of sequences of functions $b_{y,x_n} \in \mathscr{C}(S)$, where x_n is a sequence of elements of S going to infinity, are called (generalised) *Busemann functions*.

When the metric space S is proper, meaning that all closed bounded subsets of S are compact, the set of Busemann functions coincides with the max-plus Martin boundary obtained by taking $A_{zx} = A^*_{zx} = -d(z, x)$, and σ the max-plus Dirac function at the basepoint y. This follows from Ascoli's theorem, see Remark 7.10 for details. Note that our setting is more general since $-A^*$ need not have the properties of a metric, apart from the triangle inequality (the case when A^* is not symmetrical is needed in optimal control).

We note that Ballman has drawn attention in [Bal95, Ch. II] to the analogy between this boundary and the probabilistic Martin boundary.

The same boundary has recently appeared in the work of Rieffel [Rie02], who called it the *metric boundary*. Rieffel used the term *Busemann point* to designate those points of the metric boundary that are limits of what he calls "almost-geodesics". We shall see in Corollary 7.13 that these are exactly the points of the max-plus minimal Martin boundary, at least when S is a proper metric space. We also relate Busemann points to extremal eigenvectors of Lax-Oleinik semigroups, in Section 12. Rieffel asked in what cases are all boundary points Busemann points. This problem, as well as the relation between the metric boundary and other boundaries, has been studied by Webster and Winchester [WW06, WW05] and by Andreev [And04, And07]. However, representation problems like the one dealt with in Theorem 8.1 do not seem to have been treated in the metric space context.

Results similar to those of max-plus spectral theory have recently appeared in weak-KAM theory. In this context, S is a Riemannian manifold and the kernel A is replaced by a Lax-Oleinik semigroup, that is, the evolution semigroup of a Hamilton-Jacobi equation. Max-plus harmonic functions correspond to the *weak-KAM solutions* of Fathi [Fat97b, Fat97a, Fat08], which are essentially the eigenvectors of the Lax-Oleinik semigroup, or equivalently, the viscosity solutions of the ergodic Hamilton-Jacobi equation, see [Fat08, Chapter 8]. In weak-KAM theory, the analogue of the Green kernel is called the *Mañe potential*, the role of the critical graph is played by the *Mather set*, and the *Aubry set* is related to the saturation graph. In the case when the manifold is compact, Contreras [Con01, Theorem 0.2] and Fathi [Fat08, Theorem 8.6.1] gave a representation of the weak-KAM solutions, involving a supremum of fundamental solutions associated to elements of the Aubry set. The case of non-compact manifolds was considered by Contreras, who defined an analogue of the minimal max-plus Martin boundary in terms of Busemann functions, and obtained in [Con01, Theorem 0.5] a representation formula for weak-KAM solutions analogous to (2). Busemann functions also appear in [Fat03]. Other results of weak-KAM theory concerning non-compact manifolds have been obtained by Fathi and Maderna [FM02]. See also Fathi and Siconolfi [FS04]. Let us point out that some results of weak-KAM theory with a discrete flavor were established by MacKay, Slijepčević, and Stark [MSS00]. Extremality properties of the elements of the max-plus Martin boundary (Theorems 6.2 and 8.3 below) do not seem to have been considered in weak-KAM theory.

Despite the general analogy, the proofs of our representation theorem for harmonic functions (Theorem 8.1) and of the corresponding theorems in [Con01] and [Fat08] require different techniques. In order to relate both settings, it would be natural to set $A = B_s$, where $(B_t)_{t\geq 0}$ is the Lax-Oleinik semigroup, and $s > 0$ is arbitrary. However, only special kernels A can be written in this way, in particular A must have an "infinite divisibility" property. Also, not every harmonic function of B_s is a weak-KAM solution associated to the semigroup $(B_t)_{t\geq 0}$. Thus, the discrete time case is in some sense more general than the continuous-time case, but eigenvectors are more constrained in continuous time, so both settings require distinct treatments. Nevertheless, in some special cases, a representation of weak-KAM solutions follows from our results. This happens for example in Section 12, where our assumptions imply that the minimal Martin space of B_s is independent of s. We note that the Lagrangian there is not necessarily differentiable, a property which is required in [Fat08] and [Con01].

The lack of uniqueness of the representing measure is examined in a further work [Wal09], where it is shown that the set of (max-plus) measures representing a given (max-plus) harmonic function has a least element.

After the submission of the present paper, a boundary theory which has some similarities with the present one was developed by Ishii and Mitake [IM07]. The results there are in the setting of viscosity solutions and are independent of the present ones.

We note that the main results of the present paper have been announced in the final section of a companion paper, [AGW05], in which max-plus spectral theory was developed under some tightness conditions. Here, we use tightness only in Section 11. We finally note that the results of the present paper have been used in the further works [Wal07, Wal08].

Acknowledgements. We thank Albert Fathi for helpful comments, and in particular for having pointed out to us the work of Contreras [Con01]. We also thank Arnaud de la Fortelle for references on the probabilistic Martin boundary theory.

2 The max-plus Martin kernel and max-plus Martin space

To show the analogy between the boundary theory of deterministic optimal control problems and classical potential theory, it will be convenient to use max-plus notation. The *max-plus semiring*, $\mathbb{R}_{\max}$, is the set $\mathbb{R}\cup\{-\infty\}$ equipped with the addition $(a,b)\mapsto a\oplus b:=\max(a,b)$ and the multiplication $(a,b)\mapsto a\odot b:=a+b$. We denote by $\mathbb{0}:=-\infty$ and $\mathbb{1}:=0$ the zero and unit elements, respectively. We shall often write ab instead of $a\odot b$. Since the supremum of an infinite set may be infinite, we shall occasionally need to consider the *completed max-plus semiring* $\overline{\mathbb{R}}_{\max}$, obtained by adjoining to $\mathbb{R}_{\max}$ an element $+\infty$, with the convention that $\mathbb{0}=-\infty$ remains absorbing for the semiring multiplication.

The sums and products of matrices and vectors are defined in the natural way. These operators will be denoted by $\oplus$ and concatenation, respectively. For instance, if $A\in\overline{\mathbb{R}}_{\max}^{S\times S}$, $(i,j)\mapsto A_{ij}$, denotes a matrix (or kernel), and if $u\in\overline{\mathbb{R}}_{\max}^{S}$, $i\mapsto u_i$ denotes a vector, we denote by $Au\in\overline{\mathbb{R}}_{\max}^{S}$, $i\mapsto(Au)_i$, the vector defined by

$$(Au)_i:=\bigoplus_{j\in S}A_{ij}u_j \enspace ,$$

where the symbol $\oplus$ denotes the usual supremum.

We now introduce the max-plus analogue of the *potential kernel* (Green kernel).

Given any matrix $A \in \overline{\mathbb{R}}_{\max}^{S\times S}$, we define

$$A^* = I \oplus A \oplus A^2 \oplus \cdots \in \overline{\mathbb{R}}_{\max}^{S\times S} \ ,$$
$$A^+ = A \oplus A^2 \oplus A^3 \oplus \cdots \in \overline{\mathbb{R}}_{\max}^{S\times S}$$

where $I = A^0$ denotes the max-plus identity matrix, and A^k denotes the kth power of the matrix A. The following formulae are obvious:

$$A^* = I \oplus A^+, \qquad A^+ = AA^* = A^*A, \qquad \text{and} \qquad A^* = A^*A^* \ .$$

It may be useful to keep in mind the graph representation of matrices: to any matrix $A \in \overline{\mathbb{R}}_{\max}^{S\times S}$ is associated a directed graph with set of nodes S and an arc from i to j if the weight A_{ij} is different from $\mathbb{0}$. The weight of a path is by definition the max-plus product (that is, the sum) of the weights of its arcs. Then, A^+_{ij} and A^*_{ij} represent the supremum of the weights of all paths from i to j that are, respectively, of positive an nonnegative length.

Motivated by the analogy with potential theory, we will say that a vector $u \in \mathbb{R}_{\max}^S$ is (max-plus) *harmonic* if $Au = u$ and *super-harmonic* if $Au \leq u$. Note that we require the entries of a harmonic or super-harmonic vector to be distinct from $+\infty$. We shall say that a vector $\pi \in \mathbb{R}_{\max}^S$ is left (max-plus) harmonic if $\pi A = \pi$, π being thought of as a row vector. Likewise, we shall say that π is left (max-plus) super-harmonic if $\pi A \leq \pi$. Super-harmonic vectors have the following elementary characterisation.

Proposition 2.1. *A vector $u \in \mathbb{R}_{\max}^S$ is super-harmonic if and only if $u = A^*u$.*

Proof. If $u \in \mathbb{R}_{\max}^S$ is super-harmonic, then $A^k u \leq u$ for all $k \geq 1$, from which it follows that $u = A^*u$. The converse also holds, since $AA^*u = A^+u \leq A^*u$. □

From now on, we make the following assumption.

Assumption 2.2. *There exists a left super-harmonic vector with full support, in other words a row vector $\pi \in \mathbb{R}^S$ such that $\pi \geq \pi A$.*

By applying Proposition 2.1 to the transpose of A, we conclude that $\pi = \pi A^*$. Since π has no components equal to $\mathbb{0}$, we see that one consequence of the above assumption is that $A^*_{ij} \in \mathbb{R}_{\max}$ for all $i, j \in S$. A fortiori, $A_{ij} \in \mathbb{R}_{\max}$ for all $i, j \in S$.

The choice of π we make will determine which set of harmonic vectors is the focus of attention. It will be the set of harmonic vectors u that are π-*integrable*, meaning that $\pi u < \infty$. Of course, the boundary that we define will also depend on π, in general. For brevity, we shall omit the explicit dependence on π of the quantities that we introduce and shall omit the assumption on π in the statements of the theorems. We denote by $\mathscr{H}$ and $\mathscr{S}$, respectively, the set of π-integrable harmonic and π-integrable super-harmonic vectors.

It is often convenient to choose $\pi := A^*_{b\cdot}$ for some $b \in S$. (We use the notation $M_{i\cdot}$ and $M_{\cdot i}$ to denote, respectively, the ith row and ith column of any matrix

M.) We shall say that b is a *basepoint* when the vector π defined in this way has finite entries (in particular, a basepoint has access to every node in S). With this choice of π, every super-harmonic vector $u \in \mathbb{R}_{\max}^S$ is automatically π-integrable since, by Proposition 2.1, $\pi u = (A^* u)_b = u_b < +\infty$. So, in this case, $\mathscr{H}$ coincides with the set of all harmonic vectors. This conclusion remains true when $\pi := \sigma A^*$, where σ is any row vector with finite support, that is, with $\sigma_i = \mathbb{0}$ except for finitely many i.

We define the *Martin kernel* K with respect to π:

$$K_{ij} := A^*_{ij}(\pi_j)^{-1} \quad \text{for all } i, j \in S \ . \tag{3}$$

Since $\pi_i A^*_{ij} \leq (\pi A^*)_j = \pi_j$, we have

$$K_{ij} \leq (\pi_i)^{-1} \quad \text{for all } i, j \in S \ . \tag{4}$$

This shows that the columns $K_{\cdot j}$ are bounded above independently of j. By Tychonoff's theorem, the set of columns $\mathscr{K} := \{K_{\cdot j} \mid j \in S\}$ is relatively compact in the product topology of $\mathbb{R}_{\max}^S$. The *Martin space* $\mathscr{M}$ is defined to be the closure of $\mathscr{K}$. We call $\mathscr{B} := \mathscr{M} \setminus \mathscr{K}$ the *Martin boundary.* From (3) and (4), we get that $Aw \leq w$ and $\pi w \leq \mathbb{1}$ for all $w \in \mathscr{K}$. Since the set of vectors with these two properties can be written

$$\{w \in \mathbb{R}_{\max}^S \mid A_{ij} w_j \leq w_i \text{ and } \pi_k w_k \leq \mathbb{1} \text{ for all } i, j, k \in S\}$$

and this set is obviously closed in the product topology of $\mathbb{R}_{\max}^S$, we have that

$$\mathscr{M} \subset \mathscr{S} \quad \text{and} \quad \pi w \leq \mathbb{1} \quad \text{for all } w \in \mathscr{M} \ . \tag{5}$$

3 Harmonic vectors arising from recurrent nodes

Of particular interest are those column vectors of K that are harmonic. To investigate these we will need some basic notions and facts from max-plus spectral theory. Define the *maximal circuit mean* of A to be

$$\rho(A) := \bigoplus_{k \geq 1} (\operatorname{tr} A^k)^{1/k} \ ,$$

where $\operatorname{tr} A = \bigoplus_{i \in S} A_{ii}$. Thus, $\rho(A)$ is the maximum weight-to-length ratio for all the circuits of the graph of A. The existence of a super-harmonic row vector with full support, Assumption 2.2, implies that $\rho(A) \leq \mathbb{1}$ (see for instance Prop. 3.5 of [Dud92] or Lemma 2.2 of [AGW05]). Define the *normalised matrix* $\tilde{A} = \rho(A)^{-1} A$. The max-plus analogue of the notion of recurrence is defined in [AGW05]:

Definition 3.1 (Recurrence). We shall say that a node i is *recurrent* if $\tilde{A}^+_{ii} = \mathbb{1}$. We denote by $N^r(A)$ the set of recurrent nodes. We call *recurrent classes* of A the equivalence classes of $N^r(A)$ with the relation $\mathcal{R}$ defined by $i \mathcal{R} j$ if $\tilde{A}^+_{ij} \tilde{A}^+_{ji} = \mathbb{1}$.

This should be compared with the definition of recurrence for Markov chains, where a node is recurrent if one returns to it with probability one. Here, a node is recurrent if we can return to it with reward $\mathbb{1}$ in $\tilde{A}$.
Since $AA^* = A^+ \leq A^*$, every column of A^* is super-harmonic. Only those columns of A^* corresponding to recurrent nodes yield harmonic vectors:

PROPOSITION 3.2 (See [AGW05, Prop. 5.1]). *The column vector $A^*_{\cdot i}$ is harmonic if and only if $\rho(A) = \mathbb{1}$ and i is recurrent.* □

The same is true for the columns of K since they are proportional in the max-plus sense to those if A^*.
The following two results show that it makes sense to identify elements in the same recurrence class.

PROPOSITION 3.3. *Let $i, j \in S$ be distinct. Then $K_{\cdot i} = K_{\cdot j}$ if and only if $\rho(A) = \mathbb{1}$ and i and j are in the same recurrence class.*

Proof. Let $i, j \in S$ be such that $K_{\cdot i} = K_{\cdot j}$. Then, in particular, $K_{ii} = K_{ij}$, and so $A^*_{ij} = \pi_j(\pi_i)^{-1}$. Symmetrically, we obtain $A^*_{ji} = \pi_i(\pi_j)^{-1}$. Therefore, $A^*_{ij}A^*_{ji} = \mathbb{1}$. If $i \neq j$, then this implies that $A^+_{ii} \geq A^+_{ij}A^+_{ji} = A^*_{ij}A^*_{ji} = \mathbb{1}$, in which case $\rho(A) = \mathbb{1}$, i is recurrent, and i and j are in the same recurrence class. This shows the "only if" part of the proposition. Now let $\rho(A) = \mathbb{1}$ and i and j be in the same recurrence class. Then, according to [AGW05, Prop. 5.2], $A^*_{\cdot i} = A^*_{\cdot j}A^*_{ji}$, and so $K_{\cdot i} = K_{\cdot j}(\pi_i)^{-1}\pi_j A^*_{ji}$. But since $\pi = \pi A^*$, we have that $\pi_i \geq \pi_j A^*_{ji}$, and therefore $K_{\cdot i} \leq K_{\cdot j}$. The reverse inequality follows from a symmetrical argument. □

PROPOSITION 3.4. *Assume that $\rho(A) = \mathbb{1}$. Then, for all $u \in \mathscr{S}$ and i, j in the same recurrence class, we have $\pi_i u_i = \pi_j u_j$.*

Proof. Since $\pi \in \mathbb{R}^S$, we can consider the vector $\pi^{-1} := (\pi_i^{-1})_{i \in S}$. That π is super-harmonic can be expressed as $\pi_j \geq \pi_i A_{ij}$, for all $i, j \in S$. This is equivalent to $(\pi_i)^{-1} \geq A_{ij}(\pi_j)^{-1}$; in other words, that π^{-1}, seen as a column vector, is super-harmonic. Proposition 5.5 of [AGW05] states that the restriction of any two $\rho(A)$-super-eigenvectors of A to any recurrence class of A are proportional. Therefore, either $u = \mathbb{0}$ or the restrictions of u and π^{-1} to any recurrence class are proportional. In either case, the map $i \in S \mapsto \pi_i u_i$ is constant on each recurrence class. □

Remark 3.5. It follows from these two propositions that, for any $u \in \mathscr{S}$, the map $S \to \mathbb{R}_{\max}$, $i \mapsto \pi_i u_i$ induces a map $\mathscr{K} \to \mathbb{R}_{\max}$, $K_{\cdot i} \mapsto \pi_i u_i$. Thus, a super-harmonic vector may be regarded as a function defined on $\mathscr{K}$.

Let $u \in \mathbb{R}^S_{\max}$ be a π-integrable vector. We define the map $\mu_u : \mathscr{M} \to \mathbb{R}_{\max}$ by

$$\mu_u(w) := \limsup_{K_{\cdot j} \to w} \pi_j u_j := \inf_{W \ni w} \sup_{K_{\cdot j} \in W} \pi_j u_j \quad \text{for } w \in \mathscr{M} \quad ,$$

where the infimum is taken over all neighbourhoods W of w in $\mathscr{M}$. The reason why the limsup above cannot take the value $+\infty$ is that $\pi_j u_j \leq \pi u < +\infty$ for all $j \in S$. The following result shows that $\mu_u : \mathscr{M} \to \mathbb{R}_{\max}$ is an upper semi-continuous extension of the map from $\mathscr{K}$ to $\mathbb{R}_{\max}$ introduced in Remark 3.5.

Lemma 3.6. *Let u be a π-integrable super-harmonic vector. Then, $\mu_u(K_{\cdot i}) = \pi_i u_i$ for each $i \in S$ and $\mu_u(w)w \le u$ for each $w \in \mathscr{M}$. Moreover,*

$$u = \bigoplus_{w \in \mathscr{K}} \mu_u(w)w = \bigoplus_{w \in \mathscr{M}} \mu_u(w)w \enspace .$$

Proof. By Proposition 2.1, $A^* u = u$. Hence, for all $i \in S$,

$$u_i = \bigoplus_{j \in S} A^*_{ij} u_j = \bigoplus_{j \in S} K_{ij} \pi_j u_j \enspace . \tag{6}$$

We conclude that $u_i \ge K_{ij}\pi_j u_j$ for all $i, j \in S$. By taking the limsup with respect to j of this inequality, we obtain that

$$u_i \ge \limsup_{K_{\cdot j} \to w} K_{ij}\pi_j u_j \ge \liminf_{K_{\cdot j} \to w} K_{ij} \limsup_{K_{\cdot j} \to w} \pi_j u_j = w_i \mu_u(w) \enspace , \tag{7}$$

for all $w \in \mathscr{M}$ and $i \in S$. This shows the second part of the first assertion of the lemma. To prove the first part, we apply this inequality with $w = K_{\cdot i}$. We get that $u_i \ge K_{ii}\mu_u(K_{\cdot i})$. Since $K_{ii} = (\pi_i)^{-1}$, we see that $\pi_i u_i \ge \mu_u(K_{\cdot i})$. The reverse inequality follows from the definition of μ_u. The final statement of the lemma follows from Equation (6) and the first statement. □

4 The minimal Martin space

In probabilistic potential theory, one does not need the entire boundary to be able to represent harmonic vectors, a certain subset suffices. We shall see that the situation in the max-plus setting is similar. To define the (max-plus) minimal Martin space, we need to introduce another kernel:

$$K^\flat_{ij} := A^+_{ij}(\pi_j)^{-1} \quad \text{for all } i, j \in S \enspace .$$

Note that $K^\flat_{\cdot j} = AK_{\cdot j}$ is a function of $K_{\cdot j}$. For all $w \in \mathscr{M}$, we also define $w^\flat \in \mathbb{R}^S_{\max}$:

$$w^\flat_i = \liminf_{K_{\cdot j} \to w} K^\flat_{ij} \quad \text{for all } i \in S \enspace .$$

The following lemma shows that no ambiguity arises from this notation since $(K_{\cdot j})^\flat = K^\flat_{\cdot j}$.

Lemma 4.1. *We have $w^\flat = w$ for $w \in \mathscr{B}$, and $w^\flat = K^\flat_{\cdot j} = Aw$ for $w = K_{\cdot j} \in \mathscr{K}$. For all $w \in \mathscr{M}$, we have $w^\flat \in \mathscr{S}$ and $\pi w^\flat \le \mathbb{1}$.*

Proof. Let $w \in \mathscr{B}$. Then, for each $i \in S$, there exists a neighbourhood W of w such that $K_{\cdot i} \notin W$. So

$$w_i^\flat = \liminf_{K_{\cdot j} \to w} K_{ij}^\flat = \liminf_{K_{\cdot j} \to w} K_{ij} = w_i \ ,$$

proving that $w^\flat = w$.
Now let $w = K_{\cdot j}$ for some $j \in S$. Taking the sequence with constant value $K_{\cdot j}$, we see that $w^\flat \leq K_{\cdot j}^\flat$. To establish the opposite inequality, we observe that

$$w^\flat = \liminf_{K_{\cdot k} \to w} AK_{\cdot k} \geq \liminf_{K_{\cdot k} \to w} A_{\cdot i} K_{ik} = A_{\cdot i} w_i \qquad \text{for all } i \in S \ ,$$

or, in other words, $w^\flat \geq Aw$. Therefore we have shown that $w^\flat = K_{\cdot j}^\flat$.
The last assertion of the lemma follows from (5) and the fact that π is superharmonic. □

Next, we define two kernels H and $H^\flat$ over $\mathscr{M}$.

$$\begin{aligned} H(z, w) :=& \mu_w(z) = \limsup_{K_{\cdot i} \to z} \pi_i w_i = \limsup_{K_{\cdot i} \to z} \lim_{K_{\cdot j} \to w} \pi_i K_{ij} \\ H^\flat(z, w) :=& \mu_{w^\flat}(z) = \limsup_{K_{\cdot i} \to z} \pi_i w_i^\flat = \limsup_{K_{\cdot i} \to z} \liminf_{K_{\cdot j} \to w} \pi_i K_{ij}^\flat \ . \end{aligned}$$

Using the fact that $K^\flat \leq K$ and Inequality (4), we get that

$$H^\flat(z, w) \leq H(z, w) \leq \mathbb{1} \quad \text{for all } w, z \in \mathscr{M} \ .$$

If $w \in \mathscr{M}$, then both w and $w^\flat$ are elements of $\mathscr{S}$ by (5) and Lemma 4.1. Using the first assertion in Lemma 3.6, we get that

$$H(K_{\cdot i}, w) = \pi_i w_i \tag{8}$$

$$H^\flat(K_{\cdot i}, w) = \pi_i w_i^\flat \ . \tag{9}$$

In particular

$$H(K_{\cdot i}, K_{\cdot j}) = \pi_i K_{ij} = \pi_i A_{ij}^* (\pi_j)^{-1} \tag{10}$$

$$H^\flat(K_{\cdot i}, K_{\cdot j}) = \pi_i K_{ij}^\flat = \pi_i A_{ij}^+ (\pi_j)^{-1} \ . \tag{11}$$

Therefore, up to a diagonal similarity, H and $H^\flat$ are extensions to $\mathscr{M} \times \mathscr{M}$ of the kernels A^* and A^+ respectively.

LEMMA 4.2. *For all $w, z \in \mathscr{M}$, we have*

$$H(z, w) = \begin{cases} H^\flat(z, w) & \text{when } w \neq z \text{ or } w = z \in \mathscr{B} \ , \\ \mathbb{1} & \text{otherwise} \ . \end{cases}$$

Proof. If $w \in \mathscr{B}$, then $w^\flat = w$ by Lemma 4.1, and the equality of $H(z,w)$ and $H^\flat(z,w)$ for all $z \in \mathscr{M}$ follows immediately.
Let $w = K_{\cdot j}$ for some $j \in S$ and let $z \in \mathscr{M}$ be different from w. Then, there exists a neighbourhood W of z that does not contain w. Applying Lemma 4.1 again, we get that $w_i^\flat = K_{ij}^\flat = K_{ij} = w_i$ for all $i \in W$. We deduce that $H(z,w) = H^\flat(z,w)$ in this case also.
In the final case, we have $w = z \in \mathscr{K}$. The result follows from Equation (10). □

We define the *minimal Martin space* to be

$$\mathscr{M}^m := \{w \in \mathscr{M} \mid H^\flat(w,w) = \mathbb{1}\} \ .$$

From Lemma 4.2, we see that

$$\{w \in \mathscr{M} \mid H(w,w) = \mathbb{1}\} = \mathscr{M}^m \cup \mathscr{K} \ . \tag{12}$$

Lemma 4.3. *Every $w \in \mathscr{M}^m \cup \mathscr{K}$ satisfies $\pi w = \mathbb{1}$.*

Proof. We have

$$\pi w = \sup_{i \in S} \pi_i w_i \geq \limsup_{K_{\cdot i} \to w} \pi_i w_i = H(w,w) = \mathbb{1}.$$

By Equation (5), $\pi w \leq \mathbb{1}$, and the result follows. □

Proposition 4.4. *Every element of $\mathscr{M}^m$ is harmonic.*

Proof. If $\mathscr{K} \cap \mathscr{M}^m$ contains an element w, then, from Equation (11), we see that $\rho(A) = \mathbb{1}$ and w is recurrent. It follows from Proposition 3.2 that w is harmonic.
It remains to prove that the same is true for each element w of $\mathscr{B} \cap \mathscr{M}^m$. Let $i \in S$ be such that $w_i \neq \mathbb{0}$ and assume that $\beta > \mathbb{1}$ is given. Since $w \in \mathscr{B}$, w and $K_{\cdot i}$ will be different. We make two more observations. Firstly, by Lemma 4.2, $\limsup_{K_{\cdot j} \to w} \pi_j w_j = \mathbb{1}$. Secondly, $\lim_{K_{\cdot j} \to w} K_{ij} = w_i$. From these facts, we conclude that there exists $j \in S$, different from i, such that

$$\mathbb{1} \leq \beta \pi_j w_j \qquad \text{and} \qquad w_i \leq \beta K_{ij} \ . \tag{13}$$

Now, since i and j are distinct, we have $A_{ij}^* = A_{ij}^+ = (AA^*)_{ij}$. Therefore, we can find $k \in S$ such that

$$A_{ij}^* \leq \beta A_{ik} A_{kj}^* \ . \tag{14}$$

The final ingredient is that $A_{kj}^* w_j \leq w_k$ because w is super-harmonic. From this and the inequalities in (13) and (14), we deduce that $w_i \leq \beta^3 A_{ik} w_k \leq \beta^3 (Aw)_i$. Both β and i are arbitrary, so $w \leq Aw$. The reverse inequality is also true since every element of $\mathscr{M}$ is super-harmonic. Therefore w is harmonic. □

5 Martin spaces constructed from different basepoints

We shall see that when the left super-harmonic vector π is of the special form $\pi = A^*_{b\cdot}$ for some basepoint $b \in S$, the corresponding Martin boundary is independent of the basepoint.

Proposition 5.1. *The Martin spaces corresponding to different basepoints are homeomorphic. The same is true for Martin boundaries and minimal Martin spaces.*

Proof. Let $\mathscr{M}$ and $\mathscr{M}'$ denote the Martin spaces corresponding respectively to two different basepoints, b and b'. We set $\pi = A^*_{b\cdot}$ and $\pi' = A^*_{b'\cdot}$. We denote by K and K' the Martin kernels corresponding respectively to π and π'. By construction, $K_{bj} = \mathbb{1}$ holds for all $j \in S$. It follows that $w_b = \mathbb{1}$ for all $w \in \mathscr{M}$. Using the inclusion in (5), we conclude that $\mathscr{M} \subset \mathscr{S}_b := \{w \in \mathscr{S} \mid w_b = \mathbb{1}\}$, where $\mathscr{S}$ denotes the set of π-integrable super-harmonic functions. Observe that A^*_{bi} and $A^*_{b'j}$ are finite for all $i, j \in S$, since both b and b' are basepoints. Due to the inequalities $\pi' \geq A^*_{b'b}\pi$ and $\pi \geq A^*_{bb'}\pi'$, π-integrability is equivalent to π'-integrability. We deduce that $\mathscr{M}' \subset \mathscr{S}_{b'} := \{w' \in \mathscr{S} \mid w'_{b'} = \mathbb{1}\}$. Consider now the maps ϕ and ψ defined by

$$\phi(w) = w(w_{b'})^{-1},\ \forall w \in \mathscr{S}_b \qquad \psi(w') = w'(w'_b)^{-1},\ \forall w' \in \mathscr{S}_{b'} \ .$$

Observe that if $w \in \mathscr{S}_b$, then $w_{b'} \geq A^*_{b'b}w_b = A^*_{b'b} \neq \mathbb{0}$. Hence, $w \mapsto w_{b'}$ does not take the value $\mathbb{0}$ on $\mathscr{S}_b$. By symmetry, $w' \mapsto w'_b$ does not take the value zero on $\mathscr{S}_{b'}$. It follows that ϕ and ψ are mutually inverse homeomorphisms which exchange $\mathscr{S}_b$ and $\mathscr{S}_{b'}$. Since ϕ sends $K_{\cdot j}$ to $K'_{\cdot j}$, ϕ sends the the Martin space $\mathscr{M}$, which is the closure of $\mathscr{K} := \{K_{\cdot j} \mid j \in S\}$, to the Martin space $\mathscr{M}'$, which is the closure of $\mathscr{K}' := \{K'_{\cdot j} \mid j \in S\}$. Hence, ϕ sends the Martin boundary $\mathscr{M} \setminus \mathscr{K}$ to the Martin boundary $\mathscr{M}' \setminus \mathscr{K}'$.
It remains to show that the minimal Martin space corresponding to π, $\mathscr{M}^m$, is sent by ϕ to the minimal Martin space corresponding to π', $\mathscr{M}'^m$. Let

$$H'^\flat(z', w') = \limsup_{K'_{\cdot i} \to z'} \liminf_{K'_{\cdot j} \to w'} A^*_{b'i} A^+_{ij} (A^*_{b'j})^{-1} \ .$$

Since ϕ is an homeomorphism sending $K_{\cdot i}$ to $K'_{\cdot i}$, a net $(K_{\cdot i})_{i \in I}$ converges to w if and only if the net $(K'_{\cdot i})_{i \in I}$ converges to $\phi(w)$, and so

$$H'^\flat(\phi(z), \phi(w)) = \limsup_{K_{\cdot i} \to z} \liminf_{K_{\cdot j} \to w} A^*_{b'i} A^+_{ij} (A^*_{b'j})^{-1} = z_{b'} w_{b'}^{-1} H^\flat(z, w) \ .$$

It follows that $H^\flat(w, w) = \mathbb{1}$ if and only if $H'^\flat(\phi(w), \phi(w)) = \mathbb{1}$. Hence, $\phi(\mathscr{M}^m) = \mathscr{M}'^m$. □

Remark 5.2. Consider the kernel obtained by symmetrising the kernel $H^\flat$,

$$(z, w) \mapsto H^\flat(z, w) H^\flat(w, z) \ .$$

The final argument in the proof of Proposition 5.1 shows that this symmetrised kernel is independent of the basepoint, up to the identification of w and $\phi(w)$. The same is true for the kernel obtained by symmetrising H,

$$(z,w) \mapsto H(z,w)H(w,z) \ .$$

6 Martin representation of super-harmonic vectors

In probabilistic potential theory, each super-harmonic vector has a unique representation as integral over a certain set of vectors, the analogue of $\mathscr{M}^m \cup \mathscr{K}$. The situation is somewhat different in the max-plus setting. Firstly, according to Lemma 3.6, one does not need the whole of $\mathscr{M}^m \cup \mathscr{K}$ to obtain a representation: any set containing $\mathscr{K}$ will do. Secondly, the representation will not necessarily be unique. The following two theorems, however, show that $\mathscr{M}^m \cup \mathscr{K}$ still plays an important role.

Theorem 6.1 (Martin representation of super-harmonic vectors). *For each $u \in \mathscr{S}$, μ_u is the maximal $\nu : \mathscr{M}^m \cup \mathscr{K} \to \mathbb{R}_{\max}$ satisfying*

$$u = \bigoplus_{w \in \mathscr{M}^m \cup \mathscr{K}} \nu(w)w \ , \tag{15}$$

Any $\nu : \mathscr{M}^m \cup \mathscr{K} \to \mathbb{R}_{\max}$ satisfying this equation also satisfies

$$\sup_{w \in \mathscr{M}^m \cup \mathscr{K}} \nu(w) < +\infty \tag{16}$$

and any ν satisfying (16) defines by (15) an element u of $\mathscr{S}$.

Proof. By Lemma 3.6, u can be written as (15) with $\nu = \mu_u$. Suppose that $\nu : \mathscr{M}^m \cup \mathscr{K} \to \mathbb{R}_{\max}$ is an arbitrary function satisfying (15). We have

$$\pi u = \bigoplus_{w \in \mathscr{M}^m \cup \mathscr{K}} \nu(w)\pi w \ .$$

By Lemma 4.3, $\pi w = \mathbb{1}$ for each $w \in \mathscr{M}^m \cup \mathscr{K}$. Since $\pi u < +\infty$, we deduce that (16) holds.
Suppose that $\nu : \mathscr{M}^m \cup \mathscr{K} \to \mathbb{R}_{\max}$ is an arbitrary function satisfying (16) and define u by (15). Since the operation of multiplication by A commutes with arbitrary suprema, we have $Au \leq u$. Also $\pi u = \bigoplus_{w \in \mathscr{M}^m \cup \mathscr{K}} \nu(w) < +\infty$. So $u \in \mathscr{S}$.
Let $w \in \mathscr{M}^m \cup \mathscr{K}$. Then $\nu(w)w_i \leq u_i$ for all $i \in S$. So we have

$$\nu(w)H(w,w) = \nu(w) \limsup_{K \cdot i \to w} \pi_i w_i \leq \limsup_{K \cdot i \to w} \pi_i u_i = \mu_u(w) \ .$$

Since $H(w,w) = \mathbb{1}$, we obtain $\nu(w) \leq \mu_u(w)$. ∎

We shall now give another interpretation of the set $\mathscr{M}^m \cup \mathscr{K}$. Let V be a *subsemimodule* of $\mathbb{R}_{\max}^S$, that is a subset of $\mathbb{R}_{\max}^S$ stable under pointwise maximum and the addition of a constant (see [LMS01, CGQ04] for definitions and properties of semimodules). We say that a vector $\xi \in V\setminus\{\mathbb{0}\}$ is an *extremal generator* of V if $\xi = u \oplus v$ with $u, v \in V$ implies that either $\xi = u$ or $\xi = v$. This concept has, of course, an analogue in the usual algebra, where extremal generators are defined for cones. Max-plus extremal generators are also called *join irreducible* elements in the lattice literature. Clearly, if ξ is an extremal generator of V then so is $\alpha\xi$ for all $\alpha \in \mathbb{R}$. We say that a vector $u \in \mathbb{R}_{\max}^S$ is *normalised* if $\pi u = \mathbb{1}$. If V is a subset of the set of π-integrable vectors, then the set of its extremal generators is exactly the set of $\alpha\xi$, where $\alpha \in \mathbb{R}$ and ξ is a normalised extremal generator.

Theorem 6.2. *The normalised extremal generators of $\mathscr{S}$ are precisely the elements of $\mathscr{M}^m \cup \mathscr{K}$.*

The proof of this theorem relies on a series of auxiliary results.

Lemma 6.3. *Suppose that $\xi \in \mathscr{M}^m \cup \mathscr{K}$ can be written in the form $\xi = \bigoplus_{w\in\mathscr{M}} \nu(w)w$, where $\nu : \mathscr{M} \to \mathbb{R}_{\max}$ is upper semicontinuous. Then, there exists $w \in \mathscr{M}$ such that $\xi = \nu(w)w$.*

Proof. For all $i \in S$, we have $\xi_i = \bigoplus_{w\in\mathscr{M}} \nu(w)w_i$. As the conventional sum of two upper semicontinuous functions, the function $\mathscr{M} \to \mathbb{R}_{\max} : w \mapsto \nu(w)w_i$ is upper semicontinuous. Since $\mathscr{M}$ is compact, the supremum of $\nu(w)w_i$ is attained at some $w^{(i)} \in \mathscr{M}$, in other words $\xi_i = \nu(w^{(i)})w_i^{(i)}$. Since $H(\xi,\xi) = \mathbb{1}$, by definition of H, there exists a net $(i_k)_{k\in D}$ of elements of S such that $K_{\cdot i_k}$ converges to ξ and $\pi_{i_k}\xi_{i_k}$ converges to $\mathbb{1}$. The Martin space $\mathscr{M}$ is compact and so, by taking a subnet if necessary, we may assume that $(w^{(i_k)})_{k\in D}$ converges to some $w \in \mathscr{M}$. Now, for all $j \in S$,

$$K_{ji_k}\pi_{i_k}\xi_{i_k} = A^*_{ji_k}\xi_{i_k} = A^*_{ji_k}\nu(w^{(i_k)})w^{(i_k)}_{i_k} \le \nu(w^{(i_k)})w^{(i_k)}_j \quad ,$$

since $w^{(i_k)}$ is super-harmonic. Taking the limsup as $k \to \infty$, we get that $\xi_j \le \nu(w)w_j$. The reverse inequality is true by assumption and therefore $\xi_j = \nu(w)w_j$. ☐

The following consequence of this lemma proves one part of Theorem 6.2.

Corollary 6.4. *Every element of $\mathscr{M}^m \cup \mathscr{K}$ is a normalised extremal generator of $\mathscr{S}$.*

Proof. Let $\xi \in \mathscr{M}^m \cup \mathscr{K}$. We know from Lemma 4.3 that ξ is normalised. In particular, $\xi \neq \mathbb{0}$. We also know from Equation (5) that $\xi \in \mathscr{S}$. Suppose $u, v \in \mathscr{S}$ are such that $\xi = u \oplus v$. By Lemma 3.6, we have $u = \bigoplus_{w\in\mathscr{M}} \mu_u(w)w$ and $v = \bigoplus_{w\in\mathscr{M}} \mu_v(w)w$. Therefore, $\xi = \bigoplus_{w\in\mathscr{M}} \nu(w)w$, with $\nu = \mu_u \oplus \mu_v$. Since μ_u and μ_v are upper semicontinuous maps from $\mathscr{M}$ to $\mathbb{R}_{\max}$, so is ν. By the previous lemma, there exists $w \in \mathscr{M}$ such that $\xi = \nu(w)w$. Now, $\nu(w)$

must equal either $\mu_u(w)$ or $\mu_v(w)$. Without loss of generality, assume the first case. Then $\xi = \mu_u(w)w \leq u$, and since $\xi \geq u$, we deduce that $\xi = u$. This shows that ξ is an extremal generator of $\mathscr{S}$. $\square$

The following lemma will allow us to complete the proof of Theorem 6.2.

Lemma 6.5. *Let $\mathscr{F} \subset \mathbb{R}_{\max}^S$ have compact closure $\bar{\mathscr{F}}$ in the product topology. Denote by V the set whose elements are of the form*

$$\xi = \bigoplus_{w\in\mathscr{F}} \nu(w)w \in \mathbb{R}_{\max}^S, \qquad \text{with } \nu : \mathscr{F} \to \mathbb{R}_{\max},\ \sup_{w\in\mathscr{F}} \nu(w) < \infty \ . \tag{17}$$

Let ξ be an extremal generator of V, and ν be as in (17). Then, there exists $w \in \bar{\mathscr{F}}$ such that $\xi = \hat{\nu}(w)w$, where

$$\hat{\nu}(w) := \limsup_{w'\to w,\, w'\in\mathscr{F}} \nu(w').$$

Proof. Since $\nu \leq \hat{\nu}$, we have $\xi \leq \bigoplus_{w\in\mathscr{F}} \hat{\nu}(w)w \leq \bigoplus_{w\in\bar{\mathscr{F}}} \hat{\nu}(w)w$. Clearly, $\nu(w)w_i \leq \xi_i$ for all $i \in S$ and $w \in \mathscr{F}$. Taking the limsup as $w \to w'$ for any $w' \in \bar{\mathscr{F}}$, we get that

$$\xi_i \geq \hat{\nu}(w')w'_i.$$

Combined with the previous inequality, this gives us the representations

$$\xi = \bigoplus_{w\in\mathscr{F}} \hat{\nu}(w)w = \bigoplus_{w\in\bar{\mathscr{F}}} \hat{\nu}(w)w \ . \tag{18}$$

Consider now, for each $i \in S$ and $\alpha < \mathbb{1}$, the set

$$U_{i,\alpha} := \{w \in \bar{\mathscr{F}} \mid \hat{\nu}(w)w_i < \alpha\xi_i\} \ ,$$

which is open in $\bar{\mathscr{F}}$ since the map $w \mapsto \hat{\nu}(w)w_i$ is upper semicontinuous. Let $\xi \in V\backslash\{\mathbb{0}\}$ be such that $\xi \neq \hat{\nu}(w)w$ for all $w \in \bar{\mathscr{F}}$. We conclude that there exist $i \in S$ and $\alpha < \mathbb{1}$ such that $\alpha\xi_i > \hat{\nu}(w)w_i$, which shows that $(U_{i,\alpha})_{i\in S,\alpha<\mathbb{1}}$ is an open covering of $\bar{\mathscr{F}}$. Since $\bar{\mathscr{F}}$ is compact, there exists a finite sub-covering $U_{i_1,\alpha_1}, \dots, U_{i_n,\alpha_n}$.

Using (18) and the idempotency of the $\oplus$ law, we get

$$\xi = \xi^1 \oplus \cdots \oplus \xi^n \qquad \text{with } \xi^j = \bigoplus_{w\in U_{i_j,\alpha_j}\cap\mathscr{F}} \hat{\nu}(w)w \ , \tag{19}$$

for $j = 1 \dots, n$. Since the supremum of $\hat{\nu}$ over $\bar{\mathscr{F}}$ is the same as that over $\mathscr{F}$, the vectors $\xi^1, \dots, \xi^n$ all belong to V. Since ξ is an extremal generator of $\mathscr{S}$, we must have $\xi = \xi^j$ for some j. Then $U_{i_j,\alpha_j} \cap \mathscr{F}$ is non-empty, and so $\xi_{i_j} > \mathbb{0}$. But, from the definition of U_{i_j,α_j},

$$\xi^j_{i_j} = \bigoplus_{w\in U_{i_j,\alpha_j}\cap\mathscr{F}} \hat{\nu}(w)w_{i_j} \leq \alpha_{i_j}\xi_{i_j} < \xi_{i_j} \ .$$

This shows that ξ^j is different from ξ, and so Equation (19) gives the required decomposition of ξ, proving it is not an extremal generator of V. $\square$

We now conclude the proof of Theorem 6.2:

Corollary 6.6. *Every normalised extremal generator of $\mathscr{S}$ belongs to $\mathscr{M}^m \cup \mathscr{K}$.*

Proof. Take $\mathscr{F} = \mathscr{M}^m \cup \mathscr{K}$ and let V be as defined in Lemma 6.5. Then, by definition, $\bar{\mathscr{F}} = \mathscr{M}$, which is compact. By Theorem 6.1, $V = \mathscr{S}$. Let ξ be a normalised extremal generator of $\mathscr{S}$. Again by Theorem 6.1, $\xi = \oplus_{w\in\mathscr{F}}\mu_\xi(w)w$. Since μ_ξ is upper semicontinuous on $\mathscr{M}$, Lemma 6.5 yields $\xi = \mu_\xi(w)w$ for some $w \in \mathscr{M}$, with $\mu_\xi(w) \neq \mathbb{0}$ since $\xi \neq \mathbb{0}$. Note that $\mu_{\alpha u} = \alpha\mu_u$ for all $\alpha \in \mathbb{R}_{\max}$ and $u \in \mathscr{S}$. Applying this to the previous equation and evaluating at w, we deduce that $\mu_\xi(w) = \mu_\xi(w)\mu_w(w)$. Thus, $H(w, w) = \mu_w(w) = \mathbb{1}$. In addition, ξ is normalised and so, by Lemma 4.3,

$$\mathbb{1} = \pi\xi = \mu_\xi(w)\pi w = \mu_\xi(w).$$

Hence $\xi = w \in \mathscr{M}^m \cup \mathscr{K}$. □

7 Almost-geodesics

In order to prove a Martin representation theorem for harmonic vectors, we will use a notion appearing in [YK92] and [KM97, § 2.4], which we will call almost-geodesic. A variation of this notion appeared in [Rie02]. We will compare the two notions later in the section.

Let u be a super-harmonic vector, that is $u \in \mathbb{R}^S_{\max}$ and $Au \leq u$. Let $\alpha \in \mathbb{R}_{\max}$ be such that $\alpha \geq \mathbb{1}$. We say that a sequence $(i_k)_{k\geq 0}$ with values in S is an *α-almost-geodesic* with respect to u if $u_{i_0} \in \mathbb{R}$ and

$$u_{i_0} \leq \alpha A_{i_0 i_1} \cdots A_{i_{k-1} i_k} u_{i_k} \quad \text{for all } k \geq 0 \ . \tag{20}$$

Similarly, $(i_k)_{k\geq 0}$ is an *α-almost-geodesic* with respect to a left super-harmonic vector σ if $\sigma_{i_0} \in \mathbb{R}$ and

$$\sigma_{i_k} \leq \alpha\sigma_{i_0} A_{i_0 i_1} \cdots A_{i_{k-1} i_k} \quad \text{for all } k \geq 0 \ .$$

We will drop the reference to α when its value is unimportant. Observe that, if $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to some right super-harmonic vector u, then both u_{i_k} and $A_{i_{k-1} i_k}$ are in $\mathbb{R}$ for all $k \geq 0$. This is not necessarily true if $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to a left super-harmonic vector σ, however, if additionally $\sigma_{i_k} \in \mathbb{R}$ for all $k \geq 0$, then $A_{i_{k-1} i_k} \in \mathbb{R}$ for all $k \geq 0$.

Lemma 7.1. *Let $u, \sigma \in \mathbb{R}^S_{\max}$ be, respectively, right and left super-harmonic vectors and assume that u is σ-integrable, that is $\sigma u < +\infty$. If $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to u, and if $\sigma_{i_0} \in \mathbb{R}$, then $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to σ.*

Proof. Multiplying Equation (20) by $\sigma_{i_k}(u_{i_0})^{-1}$, we obtain

$$\sigma_{i_k} \leq \alpha\sigma_{i_k}u_{i_k}(u_{i_0})^{-1}A_{i_0i_1}\cdots A_{i_{k-1}i_k} \leq \alpha(\sigma u)(\sigma_{i_0}u_{i_0})^{-1}\sigma_{i_0}A_{i_0i_1}\cdots A_{i_{k-1}i_k} \enspace .$$

So $(i_k)_{k\geq 0}$ is a β-almost-geodesic with respect to σ, with $\beta := \alpha(\sigma u)(\sigma_{i_0}u_{i_0})^{-1} \geq \alpha$. ☐

Lemma 7.2. *Let $(i_k)_{k\geq 0}$ be an almost-geodesic with respect to π and let $\beta > \mathbb{1}$. Then, for ℓ large enough, $(i_k)_{k\geq\ell}$ is a β-almost-geodesic with respect to π.*

Proof. Consider the matrix $\bar{A}_{ij} := \pi_i A_{ij}(\pi_j)^{-1}$. The fact that $(i_k)_{k\geq 0}$ is an α-almost-geodesic with respect to π is equivalent to

$$p_k := (\bar{A}_{i_0i_1})^{-1}\cdots(\bar{A}_{i_{k-1}i_k})^{-1} \leq \alpha \quad \text{for all } k \geq 0 \enspace .$$

Since $(\bar{A}_{i_{\ell-1}i_\ell})^{-1} \geq \mathbb{1}$ for all $\ell \geq 1$, the sequence $\{p_k\}_{k\geq 1}$ is nondecreasing. The upper bound then implies it converges to a finite limit. The Cauchy criterion states that

$$\lim_{\ell,k\to\infty,\,\ell<k} \bar{A}_{i_\ell i_{\ell+1}}\cdots\bar{A}_{i_{k-1}i_k} = \mathbb{1} \enspace .$$

This implies that, given any $\beta > \mathbb{1}$, $\bar{A}_{i_\ell i_{\ell+1}}\cdots\bar{A}_{i_{k-1}i_k} \geq \beta^{-1}$ for k and ℓ large enough, with $k > \ell$. Writing this formula in terms of A rather than $\bar{A}$, we see that, for ℓ large enough, $(i_k)_{k\geq\ell}$ is a β-almost-geodesic with respect to π. ☐

Proposition 7.3. *If $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to π, then $K_{\cdot i_k}$ converges to some $w \in \mathscr{M}^m$.*

Proof. Let $\beta > \mathbb{1}$. By Lemma 7.2, $(i_k)_{k\geq\ell}$ is a β-almost-geodesic with respect to π, for ℓ large enough. Then, for all $k > \ell$,

$$\pi_{i_k} \leq \beta\pi_{i_\ell}A^{+}_{i_\ell i_k} \leq \beta\pi_{i_\ell}A^{*}_{i_\ell i_k} \enspace .$$

Since π is left super-harmonic, we have $\pi_{i_\ell}A^{*}_{i_\ell i_k} \leq \pi_{i_k}$. Dividing by $\beta\pi_{i_k}$ the former inequalities, we deduce that

$$\beta^{-1} \leq \pi_{i_\ell}K^{\flat}_{i_\ell i_k} \leq \pi_{i_\ell}K_{i_\ell i_k} \leq \mathbb{1} \enspace . \tag{21}$$

Since $\mathscr{M}$ is compact, it suffices to check that all convergent subnets of $K_{\cdot i_k}$ have the same limit $w \in \mathscr{M}^m$. Let $(i_{k_d})_{d\in D}$ and $(i_{\ell_e})_{e\in E}$ denote subnets of $(i_k)_{k\geq 0}$, such that the nets $(K_{\cdot i_{k_d}})_{d\in D}$ and $(K_{\cdot i_{\ell_e}})_{e\in E}$ converge to some $w \in \mathscr{M}$ and $w' \in \mathscr{M}$, respectively. Applying (21) with $\ell = \ell_e$ and $k = k_d$, and taking the limit with respect to d, we obtain $\beta^{-1} \leq \pi_{i_{\ell_e}}w_{i_{\ell_e}}$. Taking now the limit with respect to e, we get that $\beta^{-1} \leq H(w',w)$. Since this holds for all $\beta > \mathbb{1}$, we obtain $\mathbb{1} \leq H(w',w)$, thus $H(w',w) = \mathbb{1}$. From Lemma 3.6, we deduce that $w \geq \mu_w(w')w' = H(w',w)w' = w'$. By symmetry, we conclude that $w = w'$, and so $H(w,w) = \mathbb{1}$. By Equation (12), $w \in \mathscr{M}^m \cup \mathscr{K}$. Hence, $(K_{\cdot i_k})_{k\geq 0}$ converges towards some $w \in \mathscr{M}^m \cup \mathscr{K}$.

Assume by contradiction that $w \notin \mathscr{M}^m$. Then, $w = K_{\cdot j}$ for some $j \in S$, and $H^\flat(w,w) < \mathbb{1}$ by definition of $\mathscr{M}^m$. By (11), this implies that $\pi_j K^\flat_{jj} = A^+_{jj} < \mathbb{1}$. If the sequence $(i_k)_{k\geq 0}$ takes the value j infinitely often, then, we can deduce from Equation (21) that $A^+_{jj} = \mathbb{1}$, a contradiction. Hence, for k large enough, i_k does not take the value j, which implies, by Lemma 4.1, that $w_{i_k} = w^\flat_{i_k}$. Using Equation (21), we obtain $H^\flat(w,w) \geq \limsup_{k\to\infty} \pi_{i_k} w^\flat_{i_k} = \limsup_{k\to\infty} \pi_{i_k} w_{i_k} = \mathbb{1}$, which contradicts our assumption on w. We have shown that $w \in \mathscr{M}^m$. ☐

Remark 7.4. An inspection of the proof of Proposition 7.3 shows that the same conclusion holds under the weaker hypothesis that for all $\beta > \mathbb{1}$, we have $\pi_{i_k} \leq \beta \pi_{i_\ell} A^+_{i_\ell i_k}$ for all ℓ large enough and $k > \ell$.

Lemma 7.5. *If $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to π, and if w is the limit of $K_{\cdot i_k}$, then*

$$\lim_{k\to\infty} \pi_{i_k} w_{i_k} = \mathbb{1} \enspace .$$

Proof. Let $\beta > \mathbb{1}$. By Lemma 7.2, $(i_k)_{k\geq \ell}$ is a β-almost-geodesic with respect to π for ℓ large enough. Hence, for all $k \geq \ell$, $\pi_{i_k} \leq \beta \pi_{i_\ell} A^*_{i_\ell i_k}$, and so $\mathbb{1} \leq \beta \pi_{i_\ell} A^*_{i_\ell i_k} \pi^{-1}_{i_k} = \beta \pi_{i_\ell} K_{i_\ell i_k}$. Since $K_{i_\ell i_k}$ converges to w_{i_ℓ} when k tends to infinity, we deduce that $\mathbb{1} \leq \beta \liminf_{\ell\to\infty} \pi_{i_\ell} w_{i_\ell}$, and since this holds for all $\beta > \mathbb{1}$, we get $\mathbb{1} \leq \liminf_{\ell\to\infty} \pi_{i_\ell} w_{i_\ell}$. Since $\pi_j w_j \leq \mathbb{1}$ for all j, the lemma is proved. ☐

Proposition 7.6. *Let u be a π-integrable super-harmonic vector. Then, μ_u is continuous along almost-geodesics, meaning that if $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to π and if $K_{\cdot i_k}$ tends to w, then,*

$$\mu_u(w) = \lim_{k\to\infty} \mu_u(K_{\cdot i_k}) = \lim_{k\to\infty} \pi_{i_k} u_{i_k} \enspace .$$

Proof. Recall that $\pi_i u_i = \mu_u(K_{\cdot i})$ holds for all i, as shown in Lemma 3.6. It also follows from this lemma that $u \geq \mu_u(w) w$, and so $\pi_i u_i \geq \pi_i w_i \mu_u(w)$ for all $i \in S$. Hence,

$$\begin{aligned} \liminf_{k\to\infty} \pi_{i_k} u_{i_k} &\geq \liminf_{k\to\infty} \pi_{i_k} w_{i_k} \mu_u(w) \\ &= \mu_u(w) \enspace , \end{aligned}$$

by Lemma 7.5. Moreover, $\limsup_{k\to\infty} \pi_{i_k} u_{i_k} \leq \mu_u(w)$, by definition of $\mu_u(w)$. ☐

Combining Lemma 7.1 and Proposition 7.3, we deduce the following.

Corollary 7.7. *If $(i_k)_{k\geq 0}$ is an almost-geodesic with respect to a π-integrable super-harmonic vector, then $K_{\cdot i_k}$ converges to some element of $\mathscr{M}^m$.*

For brevity, we shall say sometimes that an almost-geodesic $(i_k)_{k\geq 0}$ *converges* to a vector w when $K_{\cdot i_k}$ converges to w. We state a partial converse to Proposition 7.3.

Proposition 7.8. *Assume that $\mathscr{M}$ is first-countable. For all $w \in \mathscr{M}^m$, there exists an almost-geodesic with respect to π converging to w.*

Proof. By definition, $H^\flat(w,w) = 0$. Writing this formula explicitly in terms of A_{ij} and making the transformation $\bar{A}_{ij} := \pi_i A_{ij}(\pi_j)^{-1}$, we get

$$\limsup_{K_{\cdot i}\to w} \liminf_{K_{\cdot j}\to w} \bar{A}^+_{ij} = \mathbb{1} \enspace .$$

Fix a sequence $(\alpha_k)_{k\geq 0}$ in $\mathbb{R}_{\max}$ such that $\alpha_k > \mathbb{1}$ and $\alpha := \alpha_0\alpha_1\cdots < +\infty$. Fix also a decreasing sequence $(W_k)_{k\geq 0}$ of open neighbourhoods of w. We construct a sequence $(i_k)_{k\geq 0}$ in S inductively as follows. Given i_{k-1}, we choose i_k to have the following three properties:

a. $K_{\cdot i_k} \in W_k$,

b. $\liminf_{K_{\cdot j}\to w} \bar{A}^+_{i_k j} > \alpha_k^{-1}$,

c. $\bar{A}^+_{i_{k-1} i_k} > \alpha_{k-1}^{-1}$.

Notice that it is possible to satisfy (c) because i_{k-1} was chosen to satisfy (b) at the previous step. We require i_0 to satisfy (a) and (b) but not (c). Since $\mathscr{M}$ is first-countable, one can choose the sequence $(W_k)_{k\geq 0}$ in such a way that every sequence $(w_k)_{k\geq 0}$ in $\mathscr{M}$ with $w_k \in W_k$ converges to w. By (c), one can find, for all $k \in \mathbb{N}$, a finite sequence $(i_k^\ell)_{0\leq \ell\leq N_k}$ such that $i_k^0 = i_k$, $i_k^{N_k} = i_{k+1}$, and

$$\bar{A}_{i_k^0, i_k^1}\cdots \bar{A}_{i_k^{N_k-1}, i_k^{N_k}} > \alpha_k^{-1} \quad \text{for all } k \in \mathbb{N} \enspace .$$

Since $\bar{A}_{ij} \leq \mathbb{1}$ for all $i,j \in S$, we obtain

$$\bar{A}_{i_k^0, i_k^1}\cdots \bar{A}_{i_k^{n-1}, i_k^{n}} > \alpha_k^{-1} \quad \text{for all } k \in \mathbb{N},\ 1 \leq n \leq N_k \enspace .$$

Concatenating the sequences $(i_k^\ell)_{0\leq \ell\leq N_k}$, we obtain a sequence $(j_m)_{m\geq 0}$ such that $\alpha^{-1} \leq \bar{A}_{j_0 j_1}\cdots \bar{A}_{j_{m-1} j_m}$ for all $m \in \mathbb{N}$, in other words an α-almost-geodesic with respect to π. From Lemma 7.3, we know that $K_{\cdot j_m}$ converges to some point in $\mathscr{M}$. Since (i_k) is a subsequence of (j_m) and $K_{\cdot i_k}$ converges to w, we deduce that $K_{\cdot j_m}$ also converges to w. ☐

Remark 7.9. If S is countable, the product topology on $\mathscr{M}$ is metrisable. Then, the assumption of Proposition 7.8 is satisfied.

Remark 7.10. Assume that (S,d) is a metric space, let $A_{ij} = A^*_{ij} = -d(i,j)$ for $i,j \in S$, and let $\pi = A^*_{b\cdot}$ for any $b \in S$. We have $K_{\cdot j} = -d(\cdot, j) + d(b,j)$. Using the triangle inequality for d, we see that, for all $k \in S$, the function $K_{\cdot k}$ is non-expansive, meaning that $|K_{ik} - K_{jk}| \leq d(i,j)$ for all $i,j \in S$. It follows that every map in $\mathscr{M}$ is non-expansive. By Ascoli's theorem, the topology of pointwise convergence on $\mathscr{M}$ coincides with the topology of uniform convergence on compact sets. Hence, if S is a countable union of compact sets, then $\mathscr{M}$ is metrisable and the assumption of Proposition 7.8 is satisfied.

Example 7.11. The assumption in Proposition 7.8 cannot be dispensed with. To see this, take $S = \omega_1$, the first uncountable ordinal. For all $i, j \in S$, define $A_{ij} := 0$ if $i < j$ and $A_{ij} := -1$ otherwise. Then, $\rho(A) = \mathbb{1}$ and $A = A^+$. Also A^*_{ij} equals 0 when $i \leq j$ and -1 otherwise. We take $\pi := A^*_{0\cdot}$, where 0 denotes the smallest ordinal. With this choice, $\pi_i = \mathbb{1}$ for all $i \in S$, and $K = A^*$.
Let $\mathscr{D}$ be the set of maps $S \to \{-1, 0\}$ that are non-decreasing and take the value 0 at 0. For each $z \in \mathscr{D}$, define $s(z) := \sup\{i \in S \mid z_i = 0\} \in S \cup \{\omega_1\}$. Our calculations above lead us to conclude that

$$\mathscr{K} = \{z \in \mathscr{D} \mid s(z) \in S \text{ and } z_{s(z)} = 0\} \ .$$

We note that $\mathscr{D}$ is closed in the product topology on $\{-1, 0\}^S$ and contains $\mathscr{K}$. Furthermore, every $z \in \mathscr{D} \setminus \mathscr{K}$ is the limit of the net $(A^*_{\cdot d})_{d \in D}$ indexed by the directed set $D = \{d \in S \mid d < s_z\}$. Therefore the Martin space is given by $\mathscr{M} = \mathscr{D}$. Every limit ordinal γ less than or equal to ω_1 yields one point z^γ in the Martin boundary $\mathscr{B} := \mathscr{M} \setminus \mathscr{K}$: we have $z^\gamma_i = 0$ for $i < \gamma$, and $z^\gamma_i = -1$ otherwise.
Since $A^+_{ii} = A_{ii} = -1$ for all $i \in S$, there are no recurrent points, and so $\mathscr{K} \cap \mathscr{M}^m$ is empty. For any $z \in \mathscr{B}$, we have $z_d = 0$ for all $d < s(z)$. Taking the limsup, we conclude that $H(z, z) = \mathbb{1}$, thus $\mathscr{M}^m = \mathscr{B}$. In particular, the identically zero vector z^{ω_1} is in $\mathscr{M}^m$.
Since a countable union of countable sets is countable, for any sequence $(i_k)_{k \in \mathbb{N}}$ of elements of S, the supremum $I = \sup_{k \in \mathbb{N}} i_k$ belongs to S, and so its successor ordinal, that we denote by $I+1$, also belongs to S. Since $\lim_{k \to \infty} K_{I+1, i_k} = -1$, $K_{\cdot i_k}$ cannot converge to z^{ω_1}, which shows that the point z^{ω_1} in the minimal Martin space is not the limit of an almost-geodesic.

We now compare our notion of almost-geodesic with that of Rieffel [Rie02] in the metric space case. We assume that (S, d) is a metric space and take $A_{ij} = A^*_{ij} = -d(i, j)$ and $\pi_j = -d(b, j)$, for an some $b \in S$. The compactification of S discussed in [Rie02], called there the *metric compactification*, is the closure of $\mathscr{K}$ in the topology of uniform convergence on compact sets, which, by Remark 7.10, is the same as its closure in the product topology. It thus coincides with the Martin space $\mathscr{M}$. We warn the reader that variants of the metric compactification can be found in the literature, in particular, the references [Gro81, Bal95] use the topology of uniform convergence on bounded sets rather than on compacts.
Observe that the basepoint b can be chosen in an arbitrary way: indeed, for all $b' \in S$, setting $\pi' = A^*_{b'\cdot}$, we get $\pi' \geq A^*_{b'b}\pi$ and $\pi \geq A^*_{bb'}\pi'$, which implies that almost-geodesics in our sense are the same for the basepoints b and b'. Therefore, when speaking of almost-geodesics in our sense, in a metric space, we will omit the reference to π.
Rieffel defines an almost-geodesic as an S-valued map γ from an unbounded set $\mathcal{T}$ of real nonnegative numbers containing 0, such that for all $\epsilon > 0$, for all $s \in \mathcal{T}$ large enough, and for all $t \in \mathcal{T}$ such that $t \geq s$,

$$|d(\gamma(t), \gamma(s)) + d(\gamma(s), \gamma(0)) - t| < \epsilon \ .$$

By taking $t = s$, one sees that $|d(\gamma(t), \gamma(0)) - t| < \epsilon$. Thus, almost-geodesics in the sense of Rieffel are "almost" parametrised by arc-length, unlike those in our sense.

Proposition 7.12. *Any almost-geodesic in the sense of Rieffel has a subsequence that is an almost-geodesic in our sense. Conversely, any almost-geodesic in our sense that is not bounded has a subsequence that is an almost-geodesic in the sense of Rieffel.*

Proof. Let $\gamma : \mathcal{T} \to S$ denote an almost-geodesic in the sense of Rieffel. Then, for all $\beta > \mathbb{1}$, we have

$$A^*_{\gamma(0),\gamma(t)} \leq \beta A^*_{\gamma(0),\gamma(s)} A^*_{\gamma(s)\gamma(t)} \tag{22}$$

for all $s \in \mathcal{T}$ large enough and for all $t \in \mathcal{T}$ such that $t \geq s$. Since the choice of the basepoint b is irrelevant, we may assume that $b = \gamma(0)$, so that $\pi_{\gamma(s)} = A^*_{\gamma(0),\gamma(s)}$. As in the proof of Lemma 7.2 we set $\bar{A}_{ij} = \pi_i A^*_{ij} \pi_j^{-1}$. We deduce from (22) that

$$\beta^{-1} \leq \bar{A}_{\gamma(s)\gamma(t)} \leq \mathbb{1} \enspace .$$

Let us choose a sequence $\beta_1, \beta_2, \ldots \geq \mathbb{1}$ such that the product $\beta_1\beta_2\ldots$ converges to a finite limit. We can construct a sequence $t_0 < t_1 < \ldots$ of elements of $\mathcal{T}$ such that, setting $i_k = \gamma(t_{i_k})$,

$$\bar{A}_{i_k i_{k+1}} \geq \beta_k^{-1} \enspace .$$

Then, the product $\bar{A}_{i_0 i_1} \bar{A}_{i_1 i_2} \cdots$ converges, which implies that the sequence i_0, $i_1, \ldots$ is an almost-geodesic in our sense.
Conversely, let $i_0, i_1, \ldots$ be an almost-geodesic in our sense, and assume that $t_k = d(b, i_k)$ is not bounded. After replacing i_k by a subsequence, we may assume that $t_0 < t_1 < \ldots$. We set $\mathcal{T} = \{t_0, t_1, \ldots\}$ and $\gamma(t_k) = i_k$. We choose the basepoint $b = i_0$, so that $t_0 = 0 \in \mathcal{T}$, as required in the definition of Rieffel. Lemma 7.2 implies that

$$A^*_{bi_k} \leq \beta A^*_{bi_\ell} A^*_{i_\ell i_k}$$

holds for all ℓ large enough and for all $k \geq \ell$. Since $t_k^{-1} = A^*_{bi_k}$, γ is an almost-geodesic in the sense of Rieffel. □

Rieffel called the limits of almost-geodesics in his sense *Busemann points.*

Corollary 7.13. *Let S be a proper metric space. Then the minimal Martin space is the disjoint union of $\mathscr{K}$ and of the set of Busemann points of S.*

Proof. Since $A^+_{ii} = -d(i,i) = 0$ for all i, the set $\mathscr{K}$ is included in the minimal Martin space $\mathscr{M}^m$. We next show that $\mathscr{M}^m \setminus \mathscr{K}$ is the set of Busemann points. Let $w \in \mathscr{M}$ be a Busemann point. By Proposition 7.12 we can find an almost-geodesic in our sense $i_0, i_1, \ldots$ such that $K_{\cdot i_k}$ converges to w and $d(b, i_k)$ is unbounded. We know from Proposition 7.3 that $w \in \mathscr{M}^m$. It remains to check that $w \notin \mathscr{K}$. To see this, we show that for all $z \in \mathscr{M}$,

$$\lim_{k\to\infty} H(K_{\cdot i_k}, z) = H(w, z) \ . \tag{23}$$

Indeed, for all $\beta > \mathbb{1}$, letting k tend to infinity in (21) and using (8), we get

$$\beta^{-1} \leq \pi_{i_\ell} w_{i_\ell} = H(K_{\cdot i_\ell}, w) \leq \mathbb{1} \ ,$$

for ℓ large enough. Hence, $\lim_{\ell\to\infty} H(K_{\cdot i_\ell}, w) = \mathbb{1}$. By Lemma 3.6, $z \geq H(w, z)w$. We deduce that $H(K_{\cdot i_\ell}, z) \geq H(w, z)H(K_{\cdot i_\ell}, w)$, and so $\liminf_{\ell\to\infty} H(K_{\cdot i_\ell}, z) \geq H(w, z)$. By definition of H, $\limsup_{\ell\to\infty} H(K_{\cdot i_\ell}, z) \leq \limsup_{K_{\cdot j}\to w} H(K_{\cdot j}, z) = H(w, z)$, which shows (23). Assume now that $w \in \mathscr{K}$, that is, $w = K_{\cdot j}$ for some $j \in S$, and let us apply (23) to $z = K_{\cdot b}$. We have $H(K_{\cdot i_k}, z) = A^*_{bi_k} A^*_{i_k b} = -2 \times d(b, i_k) \to -\infty$. Hence, $H(w, z) = -\infty$. But $H(w, z) = A^*_{bj} A^*_{jb} = -2 \times d(b, j) > -\infty$, which shows that $w \notin \mathscr{K}$.
Conversely, let $w \in \mathscr{M}^m \setminus \mathscr{K}$. By Proposition 7.8, w is the limit of an almost-geodesic in our sense. Observe that this almost-geodesic is unbounded. Otherwise, since S is proper, i_k would have a converging subsequence, and by continuity of the map $i \mapsto K_{\cdot i}$, we would have $w \in \mathscr{K}$, a contradiction. It follows from Proposition 7.12 that w is a Busemann point. ☐

8 Martin representation of harmonic vectors

Theorem 8.1 (Poisson-Martin representation of harmonic vectors). *Any element $u \in \mathscr{H}$ can be written as*

$$u = \bigoplus_{w\in\mathscr{M}^m} \nu(w)w \ , \tag{24}$$

with $\nu : \mathscr{M}^m \to \mathbb{R}_{\max}$, and necessarily,

$$\sup_{w\in\mathscr{M}^m} \nu(w) < +\infty \ .$$

Conversely, any $\nu : \mathscr{M}^m \to \mathbb{R}_{\max}$ satisfying the latter inequality defines by (24) an element u of $\mathscr{H}$. Moreover, given $u \in \mathscr{H}$, μ_u is the maximal ν satisfying (24).

Proof. Let $u \in \mathscr{H}$. Then u is also in $\mathscr{S}$ and so, from Lemma 3.6, we obtain that

$$u = \bigoplus_{w\in\mathscr{M}} \mu_u(w)w \geq \bigoplus_{w\in\mathscr{M}^m} \mu_u(w)w \ . \tag{25}$$

To show the opposite inequality, let us fix some $i \in S$ such that $u_i \neq \mathbb{0}$. Let us also fix some sequence $(\alpha_k)_{k\geq 0}$ in $\mathbb{R}_{\max}$ such that $\alpha_k > \mathbb{1}$ for all $k \geq 0$ and such that $\alpha := \alpha_0\alpha_1 \cdots < +\infty$. Since $u = Au$, one can construct a sequence $(i_k)_{k\geq 0}$ in S starting at $i_0 := i$, and such that

$$u_{i_k} \leq \alpha_k A_{i_k i_{k+1}} u_{i_{k+1}} \quad \text{for all } k \geq 0 \ .$$

Then,

$$u_{i_0} \leq \alpha A_{i_0 i_1} \cdots A_{i_{k-1} i_k} u_{i_k} \leq \alpha A^*_{i_0 i_k} u_{i_k} \quad \text{for all } k \geq 0 \ , \tag{26}$$

and so $(i_k)_{k\geq 0}$ is an α-almost-geodesic with respect to u. Since u is π-integrable, we deduce using Corollary 7.7 that $K_{\cdot i_k}$ converges to some $w \in \mathscr{M}^m$. From (26), we get $u_i \leq \alpha K_{i i_k} \pi_{i_k} u_{i_k}$, and letting k go to infinity, we obtain $u_i \leq \alpha w_i \mu_u(w)$. We thus obtain

$$u_i \leq \alpha \bigoplus_{w \in \mathscr{M}^m} \mu_u(w) w_i \ .$$

Since α can be chosen arbitrarily close to $\mathbb{1}$, we deduce the inequality opposite to (25), which shows that (24) holds with $\nu = \mu_u$.
The other parts of the theorem are proved in a manner similar to Theorem 6.1. □

Remark 8.2. The maximal representing measure μ_u at every point that is the limit of an almost geodesic can be computed by taking the limit of $\pi_i u_i$ along any almost-geodesic converging to this point. See Proposition 7.6.

In particular, $\mathscr{H} = \{\mathbb{0}\}$ if and only if $\mathscr{M}^m$ is empty. We now prove the analogue of Theorem 6.2 for harmonic vectors.

Theorem 8.3. *The normalised extremal generators of $\mathscr{H}$ are precisely the elements of $\mathscr{M}^m$.*

Proof. We know from Theorem 6.2 that each element of $\mathscr{M}^m$ is a normalised extremal generator of $\mathscr{S}$. Since $\mathscr{H} \subset \mathscr{S}$, and $\mathscr{M}^m \subset \mathscr{H}$ (by Proposition 4.4), this implies that each element of $\mathscr{M}^m$ is a normalised extremal generator of $\mathscr{H}$.
Conversely, by the same arguments as in the proof of Corollary 6.6, taking $\mathscr{F} = \mathscr{M}^m$ in Lemma 6.5 and using Theorem 8.1 instead of Lemma 3.6, we get that each normalised extremal generator ξ of $\mathscr{H}$ belongs to $\mathscr{M}^m \cup \mathscr{K}$. Since, by Proposition 3.2, no element of $\mathscr{K} \setminus \mathscr{M}^m$ can be harmonic, we have that $\xi \in \mathscr{M}^m$. □

Remark 8.4. Consider the situation when there are only finitely many recurrence classes and only finitely many non-recurrent nodes. Then $\mathscr{K}$ is a finite set, so that $\mathscr{B}$ is empty, $\mathscr{M} = \mathscr{K}$, and $\mathscr{M}^m$ coincides with the set of columns $K_{\cdot j}$ with j recurrent. The representation theorem (Theorem 8.1) shows in this case that each harmonic vector is a finite max-plus linear combination of the recurrent columns of A^*, as is the case in finite dimension.

9 Product Martin spaces

In this section, we study the situation where the set S is the Cartesian product of two sets, S_1 and S_2, and A and π can be decomposed as follows:

$$A = A_1 \otimes I_2 \oplus I_1 \otimes A_2 \ , \quad \pi = \pi_1 \otimes \pi_2 \ . \tag{27}$$

Here, $\otimes$ denotes the max-plus tensor product of matrices or vectors, A_i is an $S_i \times S_i$ matrix, π_i is a vector indexed by S_i, and I_i denotes the $S_i \times S_i$ max-plus identity matrix. For instance, $(A_1 \otimes I_2)_{(i_1,i_2),(j_1,j_2)} = (A_1)_{i_1 j_1}(I_2)_{i_2 j_2}$, which is equal to $(A_1)_{i_1 j_1}$ if $i_2 = j_2$, and to $\mathbb{0}$ otherwise. We shall always assume that π_i is left super-harmonic with respect to A_i, for $i = 1, 2$. We denote by $\mathscr{M}_i$ the corresponding Martin space, by K_i the corresponding Martin kernel, etc.
We introduce the map

$$\imath : \mathbb{R}_{\max}^{S_1} \times \mathbb{R}_{\max}^{S_2} \to \mathbb{R}_{\max}^{S},\ \imath(w_1, w_2) = w_1 \otimes w_2 \enspace ,$$

which is obviously continuous for the product topologies. The restriction of $\imath$ to the set of (w_1, w_2) such that $\pi_1 w_1 = \pi_2 w_2 = \mathbb{1}$ is injective. Indeed, if $w_1 \otimes w_2 = w_1' \otimes w_2'$, applying the operator $I_1 \otimes \pi_2$ on both sides of the equality, we get $w_1 \otimes \pi_2 w_2 = w_1' \otimes \pi_2 w_2'$, from which we deduce that $w_1 = w_1'$ if $\pi_2 w_2 = \pi_2 w_2' = \mathbb{1}$.

Proposition 9.1. *Assume that A and π are of the form (27), and that $\pi_i w_i = \mathbb{1}$ for all $w_i \in \mathscr{M}_i$ and $i = 1, 2$. Then, the map $\imath$ is a homeomorphism from $\mathscr{M}_1 \times \mathscr{M}_2$ to the Martin space $\mathscr{M}$ of A, and sends $\mathscr{K}_1 \times \mathscr{K}_2$ to $\mathscr{K}$. Moreover, the same map sends*

$$\mathscr{M}_1^m \times (\mathscr{K}_2 \cup \mathscr{M}_2^m) \ \cup\ (\mathscr{K}_1 \cup \mathscr{M}_1^m) \times \mathscr{M}_2^m$$

to the minimal Martin space $\mathscr{M}^m$ of A.

The proof of Proposition 9.1 relies on several lemmas.

Lemma 9.2. *If A is given by (27), then, $A^* = A_1^* \otimes A_2^*$ and*

$$A^+ = A_1^+ \otimes A_2^* \oplus A_1^* \otimes A_2^+ \enspace .$$

Proof. Summing the equalities $A^k = \bigoplus_{1 \le \ell \le k} A_1^\ell \otimes A_2^{k-\ell}$, we obtain $A^* = A_1^* \otimes A_2^*$. Hence, $A^+ = AA^* = (A_1 \otimes I_2 \oplus I_1 \otimes A_2)(A_1^* \otimes A_2^*) = A_1^+ \otimes A_2^* \oplus A_1^* \otimes A_2^+$. □

We define the kernel $H \circ \imath$ from $(\mathscr{M}_1 \times \mathscr{M}_2)^2$ to $\mathbb{R}_{\max}$, by $H \circ \imath((z_1, z_2), (w_1, w_2)) = H(\imath(z_1, z_2), \imath(w_1, w_2))$. The kernel $H^\flat \circ \imath$ is defined from $H^\flat$ in the same way.

Lemma 9.3. *If $A^* = A_1^* \otimes A_2^*$ and $\pi = \pi_1 \otimes \pi_2$, then $\mathscr{K} = \imath(\mathscr{K}_1 \times \mathscr{K}_2)$ and $\imath(\mathscr{M}_1 \times \mathscr{M}_2) = \mathscr{M}$. Moreover, if $\pi_i w_i = \mathbb{1}$ for all $w_i \in \mathscr{M}_i$ and $i = 1, 2$, then $\imath$ is an homeomorphism from $\mathscr{M}_1 \times \mathscr{M}_2$ to $\mathscr{M}$, and $H \circ \imath = H_1 \otimes H_2$.*

Proof. Observe that $K = K_1 \otimes K_2$. Hence, $\mathscr{K} = \imath(\mathscr{K}_1 \times \mathscr{K}_2)$. Let $\overline{X}$ denote the closure of any set X. Since $\overline{\mathscr{K}_i} = \mathscr{M}_i$, we get $\overline{\mathscr{K}_1 \times \mathscr{K}_2} = \mathscr{M}_1 \times \mathscr{M}_2$, and so $\overline{\mathscr{K}_1 \times \mathscr{K}_2}$ is compact. Since $\imath$ is continuous, we deduce that $\imath(\overline{\mathscr{K}_1 \times \mathscr{K}_2}) = \overline{\imath(\mathscr{K}_1 \times \mathscr{K}_2)}$. Hence, $\imath(\mathscr{M}_1 \times \mathscr{M}_2) = \overline{\mathscr{K}} = \mathscr{M}$. Assume now that $\pi_i w_i = \mathbb{1}$ for all $w_i \in \mathscr{M}_i$ and $i = 1, 2$, so that the restriction of $\imath$ to $\mathscr{M}_1 \times \mathscr{M}_2$ is injective. Since $\mathscr{M}_1 \times \mathscr{M}_2$ is compact, we deduce that $\imath$ is an homeomorphism from $\mathscr{M}_1 \times \mathscr{M}_2$ to its image, that is, $\mathscr{M}$. Finally, let $z = \imath(z_1, z_2)$ and $w = \imath(w_1, w_2)$, with

$z_1, w_1 \in \mathscr{M}_1$ and $z_2, w_2 \in \mathscr{M}_2$. Since $\imath$ is an homeomorphism from $\mathscr{M}_1 \times \mathscr{M}_2$ to $\mathscr{M}$, we can write $H(z,w)$ in terms of limsup and limit for the product topology of $\mathscr{M}_1 \times \mathscr{M}_2$:

$$H(z,w) = \limsup_{\substack{(K_1)_{\cdot i_1} \to z_1 \\ (K_2)_{\cdot i_2} \to z_2}} \lim_{\substack{(K_1)_{\cdot j_1} \to w_1 \\ (K_2)_{\cdot j_2} \to w_2}} \pi_{(i_1,i_2)} K_{(i_1,i_2),(j_1,j_2)} \ . \tag{28}$$

Since $A^* = A_1^* \otimes A_2^*$ and $\pi = \pi_1 \otimes \pi_2$, we can write the right hand side term of (28) as the product of two terms that are both bounded from above:

$$\pi_{(i_1,i_2)} K_{(i_1,i_2),(j_1,j_2)} = ((\pi_1)_{i_1} (K_1)_{i_1,j_1}) ((\pi_2)_{i_2} (K_2)_{i_2,j_2}) \ .$$

Hence, the limit and limsup in (28) become a product of limits and limsups, respectively, and so $H(z,w) = H_1(z_1,w_1) H_2(z_2,w_2)$. □

Lemma 9.4. *Assume that A and π are of the form (27) and that $\pi_i w_i = \mathbb{1}$ for all $w_i \in \mathscr{M}_i$ and $i = 1,2$. Then*

$$H^\flat \circ \imath = H_1^\flat \otimes H_2 \oplus H_1 \otimes H_2^\flat \ . \tag{29}$$

Proof. By Lemma 9.2, $A^+ = A_1^+ \otimes A_2^* \oplus A_1^* \otimes A_2^+$, and so

$$K^\flat = K_1^\flat \otimes K_2 \oplus K_1 \otimes K_2^\flat \ .$$

Let $z = \imath(z_1,z_2)$ and $w = \imath(w_1,w_2)$, with $z_1, w_1 \in \mathscr{M}_1$, $z_2, w_2 \in \mathscr{M}_2$. In a way similar to (28), we can write $H^\flat$ as

$$H^\flat(z,w) = \limsup_{\substack{(K_1)_{\cdot i_1} \to z_1 \\ (K_2)_{\cdot i_2} \to z_2}} \liminf_{\substack{(K_1)_{\cdot j_1} \to w_1 \\ (K_2)_{\cdot j_2} \to w_2}} \pi_{(i_1,i_2)} K^\flat_{(i_1,i_2),(j_1,j_2)} \ .$$

The right hand side term is a sum of products:

$$\begin{aligned} &\pi_{(i_1,i_2)} K^\flat_{(i_1,i_2),(j_1,j_2)} \\ = \quad & (\pi_1)_{i_1} (K_1^\flat)_{i_1 j_1} (\pi_2)_{i_2} (K_2)_{i_2 j_2} \oplus (\pi_1)_{i_1} (K_1)_{i_1 j_1} (\pi_2)_{i_2} (K_2^\flat)_{i_2 j_2} \ . \end{aligned}$$

We now use the following two general observations. Let $(\alpha_d)_{d\in D}$, $(\beta_e)_{e\in E}$, $(\gamma_d)_{d\in D}$, $(\delta_e)_{e\in E}$ be nets of elements of $\mathbb{R}_{\max}$ that are bounded from above. Then,

$$\limsup_{d,e} \alpha_d \beta_e \oplus \gamma_d \delta_e = (\limsup_d \alpha_d)(\limsup_e \beta_e) \oplus (\limsup_d \gamma_d)(\limsup_e \delta_e) \ .$$

If additionally the nets $(\beta_e)_{e\in E}$ and $(\gamma_d)_{d\in D}$ converge, we have

$$\liminf_{d,e} \alpha_d \beta_e \oplus \gamma_d \delta_e = (\liminf_d \alpha_d)(\lim_e \beta_e) \oplus (\lim_d \gamma_d)(\liminf_e \delta_e) \ .$$

Using both identities, we deduce that $H^\flat$ is given by (29). □

Proof of Proposition 9.1. We know from Lemma 9.2 that $A^* = A_1^* \otimes A_2^*$, and so, by Lemma 9.3, $\imath$ is an homeomorphism from $\mathscr{M}_1 \times \mathscr{M}_2$ to $\mathscr{M}$. Since the kernels $H_1, H_1^\flat, H_2$ and $H_2^\flat$ all take values less than or equal to $\mathbb{1}$, we conclude from (29) that, when $z = \imath(z_1, z_2)$, $H^\flat(z,z) = \mathbb{1}$ if and only if $H_1^\flat(z_1,z_1) = H_2(z_2,z_2) = \mathbb{1}$ or $H_1(z_1,z_1) = H_2^\flat(z_2,z_2) = \mathbb{1}$. Using Equation (12) and the definition of the minimal Martin space, we deduce that

$$\mathscr{M}^m = \imath\big(\mathscr{M}_1^m \times (\mathscr{K}_2 \cup \mathscr{M}_2^m) \ \cup\ (\mathscr{K}_1 \cup \mathscr{M}_1^m) \times \mathscr{M}_2^m\big)\ . \qquad \square$$

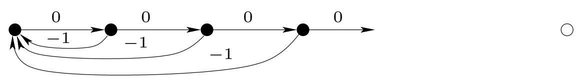
Remark 9.5. The assumption that $\pi_i w_i = \mathbb{1}$ for all $w_i \in \mathscr{M}_i$ is automatically satisfied when the left super-harmonic vectors π_i originate from basepoints, that is, when $\pi_i = (A_i)^*_{b_i,\cdot}$ for some basepoint b_i. Indeed, we already observed in the proof of Proposition 5.1 that every vector $w_i \in \mathscr{M}_i$ satisfies $(\pi_i)_{b_i}(w_i)_{b_i} = \mathbb{1}$. By (5), $\pi_i w_i \leq \mathbb{1}$. We deduce that $\pi_i w_i = \mathbb{1}$.

Remark 9.6. Rieffel [Rie02, Prop. 4.11] obtained a version of the first part of Lemma 9.3 for metric spaces. His result states that if (S_1, d_1) and (S_2, d_2) are locally compact metric spaces, and if their product S is equipped with the sum of the metrics, $d((i_1,i_2),(j_1,j_2)) = d_1(i_1,j_1) + d_2(i_2,j_2)$, then the metric compactification of S can be identified with the Cartesian product of the metric compactifications of S_1 and S_2. This result can be re-obtained from Lemma 9.3 by taking $(A_1)_{i_1,j_i} = -d_1(i_1,j_1)$, $(A_2)_{i_2,j_2} = -d_2(i_2,j_2)$, $\pi_{i_1} = -d_1(i_1,b_1)$, and $\pi_{i_2} = -d(i_2,b_2)$, for arbitrary basepoints $b_1, b_2 \in \mathbb{Z}$. We shall illustrate this in Example 10.4.

10 Examples and Counter-Examples

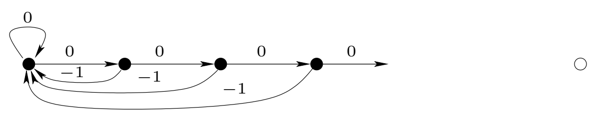
We now illustrate our results and show various features that the Martin space may have.

Example 10.1. Let $S = \mathbb{N}$, $A_{i,i+1} = 0$ for all $i \in \mathbb{N}$, $A_{i,0} = -1$ for all $i \in \mathbb{N}\setminus\{0\}$ and $A_{ij} = -\infty$ elsewhere. We choose the basepoint 0, so that $\pi = A^*_{0,\cdot}$. The graph of A is:



States (elements of S) are represented by black dots. The white circle represents the extremal boundary element ξ, that we next determine. In this example, $\rho(A) = \mathbb{1}$, and A has no recurrent class. We have $A^*_{ij} = \mathbb{1}$ for $i \leq j$ and $A^*_{ij} = -1$ for $i > j$, so the Martin space of A corresponding to $\pi = A^*_{0\cdot}$ consists of the columns $A^*_{\cdot j}$, with $j \in \mathbb{N}$, together with the vector ξ whose entries are all equal to $\mathbb{1}$. We have $\mathscr{B} = \{\xi\}$. One can easily check that $H(\xi,\xi) = \mathbb{1}$. Therefore, $\mathscr{M}^m = \{\xi\}$. Alternatively, we may use Proposition 7.3 to show that $\xi \in \mathscr{M}^m$, since ξ is the limit of the almost-geodesic $0, 1, 2, \ldots$. Theorem 8.1 says that ξ is the unique (up to a multiplicative constant) non-zero harmonic vector.

Example 10.2. Let us modify Example 10.1 by setting $A_{00} = 0$, so that the previous graph becomes:



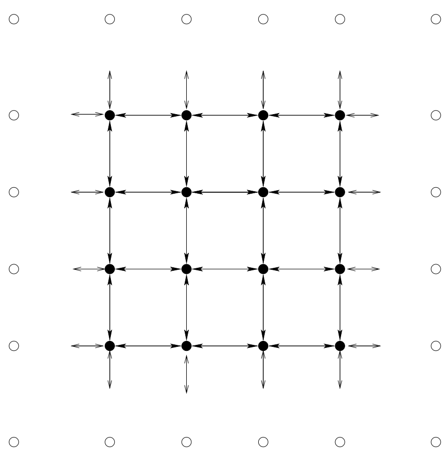
We still have $\rho(A) = \mathbb{1}$, the node 0 becomes recurrent, and the minimal Martin space is now $\mathscr{M}^m = \{K_{\cdot 0}, \xi\}$, where ξ is defined in Example 10.1. Theorem 8.1 says that every harmonic vector is of the form $\alpha K_{\cdot 0} \oplus \beta\xi$, that is $\sup(\alpha + K_{\cdot 0}, \beta + \xi)$ with the notation of classical algebra, for some $\alpha, \beta \in \mathbb{R} \cup \{-\infty\}$.

Example 10.3. Let $S = \mathbb{Z}$, $A_{i,i+1} = A_{i+1,i} = -1$ for $i \in \mathbb{Z}$, and $A_{ij} = \mathbb{0}$ elsewhere. We choose 0 to be the basepoint, so that $\pi = A^*_{0,\cdot}$. The graph of A is:

We are using the same conventions as in the previous examples, together with the following additional conventions: the arrows are bidirectional since the matrix is symmetric, and each arc has weight -1 unless otherwise specified. This example and the next were considered by Rieffel [Rie02].

We have $\rho(A) = -1 < \mathbb{1}$, which implies there are no recurrent nodes. We have $A^*_{i,j} = -|i-j|$, and so $K_{i,j} = |j| - |i-j|$. There are two Martin boundary points, $\xi^+ = \lim_{j\to\infty} K_{\cdot j}$ and $\xi^- = \lim_{j\to-\infty} K_{\cdot j}$, which are given by $\xi^+_i = i$ and $\xi^-_i = -i$. Thus, the Martin space $\mathscr{M}$ is homeomorphic to $\overline{\mathbb{Z}} := \mathbb{Z} \cup \{\pm\infty\}$ equipped with the usual topology. Since both ξ^+ and ξ^- are limits of almost-geodesics, $\mathscr{M}^m = \{\xi^+, \xi^-\}$. Theorem 8.1 says that every harmonic vector is of the form $\alpha\xi^+ \oplus \beta\xi^-$, for some $\alpha, \beta \in \mathbb{R}_{\max}$.

Example 10.4. Consider $S := \mathbb{Z}\times\mathbb{Z}$ and the operator A given by $A_{(i,j),(i,j\pm1)} = -1$ and $A_{(i,j),(i\pm1,j)} = -1$, for each $i, j \in \mathbb{Z}$, with all other entries equal to $-\infty$. We choose the basepoint $(0,0)$. We represent the graph of A with the same conventions as in Example 10.3:



For all $i, j, k, l \in \mathbb{Z}$,

$$A^*_{(i,j),(k,l)} = -|i-k| - |j-l| \enspace .$$

Note that this is the negative of the distance in the ℓ_1 norm between (i,j) and (k,l). The matrix A can be decomposed as $A = A_1 \otimes I \oplus I \otimes A_2$, where A_1, A_2 are two copies of the matrix of Example 10.3, and I denotes the $\mathbb{Z} \times \mathbb{Z}$ identity matrix (recall that $\otimes$ denotes the tensor product of matrices, see Section 9 for details). The vector π can be written as $\pi_1 \otimes \pi_2$, with $\pi_1 = (A_1)^*_{0,\cdot}$ and $\pi_2 = (A_2)^*_{0,\cdot}$. Hence, Proposition 9.1 shows that the Martin space of A is homeomorphic to the Cartesian product of two copies of the Martin space of Example 10.3, in other words, that there is an homeomorphism from $\mathscr{M}$ to $\overline{\mathbb{Z}} \times \overline{\mathbb{Z}}$. Proposition 9.1 also shows that the same homeomorphism sends $\mathscr{K}$ to $\mathbb{Z} \times \mathbb{Z}$ and the minimal Martin space to $(\{\pm\infty\} \times \overline{\mathbb{Z}}) \cup (\overline{\mathbb{Z}} \times \{\pm\infty\})$. Thus, the Martin boundary and the minimal Martin space are the same. This example may be considered to be the max-plus analogue of the random walk on the 2-dimensional integer lattice. The Martin boundary for the latter (with respect to eigenvalues strictly greater than the spectral radius) is known [NS66] to be the circle.

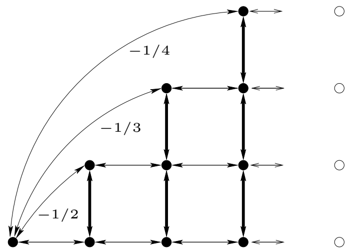
Example 10.5. Let $S = \mathbb{Q}$ and $A_{ij} = -|i-j|$. Choosing 0 to be the basepoint, we get $K_{ij} = -|i-j| + |j|$ for all $j \in \mathbb{Q}$. The Martin boundary $\mathscr{B}$ consists of the functions $i \mapsto -|i-j| + |j|$ with $j \in \mathbb{R} \setminus \mathbb{Q}$, together with the functions $i \mapsto i$ and $i \mapsto -i$. The Martin space $\mathscr{M}$ is homeomorphic to $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ equipped with its usual topology.

Example 10.6. We give an example of a complete locally compact metric space (S,d) such that the canonical injection from S to the Martin space $\mathscr{M}$ is not an embedding, and such that the Martin boundary $\mathscr{B} = \mathscr{M} \setminus \mathscr{K}$ is not closed. Consider $S = \{(i,j) \mid i \geq j \geq 1\}$ and the operator A given by

$$A_{(i,j),(i+1,j)} = A_{(i+1,j),(i,j)} = -1, \text{ for } i \geq j \geq 1,$$

$$A_{(i,j),(i,j+1)} = A_{(i,j+1),(i,j)} = -2, \text{ for } i-1 \geq j \geq 1,$$
$$A_{(1,1),(i,i)} = A_{(i,i),(0,0)} = -1/i, \text{ for } i \geq 2,$$

with all other entries equal to $-\infty$. We choose the basepoint $(1,1)$. The graph of A is depicted in the following diagram:



We are using the same conventions as before. The arcs with weight -2 are drawn in bold. One can check that

$$A^*_{(i,j),(k,\ell)} = \max\big(-|i-k| - 2|j-\ell|, -(i-j)-(k-\ell)-\phi(j)-\phi(\ell)\big)$$

where $\phi(j) = 1/j$ if $j \geq 2$, and $\phi(j) = 0$ if $j = 1$. In other words, an optimal path from (i,j) to (k,ℓ) is either an optimal path for the metric of the weighted ℓ_1 norm $(i,j) \mapsto |i| + 2|j|$, or a path consisting of an horizontal move to the diagonal point (j,j), followed by moves from (j,j) to $(1,1)$, from $(1,1)$ to (ℓ,ℓ), and by an horizontal move from (ℓ,ℓ) to (k,ℓ). Since A is symmetric and A^* is zero only on the diagonal, $d((i,j),(k,\ell)) := -A^*_{(i,j),(k,\ell)}$ is a metric on S. The metric space (S,d) is complete since any Cauchy sequence is either ultimately constant or converges to the point $(1,1)$. It is also locally compact since any point distinct from $(1,1)$ is isolated, whereas the point $(1,1)$ has the basis of neighbourhoods consisting of the compact sets $V_j = \{(i,i) \mid i \geq j\} \cup \{(1,1)\}$, for $j \geq 2$.

If $((i_m, j_m))_{m\geq 1}$ is any sequence of elements of S such that both i_m and j_m tend to infinity, then, for any $(k,\ell) \in S$,

$$A^*_{(k,\ell),(i_m,j_m)} = A^*_{(k,\ell),(1,1)} A^*_{(1,1),(i_m,j_m)} \qquad \text{for } m \text{ large enough.}$$

(Intuitively, this is related to the fact that, for m large enough, every optimal path from (k,ℓ) to (i_m,j_m) passes through the point $(1,1)$). It follows that $K_{\cdot,(i_m,j_m)}$ converges to $K_{\cdot,(1,1)}$ as $m \to \infty$. However, the sequence (i_m,j_m) does not converge to the point $(1,1)$ in the metric topology unless $i_m = j_m$ for m large enough. This shows that the map $(i,j) \to K_{\cdot,(i,j)}$ is not an homeomorphism from S to its image.

The Martin boundary consists of the points $\xi^1, \xi^2, \ldots$, obtained as limits of horizontal half-lines, which are almost-geodesics. We have

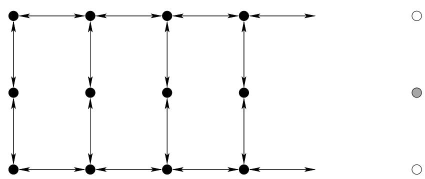
$$\xi^\ell_{(i,j)} := \lim_{k\to\infty} K_{(i,j),(k,\ell)} = \max\big(i - \ell - 2|j-\ell| + \phi(\ell), -(i-j) - \phi(j)\big) \ .$$

The functions ξ^ℓ are all distinct because $i \mapsto \xi^\ell_{(i,i)}$ has a unique maximum attained at $i = \ell$. The functions ξ^ℓ do not belong to $\mathscr{K}$ because $\xi^\ell_{(3j,j)} = j + \ell + \phi(\ell) \sim j$ as j tends to infinity, whereas for any $w \in \mathscr{K}$, $w_{(3j,j)} = -2j - \phi(j) \sim -2j$ as j tends to infinity,. The sequence ξ^ℓ converges to $K_{\cdot,(1,1)}$ as ℓ tends to infinity, which shows that the Martin boundary $\mathscr{B} = \mathscr{M} \setminus \mathscr{K}$ is not closed.

Example 10.7. We next give an example of a Martin space having a boundary point which is not an extremal generator. The same example has been found independently by Webster and Winchester [WW06]. Consider $S := \mathbb{N} \times \{0, 1, 2\}$ and the operator A given by

$$A_{(i,j),(i+1,j)} = A_{(i+1,j),(i,j)} = A_{(i,1),(i,j)} = A_{(i,j),(i,1)} = -1,$$

for all $i \in \mathbb{N}$ and $j \in \{0, 2\}$, with all other entries equal to $-\infty$. We choose $(0, 1)$ as basepoint, so that $\pi := A^*_{(0,1),\cdot}$ is such that $\pi_{(i,j)} = -(i + 1)$ if $j = 0$ or 2, and $\pi_{(i,j)} = -(i + 2)$ if $j = 1$ and $i \neq 0$. The graph associated to the matrix A is depicted in the following diagram, with the same conventions as in the previous example.



There are three boundary points. They may be obtained by taking the limits

$$\xi^0 := \lim_{i\to\infty} K_{\cdot,(i,0)}, \qquad \xi^1 := \lim_{i\to\infty} K_{\cdot,(i,1)}, \qquad \text{and } \xi^2 := \lim_{i\to\infty} K_{\cdot,(i,2)}.$$

Calculating, we find that

$$\xi^0_{(i,j)} = i - j + 1, \qquad \xi^2_{(i,j)} = i + j - 1, \qquad \text{and } \xi^1 = \xi^0 \oplus \xi^2.$$

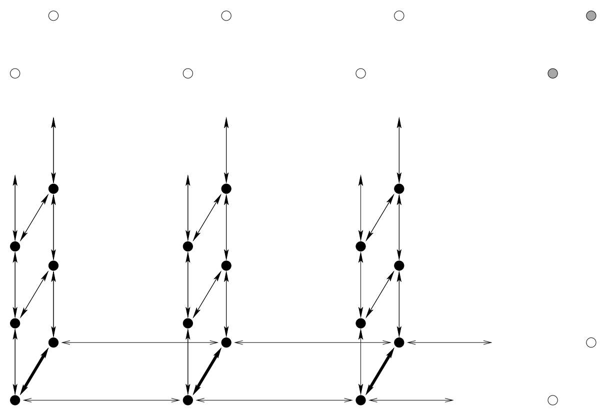
We have $H(\xi^0, \xi^0) = H(\xi^2, \xi^2) = H(\xi^2, \xi^1) = H(\xi^0, \xi^1) = 0$. For all other pairs $(\xi', \xi) \in \mathscr{B} \times \mathscr{B}$, we have $H(\xi', \xi) = -2$. Therefore, the minimal Martin boundary is $\mathscr{M}^m = \{\xi^0, \xi^2\}$, and there is a non-extremal boundary point, ξ^1, represented above by a gray circle. The sequences $((i, 0))_{i\in\mathbb{N}}$ and $((i, 2))_{i\in\mathbb{N}}$ are almost-geodesics, while it should be clear from the diagram that there are no almost-geodesics converging to ξ^1. So this example provides an illustration of Propositions 7.3 and 7.8.

Example 10.8. Finally, we will give an example of a non-compact minimal

Martin space. Consider $S := \mathbb{N} \times \mathbb{N} \times \{0,1\}$ and the operator A given by

$$\begin{aligned} A_{(i,j,k),(i,j+1,k)} = A_{(i,j+1,k),(i,j,k)} &= -1, \text{ for all } i,j \in \mathbb{N} \text{ and } k \in \{0,1\}, \\ A_{(i,j,k),(i,j,1-k)} &= -1, \text{ for all } i \in \mathbb{N}, j \in \mathbb{N}\backslash\{0\} \text{ and } k \in \{0,1\}, \\ A_{(i,0,k),(i,0,1-k)} &= -2, \text{ for all } i \in \mathbb{N} \text{ and } k \in \{0,1\}, \\ A_{(i,0,k),(i+1,0,k)} = A_{(i+1,0,k),(i,0,k)} &= -1, \text{ for all } i \in \mathbb{N} \text{ and } k \in \{0,1\}, \end{aligned}$$

with all other entries equal to $-\infty$. We take $\pi := A^*_{(0,0,0),\cdot}$. With the same conventions as in Examples 10.4 and 10.7, the graph of A is



Recall that arcs of weight -1 are drawn with thin lines whereas those of weight -2 are drawn in bold.
For all $(i,j,k),(i',j',k') \in S$,

$$A^*_{(i,j,k),(i',j',k')} = -|k'-k| - |i'-i| - |j'-j|\chi_{i=i'} - (j+j')\chi_{i\neq i'} - \chi_{j=j'=0,\, k\neq k'} \quad ,$$

where χ_E takes the value 1 when condition E holds, and 0 otherwise. Hence,

$$\begin{aligned} K_{(i,j,k),(i',j',k')} =& k' - |k'-k| + i' - |i'-i| + j' - |j'-j|\chi_{i=i'} - (j+j')\chi_{i\neq i'} \\ &+ \chi_{j'=0,k'=1} - \chi_{j=j'=0,\, k\neq k'} \quad . \end{aligned}$$

By computing the limits of $K_{\cdot,(i',j',k')}$ when i' and/or j' go to $+\infty$, we readily check that the Martin boundary is composed of the vectors

$$\begin{aligned} \xi^{i',\infty,k'} &:= \lim_{j'\to\infty} K_{\cdot,(i',j',k')}, \\ \xi^{\infty,\infty,k'} &:= \lim_{i',j'\to\infty} K_{\cdot,(i',j',k')} \\ \xi^{\infty,0,k'} &:= \lim_{i'\to\infty} K_{\cdot,(i',0,k')}. \end{aligned}$$

where the limit in i and j' in the second line can be taken in either order. Note that $\lim_{i'\to\infty} K_{\cdot,(i',j',k')} = \xi^{\infty,\infty,k'}$ for any $j' \in \mathbb{N}\setminus\{0\}$ and $k' \in \{0,1\}$. The minimal Martin space is composed of the vectors $\xi^{i',\infty,k'}$ and $\xi^{\infty,0,k'}$ with $i' \in \mathbb{N}$ and $k' \in \{0,1\}$. The two boundary points $\xi^{\infty,\infty,0}$ and $\xi^{\infty,\infty,1}$ are non-extremal and have representations

$$\xi^{\infty,\infty,0} = \xi^{\infty,0,0} \oplus -3\xi^{\infty,0,1} \ ,$$
$$\xi^{\infty,\infty,1} = \xi^{\infty,0,0} \oplus -1\xi^{\infty,0,1}.$$

For $k' \in \{0,1\}$, the sequence $(\xi^{i',\infty,k'})_{i\in\mathbb{N}}$ converges to $\xi^{\infty,\infty,k'}$ as i goes to infinity. Since this point is not in $\mathscr{M}^m$, we see that $\mathscr{M}^m$ is not compact.

11 Tightness and existence of harmonic vectors

We now show how the Martin boundary can be used to obtain existence results for eigenvectors. As in [AGW05], we restrict our attention to the case where S is equipped with the discrete topology. We say that a vector $u \in \mathbb{R}^S_{\max}$ is *A-tight* if, for all $i \in S$ and $\beta \in \mathbb{R}$, the super-level set $\{j \in S \mid A_{ij}u_j \geq \beta\}$ is finite. We say that a family of vectors $\{u^\ell\}_{\ell\in L} \subset \mathbb{R}^S_{\max}$ is A-tight if $\sup_{\ell\in L} u^\ell$ is A-tight. The notion of tightness is motivated by the following property.

Lemma 11.1. *If a net $\{u^\ell\}_{\ell\in L} \subset \mathbb{R}^S_{\max}$ is A-tight and converges pointwise to u, then Au^ℓ converges pointwise to Au.*

Proof. This may be checked elementarily, or obtained as a special case of general results for idempotent measures [Aki95, AQV98, Aki99, Puh01] or, even more generally, capacities [OV91]. We may regard u and u^l as the densities of the idempotent measures defined by

$$Q_u(J) = \sup_{j\in J} u_j \qquad \text{and} \qquad Q_{u^l}(J) = \sup_{j\in J} u^l_j \ ,$$

for any $J \subset S$. When S is equipped with the discrete topology, pointwise convergence of $(u^\ell)_{\ell\in L}$ is equivalent to convergence in the hypograph sense of convex analysis. It is shown in [AQV98] that this is then equivalent to convergence of $(Q_{u^l})_{\ell\in L}$ in a sense analogous to the vague convergence of probability theory. It is also shown that, when combined with the tightness of $(u_l)_{\ell\in L}$, this implies convergence in a sense analogous to weak convergence. The result follows as a special case. □

Proposition 11.2. *Assume that S is infinite and that the vector $\pi^{-1} := (\pi_i^{-1})_{i\in S}$ is A-tight. Then, some element of $\mathscr{M}$ is harmonic and, if $\mathbb{0} \notin \mathscr{M}$, then $\mathscr{M}^m$ is non-empty. Furthermore, each element of $\mathscr{B}$ is harmonic.*

Proof. Since S is infinite, there exists an injective map $n \in \mathbb{N} \mapsto i_n \in S$. Consider the sequence $(i_n)_{n\in\mathbb{N}}$. Since $\mathscr{M}$ is compact, it has a subnet $(j_k)_{k\in D}$,

$j_k := i_{n_k}$ such that $\{K_{\cdot j_k}\}_{k\in K}$ converges to some $w \in \mathscr{M}$. Let $i \in S$. Since $(AA^*)_{ij} = A^+_{ij} = A^*_{ij}$ for all $j \neq i$, we have

$$(AK_{\cdot j_k})_i = K_{ij_k}$$

when $j_k \neq i$. But, by construction, the net $(j_k)_{k\in D}$ is eventually in $S\backslash\{i\}$ and so we may pass to the limit, obtaining $\lim_{k\in K} AK_{\cdot j_k} = w$. Since π^{-1} is A-tight, it follows from (4) that the family $(K_{\cdot j})_{j\in S}$ is A-tight. Therefore, by Lemma 11.1, we get $w = Aw$. If $\mathbb{0} \notin \mathscr{M}$, then $\mathscr{H}$ contains a non-zero vector, and applying the representation formula (24) to this vector, we see that $\mathscr{M}^m$ cannot be empty.

It remains to show that $\mathscr{B} \subset \mathscr{H}$. Any $w \in \mathscr{B}$ is the limit of a net $\{K_{\cdot j_k}\}_{k\in D}$. Let $i \in S$. Since $w \neq K_{\cdot i}$, the net $\{K_{\cdot j_k}\}_{k\in D}$ is eventually in some neighbourhood of w not containing $K_{\cdot i}$. We deduce as before that w is harmonic. □

Corollary 11.3 (Existence of harmonic vectors). *Assume that S is infinite, that $\pi = A^*_{b\cdot} \in \mathbb{R}^S$ for some $b \in S$, and that π^{-1} is A-tight. Then, $\mathscr{H}$ contains a non-zero vector.*

Proof. We have $K_{bj} = \mathbb{1}$ for all $j \in S$ and hence, by continuity, $w_b = \mathbb{1}$ for all $w \in \mathscr{M}$. In particular, $\mathscr{M}$ does not contain $\mathbb{0}$. The result follows from an application of the proposition. □

We finally derive a characterisation of the spectrum of A. We say that λ is a *(right)-eigenvalue* of A if $Au = \lambda u$ for some vector u such that $u \neq \mathbb{0}$.

Corollary 11.4. *Assume that S is infinite, A is irreducible, and for each $i \in S$, there are only finitely many $j \in S$ with $A_{ij} > \mathbb{0}$. Then the set of right eigenvalues of A is $[\rho(A), \infty[$.*

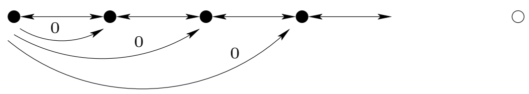
Proof. Since A is irreducible, no eigenvector of A can have a component equal to $\mathbb{0}$. It follows from [Dud92, Prop. 3.5] that every eigenvalue of A must be greater than or equal to $\rho(A)$.

Conversely, for all $\lambda \geq \rho(A)$, we have $\rho(\lambda^{-1}A) \leq \mathbb{1}$. Combined with the irreducibility of A, this implies [AGW05, Proposition 2.3] that all the entries of $(\lambda^{-1}A)^*$ are finite. In particular, for any $b \in S$, the vector $\pi := (\lambda^{-1}A)^*_{b\cdot}$ is in $\mathbb{R}^S$. The last of our three assumptions ensures that π^{-1} is $(\lambda^{-1}A)$-tight and so, by Corollary 11.3, $(\lambda^{-1}A)$ has a non-zero harmonic vector. This vector will necessarily be an eigenvector of A with eigenvalue λ. □

Example 11.5. The following example shows that when π^{-1} is not A-tight, a Martin boundary point need not be an eigenvector. Consider $S := \mathbb{N}$ and the operator A given by

$$A_{i,i+1} = A_{i+1,i} := -1 \qquad \text{and} \qquad A_{0i} := 0 \qquad \text{for all } i \in \mathbb{N},$$

with all other entries of equal to $-\infty$. We take $\pi := A^*_{0,\cdot}$. With the same conventions as in Example 10.7, the graph of A is



We have $A^*_{i,j} = \max(-i, -|i-j|)$ and $\pi_i = 0$ for all $i, j \in \mathbb{N}$. There is only one boundary point, $b := \lim_{k\to\infty} K_{\cdot k}$, which is given by $b_i = -i$ for all $i \in \mathbb{N}$. One readily checks that b is not an harmonic vector and, in fact, A has no non-zero harmonic vectors.

12 Eigenvectors of Lax-Oleinik semigroups and Busemann points of normed spaces

We now use the Martin boundary to solve a class of continuous-time deterministic optimal control problems. Consider the value function v defined by:

$$v(t,x) := \sup_{X(\cdot),\, X(0)=x} \phi(X(t)) - \int_0^t L(\dot{X}(s))\,\mathrm{d}s \enspace .$$

Here, x is a point in $\mathbb{R}^n$, t is a nonnegative real number, the *Lagrangian* L is a Borel measurable map $\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, bounded from below, the *terminal reward* ϕ is an arbitrary map $\mathbb{R}^n \to \mathbb{R}\cup\{-\infty\}$, and the supremum is taken over all absolutely continuous functions $X : [0,t] \to \mathbb{R}^n$ such that $X(0) = x$. This is a special case of the classical Lagrange problem of calculus of variations.
The *Lax-Oleinik semigroup* $(T^t)_{t\geq 0}$ is composed of the maps T^t sending the value function at time 0, $v(0,\cdot) = \phi$ to the value function at time t, $v(t,\cdot)$. The semigroup property $T^{t+s} = T^t \circ T^s$ follows from the dynamic programming principle. The kernel of the operator T^t is given by

$$(x,y) \mapsto T^t_{x,y} = \sup_{X(\cdot),\, X(0)=x,\, X(t)=y} - \int_0^t L(\dot{X}(s))\,\mathrm{d}s \enspace ,$$

where the supremum is taken over all absolutely continuous functions $X : [0,t] \to \mathbb{R}^n$ such that $X(0) = x$ and $X(t) = y$.
The classical Hopf-Lax formula states that

$$T^t_{x,y} = -t \operatorname{co} L\Big(\frac{y-x}{t}\Big), \qquad \text{for } t > 0 \enspace ,$$

where $\operatorname{co} L$ denotes the convex lower semicontinuous hull of L. This is proved, for instance, in [Eva98, §3.3, Th. 4] when L is convex and finite valued, and when the curves $X(\cdot)$ are required to be continuously differentiable. The extension to the present setting is not difficult.
Since T^t only depends on $\operatorname{co} L$, we shall assume that L is convex, lower semicontinuous, and bounded from below. Moreover, we shall always assume that $L(0)$ is finite.

We say that a function $u : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$, not identically $-\infty$, is an *eigenvector* of the semigroup $(T^t)_{t\geq 0}$ with *eigenvalue* λ if

$$T^t u = u + \lambda t, \quad \text{for all } t > 0 \ .$$

We shall say that u is *extremal* if it is an extremal generator of the eigenspace of the semigroup $(T^t)_{t\geq 0}$ with eigenvalue λ, meaning that u cannot be written as the supremum of two eigenvectors with the same eigenvalue that are both different from it.

One easily checks, using the convexity of L, that for all $t > 0$, the maximal circuit mean of the operator T^t is given by

$$\rho(T^t) = -tL(0) \ .$$

By Proposition 3.5 of [Dud92] or Lemma 2.2 of [AGW05], any eigenvalue μ of T^t must satisfy $\mu \geq \rho(T^t)$, and so any eigenvalue λ of the semigroup $(T^t)_{t\geq 0}$ satisfies

$$\lambda \geq -L(0) \ .$$

We denote by $\zeta(x)$ the one sided directional derivative of L at the origin in the direction x:

$$\zeta(x) = \lim_{t\to 0^+} t^{-1}(L(tx) - L(0)) = \inf_{t>0} t^{-1}(L(tx) - L(0)) \in \mathbb{R} \cup \{\pm\infty\} \ , \quad (30)$$

which always exists since L is convex.

Proposition 12.1. *Assume that ζ does not take the value $-\infty$. Then, the eigenvectors of the Lax-Oleinik semigroup $(T^t)_{t\geq 0}$ with eigenvalue $-L(0)$ are precisely the functions $u : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$, not identically $-\infty$, such that*

$$-\zeta(y - x) + u(y) \leq u(x) \ , \ \textit{for all } x, y \in \mathbb{R}^n \ . \quad (31)$$

Moreover, when ζ only takes finite values, the extremal eigenvectors with eigenvalue $-L(0)$ are of the form $c + w$, where $c \in \mathbb{R}$ and w belongs to the minimal Martin space of the kernel $(x, y) \mapsto -\zeta(y - x)$ with respect to any basepoint.

Proof. Let us introduce the kernels

$$A_s := T^s + sL(0), \text{ for all } s \geq 0.$$

Using the Hopf-Lax formula, we get

$$(A_s)^+_{xy} = \sup_{k\in\mathbb{N}\setminus\{0\}} -ksL\Big(\frac{y-x}{ks}\Big) + ksL(0) \ .$$

Using (30) and the fact that $\zeta(0) = 0$, we deduce that

$$(A_s)^*_{xy} = (A_s)^+_{xy} = -\zeta(y - x) \ . \quad (32)$$

The eigenvectors of the semigroup $(T^t)_{t\geq 0}$ are precisely the functions that are harmonic with respect to all the kernels A_s, with $s>0$. Since $(A_s)_{xx}=0$ for all $x\in\mathbb{R}^n$, the harmonic and super-harmonic functions of A_s coincide. It follows from Proposition 2.1 that u is a super-harmonic function of A_s if and only if $u\geq A_s^*u$. Since the latter condition can be written as (31) and is independent of s, the first assertion of the corollary is proved.

By (32), when ζ is finite, any point can be taken as the basepoint. The kernels A_s and $(x,y)\mapsto -\zeta(y-x)$ have the same Martin and minimal Martin spaces with respect to any given basepoint, and so the final assertion of the corollary follows from Theorem 6.2. $\square$

Remark 12.2. When $\partial L(0)$, the subdifferential of L at the origin, is non-empty, ζ does not take the value $-\infty$. This is the case when the origin is in the relative interior of the domain of L. Then, ζ coincides with the support function of $\partial L(0)$:

$$\zeta(x)=\sup_{y\in\partial L(0)} y\cdot x, \qquad \text{for all } x\in\mathbb{R}^n \enspace ,$$

see [Roc70, Th. 23.4]. If in addition the origin is in the interior of the domain of L, then $\partial L(0)$ is non-empty and compact, and so the function ζ is everywhere finite.

Corollary 12.3. *When ζ is a norm on $\mathbb{R}^n$, the extremal eigenvectors with eigenvalue $-L(0)$ of the Lax-Oleinik semigroup $(T^t)_{t\geq 0}$ are precisely the functions $x\mapsto c-\zeta(y-x)$, where $c\in\mathbb{R}$ and $y\in\mathbb{R}^n$, together with the functions $c+w$, where $c\in\mathbb{R}$ and w is a Busemann point of the normed space $(\mathbb{R}^n,\zeta)$.*

Proof. This follows from Proposition 12.1 and Corollary 7.13. $\square$

Remark 12.4. The map ζ is a norm when the origin is in the interior of the domain of L and the subdifferential $\partial L(0)$ is symmetric, meaning that $p\in\partial L(0)$ implies $-p\in\partial L(0)$. When ζ is a norm, condition (31) means that u is Lipschitz-continuous with respect to ζ or that u is identically $-\infty$.

We next study the eigenspace of $(T^t)_{t\geq 0}$ for an eigenvalue $\lambda>-L(0)$ in the special case where L is of the form

$$L(x)=\frac{\|x\|^p}{p} \enspace ,$$

where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^n$ and $p>1$. For all $\lambda>0$, we set

$$\vartheta_\lambda:=(q\lambda)^{\frac{1}{q}} \quad \text{where } \frac{1}{p}+\frac{1}{q}=1 \enspace .$$

Theorem 12.5. *Let $s>0$ and $\lambda>0$. Any eigenvector of T^s with eigenvalue λs is an eigenvector of the Lax-Oleinik semigroup $(T^t)_{t\geq 0}$ with eigenvalue λ. Such an eigenvector can be written as*

$$u=\sup_{w\in\mathscr{M}_{\mathrm{bu}}} \nu(w)+\vartheta_\lambda w \enspace , \tag{33}$$

where $\mathscr{M}_{\mathrm{bu}}$ denotes the set of Busemann points of the normed space $(\mathbb{R}^n, \|\cdot\|)$ and ν is an arbitrary map $\mathscr{M}_{\mathrm{bu}} \to \mathbb{R}\cup\{-\infty\}$ bounded from above. The maximal map ν satisfying (33) is given by μ_u. Moreover, the extremal eigenvectors with eigenvalue λ are of the form $c + \vartheta_\lambda w$, where $c \in \mathbb{R}$ and $w \in \mathscr{M}_{\mathrm{bu}}$.

This theorem follows from Theorem 8.1, Theorem 8.3, and the next lemma.

Lemma 12.6. *For all $s > 0$, the minimal Martin space of the kernel $A_s := T^s - s\lambda$, with respect to any basepoint, coincides with the set of functions $\vartheta_\lambda w$, where w is a Busemann point of the normed space $(\mathbb{R}^n, \|\cdot\|)$ equipped with the same basepoint.*

Proof. For all $x, y \in \mathbb{R}^n$, we set

$$\psi(t) := -t^{1-p} L(y-x) - t\lambda \ .$$

It follows from the Hopf-Lax formula that

$$(A_s)^+_{xy} = \sup_{k\in\mathbb{N}\setminus\{0\}} \psi(ks) \ . \tag{34}$$

Since ψ is concave, the supremum of $\psi(t)$ over all $t > 0$ is attained at the point $\bar{t}$ such that

$$\psi'(\bar{t}) = \bar{t}^{-p}(p-1)L(y-x) - \lambda = 0 \ .$$

It follows that

$$\psi(\bar{t}) = -\vartheta_\lambda \|y - x\| \ .$$

Since ψ is concave, we have $\psi(t) \geq \psi(\bar{t}) + \psi'(t)(t-\bar{t})$, and so, for $t \geq \bar{t}$,

$$\begin{aligned}\psi(t) - \psi(\bar{t}) &= \psi(t) - \psi(\bar{t}) - \psi'(\bar{t})(t-\bar{t}) \\ &\geq (\psi'(t) - \psi'(\bar{t}))(t-\bar{t}) \geq \psi''(\bar{t})(t-\bar{t})^2\end{aligned}$$

since ψ' is convex. Let k denote the smallest integer such that $\bar{t} \leq ks$, and let $t = ks$. We deduce that

$$0 \geq \psi(t) - \psi(\bar{t}) \geq -p(p-1)L(y-x)\bar{t}^{-1-p}(t-\bar{t})^2 = -p\lambda\bar{t}^{-1}(t-\bar{t})^2 \ .$$

Since $\bar{t} \leq t \leq \bar{t} + s$, since $\bar{t} = (q\lambda)^{-1/p}\|y-x\|$, and since

$$\psi(\bar{t}) \geq (A_s)^*_{xy} \geq (A_s)^+_{xy} \geq \psi(t) \ ,$$

we get

$$(A_s)^*_{xy} = -\vartheta_\lambda\|y-x\| + \epsilon(\|y-x\|) \ , \tag{35}$$

where ϵ is a function tending to 0 at infinity. Observe that the supremum in (34) is always attained by an integer k which can be bounded by an increasing function of $\|y-x\|$. Hence, for all $x \in \mathbb{R}^n$ and every compact set C, we can find an integer N such that $(A_s)^+_{xy} = \sup_{1\leq k\leq N} \psi(ks)$ for all $y \in C$. Since every $\psi(ks)$ is a continuous function of $y-x$, we deduce that the map $y \mapsto (A_s)^+_{xy}$ is continuous.

Denote by K the Martin kernel of A_s with respect to this basepoint and denote by $\mathscr{M}$, $\mathscr{M}^m$, and $\mathscr{K}$, the corresponding Martin space, minimal Martin space, and set of columns of the Martin kernel. Also, we denote by H the kernel constructed from K as in Section 4. Define the kernel $A' : (x,y) \mapsto -\vartheta_\lambda\|y - x\|$. We use $K', \mathscr{M}', \mathscr{M}'^m, \mathscr{K}'$ and H' to denote the corresponding objects constructed from A'.

We next show that $\mathscr{M}^m = \mathscr{M}'^m \setminus \mathscr{K}'$.

An element w of $\mathscr{M}^m$ is the limit of a net $(K_{\cdot y_d})_{d\in D}$. If the net $(y_d)_{d\in D}$ had a bounded subnet, it would have a subnet converging to some $y \in \mathbb{R}^d$. Then, by continuity of the map $z \mapsto (A_s)^+_{\cdot z}$, the element w would be proportional in the max-plus sense either to $f := (A_s)^*_{\cdot y}$ or to $g := (A_s)^+_{\cdot y}$ (the first case arises if the subnet is ultimately constant). Both cases can be ruled out: we know from Proposition 4.4 that an element of the minimal Martin space is harmonic, but $f_y = 0 \neq g_y = (A_s f)_y = -s\lambda \neq (A_s g)_y = -2s\lambda$, and so f and g are not harmonic. This shows that $(y_d)_{d\in D}$ tends to infinity.

By (35), we deduce that $K'_{\cdot y_d}$ tends to w. Thus, any net $(y_d)_{d\in D}$ such that $K_{\cdot y_d}$ tends to w is such that y_d tends to infinity and $K'_{\cdot y_d}$ tends to w. We deduce that $w \in \mathscr{M}'$ and $H'(w,w) \geq H(w,w) = \mathbb{1}$, and so, by (12), $\mathscr{M}^m \subset \mathscr{M}'^m \cup \mathscr{K}'$. We proved that the columns of $(A_s)^*$ are not harmonic, and so $\mathscr{M}^m \subset \mathscr{M} \setminus \mathscr{K}$. We claim that $\mathscr{M}^m \subset \mathscr{M}'^m \setminus \mathscr{K}'$. Indeed, if a net $K_{\cdot y_d}$ converges to $w \in \mathscr{M}^m$, we showed that $(y_d)_{d\in D}$ tends to infinity, and that $K'_{\cdot y_d}$ tends to w. But $K'_{\cdot y_d}$ cannot converge to an element $K'_{\cdot y} \in \mathscr{K}'$ because the map sending an element of a finite-dimensional normed space to its column of the Martin kernel is an embedding (see [Bal95, Ch. II,§1] for a more general result). So $w \notin \mathscr{K}'$.

Let us take now $w' \in \mathscr{M}'^m \setminus \mathscr{K}'$. Then, w' is the limit of some net $(K'_{\cdot y'_d})_{d\in D'}$, where $(y'_d)_{d\in D'}$ necessarily tends to infinity, since otherwise, there would be a subnet of $(y'_d)_{d\in D'}$ converging to some $z \in \mathbb{R}^n$, and so we would have $w' = K'_{\cdot z} \in \mathscr{K}'$. It follows from (35) that w' is the limit of $K_{\cdot y'_d}$, and hence $w' \in \mathscr{M}$. These properties also imply that $H'(w',w') \leq H(w',w')$. Since $w' \in \mathscr{M}'^m$, we have $H'(w',w') = \mathbb{1}$, and so $H(w',w') = \mathbb{1}$, and by (12), $w' \in \mathscr{M}^m \cup \mathscr{K}$. Observe that the map $z \mapsto w'_z$ is continuous because it is a pointwise limit of elements of $\mathscr{K}'$, all of which are Lipschitz continuous with constant ϑ_λ with respect to the norm $\|\cdot\|$. For all $y \in \mathbb{R}^n$, the map $x \mapsto A^*_{xy}$ takes the value 0 when $x = y$ and the value $(A_s)^+_{xy} \leq -s\lambda < 0$ when $x \neq y$. Thus, the elements of $\mathscr{K}$ are not continuous, and so, $w' \notin \mathscr{K}$. It follows that $w' \in \mathscr{M}^m \setminus \mathscr{K} = \mathscr{M}^m$. We have shown that $\mathscr{M}^m = \mathscr{M}'^m \setminus \mathscr{K}'$.

By Corollary 7.13, $\mathscr{M}'^m \setminus \mathscr{K}'$ is the set of Busemann points of the normed space $(\mathbb{R}^n, \vartheta_\lambda\|\cdot\|)$. These are precisely the functions of the form $\vartheta_\lambda w$, where w is a Busemann point of $(\mathbb{R}^n, \|\cdot\|)$. $\square$

Remark 12.7. Lemma 12.6 identifies a special situation where the minimal Martin space of $T^s - s\lambda$ is independent of s. This seems related to the fact that the set of functions of the form $x \mapsto a\|x\|^p$ with $a > 0$ is stable by inf-convolution. One may still obtain a representation of the eigenvectors for more general semigroups $(T^t)_{t\geq 0}$, but this requires adapting some of the present results to the continuous-time setting. We shall present this elsewhere.

Example 12.8. Consider the Euclidean norm on $\mathbb{R}^n$, $\|x\| := (x \cdot x)^{1/2}$, and $L(x) := \|x\|^p/p$ with $p > 1$. The set of Busemann points of the normed space $(\mathbb{R}^n, \|\cdot\|)$, with respect to the basepoint 0, coincides with the set of functions

$$w : x \mapsto x \cdot y \ ,$$

where y is an arbitrary vector of norm 1. It follows from Theorem 12.5 that the extremal eigenvectors with eigenvalue $\lambda > 0$ of the Lax-Oleinik semigroup are of the form $c + \vartheta_\lambda w$, with $c \in \mathbb{R}$, and that any eigenvector with eigenvalue λ is a supremum of maps of this form. In particular, when $n = 1$, there are two Busemann points, $w^\pm(x) = \pm\vartheta_\lambda x$, and any eigenvector u with eigenvalue λ can be written as

$$x \mapsto \max(c^+ + \vartheta_\lambda x, c^- - \vartheta_\lambda x) \ ,$$

with $c^\pm \in \mathbb{R} \cup \{-\infty\}$. The Busemann points $w^\pm$ are the limits of the geodesics $t \mapsto \pm t$, from $[0,\infty[$ to $\mathbb{R}$. Hence, Proposition 7.6 allows us to determine the maximal representing measure μ_u, or equivalently, the maximal value of the scalars $c^\pm$, as follows:

$$c^\pm = \lim_{t\to\pm\infty} u(t) \mp \vartheta_\lambda t \ .$$

In this special case, the representing measure is unique.

In order to give another example, we characterise the Busemann points of a polyhedral norm. We call *proper face* of a polytope the intersection of this polytope with a supporting half-space.

Proposition 12.9. *Let $\|\cdot\|$ denote a polyhedral norm on $\mathbb{R}^n$, so that*

$$\|x\| = \max_{i\in I} x'_i \cdot x \ ,$$

where $(x'_i)_{i\in I}$ is the finite family of the extreme points of the dual unit ball. The Martin boundary of the kernel $(x, y) \mapsto -\|x - y\|$, taking the origin as the basepoint, is precisely the set of functions of the form

$$x \mapsto \min_{j\in J} x'_j \cdot (x - X) + \max_{j\in J} x'_j \cdot X \ , \tag{36}$$

where $X \in \mathbb{R}^n$ and $(x'_j)_{j\in J}$ is the set of extreme points of a proper face of the dual unit ball. Moreover, all the points of the Martin boundary are Busemann points.

Proof. Any point f of the Martin boundary is the limit of a sequence of functions

$$x \mapsto f^k(x) = \|X^k\| - \|X^k - x\| \ ,$$

where $X^k \in \mathbb{R}^n$ and $\|X^k\| \to \infty$ when $k \to \infty$. Consider the sequence of vectors

$$u^k = (x'_i \cdot X^k - \|X^k\|)_{i\in I} \ .$$

These vectors lie in $[-\infty, 0]^I$, which is compact and metrisable, and so, we may assume, by taking a subsequence if necessary, that u^k converges to some vector $u \in [-\infty, 0]^I$. Since I is finite, we may also assume, again taking a subsequence if necessary, that there exists an index $j_0 \in I$ such that $x'_{j_0} \cdot X^k = \|X^k\|$ for all k. Let $J := \{i \in I \mid u_i > -\infty\}$. Observe that J is non-empty since $u_{j_0} = 0$. We have

$$\begin{aligned} f(x) &= \lim_{k\to\infty} f^k(x) = \lim_{k\to\infty} -\max_{i\in I}(x'_i \cdot X^k - \|X^k\| - x'_i \cdot x) \\ &= -\max_{j\in J}(u_j - x'_j \cdot x) \ . \end{aligned}$$

Observe that the set $E := \{((x'_j - x'_{j_0}) \cdot X)_{j\in J} \mid X \in \mathbb{R}^n\}$ is closed, since it is a finite-dimensional vector space. Since the vector $(u^k)_{j\in J}$ belongs to E and has a finite limit when $k \to \infty$, this limit belongs to E, and so there exists some $X \in \mathbb{R}^n$ such that $u_j = x'_j \cdot X - x'_{j_0} \cdot X$ for all $j \in J$. Thus,

$$f(x) = -\max_{j\in J} x'_j \cdot (X - x) + x'_{j_0} \cdot X \ .$$

Since $f(0) = 0$, we have $\max_{j\in J} x'_j \cdot X = x'_{j_0} \cdot X$, and so

$$f(x) = -\max_{j\in J} x'_j \cdot (X - x) + \max_{j\in J} x'_j \cdot X \ ,$$

which is of the form (36).

We now have to show that $(x'_j)_{j\in J}$ is the set of extreme points of a face of the dual unit ball. Let E' denote the set of vectors $x' \in \mathbb{R}^n$ such that $x' \cdot X^k - \|X^k\|$ remains bounded when k tends to infinity. This is an affine space. Let B' denote the dual unit ball. We claim that $F' := E' \cap B'$ is an extreme subset of B', meaning that

$$\alpha x' + (1-\alpha)y' \in F' \implies x', y' \in F', \qquad \text{for all } x', y' \in B' \text{ and } 0 < \alpha < 1. \tag{37}$$

Indeed, let $x', y' \in B'$ and $0 < \alpha < 1$. Since $x' \in B'$, we have $x' \cdot X \leq \|X\|$ for all $X \in \mathbb{R}^n$. In particular, $x' \cdot X^k - \|X^k\| \leq 0$ for all k. Similarly, $y' \cdot X^k - \|X^k\| \leq 0$ for all k. Since

$$\begin{aligned} &(\alpha x' + (1-\alpha)y') \cdot X^k - \|X^k\| \\ &= \alpha(x' \cdot X^k - \|X^k\|) + (1-\alpha)(y' \cdot X^k - \|X^k\|) \\ &\leq \alpha(x' \cdot X^k - \|X^k\|) \\ &\leq 0 \ , \end{aligned}$$

we deduce that $x' \cdot X^k - \|X^k\|$ is bounded if $\alpha x' + (1-\alpha)y' \in F'$. Similarly, $y' \cdot X^k - \|X^k\|$ is bounded. This shows (37).
Let z denote any accumulation point of the sequence $\|X^k\|^{-1}X^k$. We have $F' \subset \{x' \in B' \mid x' \cdot z = 1\}$, and so, $F' \neq B'$.
Since the dual ball B' is a polytope, the convex extreme subset $F' \neq B'$ is a proper face of B'. Therefore, the vectors x'_i, with $i \in I$, such that $x'_i \cdot X^k - \|X^k\|$ remains bounded are precisely the x'_i that belong to the proper face F'. Hence, these x'_i are the extreme points of the proper face F'.
Every proper face F' of the dual ball is the intersection of the dual ball with a supporting hyperplane, so $F' = \{x' \in B' \mid x' \cdot y = 1\}$ for some $y \in B$. Observe that the set J of x'_i such that $x'_i \cdot y = 1$ is precisely the set of extreme points of F'. Consider now $X \in \mathbb{R}^n$ and the ray $t \mapsto X + ty$, which is a geodesic, and a fortiori an almost-geodesic. One readily checks that the function $x \mapsto \|X + ty\| - \|X + ty - x\|$ converges to the function (36) when t tends to $+\infty$, and so, every point of the Martin boundary is a Busemann point. $\square$

Remark 12.10. Karlsson, Metz, and Noskov [KMN06] have shown previously that every boundary point of a polyhedral normed space is the limit of a geodesic, and hence a Busemann point. They did this by characterising the sequences which converge to a boundary point.

Example 12.11. Consider now $L(x) := \|x\|_\infty^p/p$ with $\|x\|_\infty := \max(|x_1|, \cdots, |x_n|)$ and $p > 1$. By Proposition 12.9, the Busemann points of $(\mathbb{R}^n, \|\cdot\|_\infty)$ with respect to the basepoint 0 are of the form:

$$w:\ x \mapsto \min_{i\in I} \epsilon_i(x_i - X_i) + \max_{i\in I} \epsilon_i X_i \ \ ,$$

where I is a non-empty subset of $\{1, \ldots, n\}$, $\epsilon_i = \pm 1$, and the X_i are arbitrary reals. Theorem 12.5 shows that any eigenvector with eigenvalue $\lambda > 0$ of the Lax-Oleinik semigroup can be written as a supremum of maps $c + \vartheta_\lambda w$, where $c \in \mathbb{R} \cup \{-\infty\}$ and w is of the above form. For instance, when $n = 2$, the functions w are of one of the following forms:

$$\epsilon_1 x_1, \quad \epsilon_2 x_2, \text{ or } \min(\epsilon_1(x_1 - X_1), \epsilon_2(x_2 - X_2)) + \max(\epsilon_1 X_1, \epsilon_2 X_2) \ \ ,$$

with $X_1, X_2 \in \mathbb{R}$ and $\epsilon_1 = \pm 1, \epsilon_2 = \pm 1$.

Remark 12.12. It is natural to ask whether the eigenvectors of the Lax-Oleinik semigroup $(T^t)_{t\geq 0}$ coincide with the viscosity solutions of the ergodic Hamilton-Jacobi equation

$$L^\star(\nabla u) = \lambda \ \ ,$$

where $L^\star$ denotes the Legendre-Fenchel transform of L. This is proved in [Fat08, Chapter 7] in the different setting where the space is a compact manifold and the Lagrangian L can depend on both the position and the speed but must satisfy certain regularity and coercivity conditions.

References

[AG03] M. Akian and S. Gaubert. Spectral theorem for convex monotone homogeneous maps, and ergodic control. *Nonlinear Analysis. Theory, Methods & Applications*, 52(2):637–679, 2003.

[AGW05] M. Akian, S. Gaubert, and C. Walsh. Discrete max-plus spectral theory. In G. L. Litvinov and V. P. Maslov, editors, *Idempotent Mathematics and Mathematical Physics*, Contemporary Mathematics, pages 19–51. American Mathematical Society, 2005. arXiv:math.SP/0405225.

[Aki95] Marianne Akian. Theory of cost measures: convergence of decision variables. Rapport de recherche 2611, INRIA, 1995.

[Aki99] Marianne Akian. Densities of idempotent measures and large deviations. *Trans. Amer. Math. Soc.*, 351(11):4515–4543, 1999.

[And04] P.D. Andreev. Ideal closures of Busemann space and singular Minkowski space. arXiv:math.GT/0405121, 2004.

[And07] P. D. Andreev. Geometry of ideal boundaries of geodesic spaces with nonpositive curvature in the sense of Busemann. *Mat. Tr.*, 10(1):16–28, 2007.

[AQV98] M. Akian, Jean-Pierre Quadrat, and M. Viot. Duality between probability and optimization. In J. Gunawardena, editor, *Idempotency*, pages 331–353. Cambridge University Press, 1998.

[Bal95] Weerner Ballmann. *Lecture on spaces of nonpositive curvature.* Birkhäuser, 1995.

[Bap98] R. B. Bapat. A max version of the Perron-Frobenius theorem. *Linear Algebra Appl.*, 275/276:3–18, 1998.

[BCOQ92] F. Baccelli, G. Cohen, G. J. Olsder, and Jean-Pierre Quadrat. *Synchronization and linearity : an algebra for discrete events systems.* John Wiley & Sons, New-York, 1992.

[BGS85] Werner Ballmann, Mikhael Gromov, and Viktor Schroeder. *Manifolds of nonpositive curvature*, volume 61 of *Progress in Mathematics.* Birkhäuser Boston Inc., Boston, MA, 1985.

[CG79] Raymond Cuninghame-Green. *Minimax algebra.* Springer-Verlag, Berlin, 1979.

[CGQ04] G. Cohen, S. Gaubert, and J.-P. Quadrat. Duality and separation theorems in idempotent semimodules. *Linear Algebra and Appl.*, 379:395–422, 2004. arXiv:math.FA/0212294.

[Con01] G. Contreras. Action potential and weak KAM solutions. *Calc. Var. Partial Differential Equations*, 13(4):427–458, 2001.

[Doo59] J. L. Doob. Discrete potential theory and boundaries. *J. Math. Mech.*, 8:433–458; erratum 993, 1959.

[Dud92] P.I. Dudnikov. Endomorphisms of the semimodules of bounded functions. In *Idempotent analysis*, volume 13 of *Advances In Soviet Mathematics*, pages 47–63. Amer. Math. Soc., Providence, 1992.

[Dyn69] E.B. Dynkin. Boundary theory of Markov processes (the discrete case). *Russian Math. Surveys*, 24(7):1–42, 1969.

[Eva98] Laurence C. Evans. *Partial differential equations.* AMS, 1998.

[Fat97a] Albert Fathi. Solutions KAM faibles conjuguées et barrières de Peierls. *C. R. Acad. Sci. Paris Sér. I Math.*, 325(6):649–652, 1997.

[Fat97b] Albert Fathi. Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens. *C. R. Acad. Sci. Paris Sér. I Math.*, 324(9):1043–1046, 1997.

[Fat03] Albert Fathi. Regularity of C^1 solutions of the Hamilton-Jacobi equation. *Ann. Fac. Sci. Toulouse Math. (6)*, 12(4):479–516, 2003.

[Fat08] A. Fathi. *Weak KAM theorem in Lagrangian Dynamics.* Cambridge University Press, 2008. To appear.

[FM02] A. Fathi and E. Maderna. Weak KAM theorem on non compact manifolds. Preprint. To appear in Nonlinear Differ. Equ. Appl., 2002.

[FS04] Albert Fathi and Antonio Siconolfi. Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation. *Invent. Math.*, 155(2):363–388, 2004.

[GM77] M. Gondran and M. Minoux. Valeurs propres et vecteurs propres dans les dioïdes et leur interprétation en théorie des graphes. *Bull. Direction Études Recherches Sér. C Math. Informat.*, 2:i, 25–41, 1977.

[GM02] M. Gondran and M. Minoux. *Graphes, Dioïdes et semi-anneaux.* TEC & DOC, Paris, 2002.

[Gro81] M. Gromov. Hyperbolic manifolds, groups and actions. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978)*, volume 97 of *Ann. of Math. Stud.*, pages 183–213, Princeton, N.J., 1981. Princeton Univ. Press.

[Hun60] G. A. Hunt. Markoff chains and Martin boundaries. *Illinois J. Math.*, 4:313–340, 1960.

[IM07] H. Ishii and H. Mitake. Representation formulas for solutions of Hamilton-Jacobi equations with convex hamiltonians. *Indiana Univ. Math. J.*, 56(5), 2007.

[KM97] V. N. Kolokoltsov and V. P. Maslov. *Idempotent analysis and applications.* Kluwer Acad. Publisher, 1997.

[KMN06] A. Karlsson, V. Metz, and G. A. Noskov. Horoballs in simplices and Minkowski spaces. *Int. J. Math. Math. Sci.*, pages Art. ID 23656, 20, 2006.

[LMS01] G. L. Litvinov, V. P. Maslov, and G. B. Shpiz. Idempotent functional analysis: An algebraical approach. *Mat. Notes*, 69(5):696–729, 2001.

[Mar41] Robert S. Martin. Minimal positive harmonic functions. *Trans. Amer. Math. Soc.*, 49:137–172, 1941.

[Mas87] V. Maslov. *Méthodes opératorielles.* Éditions Mir, Moscow, 1987. Translated from the Russian.

[MS92] V. P. Maslov and S. N. Samborskiĭ. *Idempotent analysis*, volume 13 of *Advances In Soviet Mathematics.* Amer. Math. Soc., Providence, 1992.

[MSS00] R. S. MacKay, S. Slijepčević, and J. Stark. Optimal scheduling in a periodic environment. *Nonlinearity*, 13(1):257–297, 2000.

[NS66] P. Ney and F. Spitzer. The Martin boundary for random walk. *Trans. A.M.S.*, 121:116–132, 1966.

[OV91] George L. O'Brien and Wim Vervaat. Capacities, large deviations and loglog laws. In *Stable processes and related topics (Ithaca, NY, 1990)*, volume 25 of *Progr. Probab.*, pages 43–83. Birkhäuser Boston, Boston, MA, 1991.

[Puh01] Anatoly Puhalskiĭ. *Large Deviations and Idempotent Probability.* Number 119 in Monographs and Surveys in Pure and Applied Mathematics. Chapman & Hall, 2001.

[Rie02] Marc A. Rieffel. Group C^*-algebras as compact quantum metric spaces. *Doc. Math.*, 7:605–651 (electronic), 2002. arXiv:math.OA/0205195.

[Roc70] R. T. Rockafellar. *Convex analysis.* Princeton University Press Princeton, N.J., 1970.

[Rom67] I. V. Romanovskiĭ. Optimization of stationary control of discrete deterministic process in dynamic programming. *Kibernetika*, 3(2):66–78, 1967.

[Wal07] Cormac Walsh. The horofunction boundary of finite-dimensional normed spaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, 142(03):497–507, 2007.

[Wal08] C. Walsh. The horofunction boundary of the Hilbert geometry. *Advances in Geometry*, 8(4):503–529, 2008.

[Wal09] Cormac Walsh. Minimum representing measures in idempotent analysis. In G.L. Litvinov and S.N. Sergeev, editors, *Proceedings of the International Conference on Tropical and Idempotent Mathematics*, Contemporary Mathematics. American Mathematical Society, 2009. To appear. Also arXiv:math.MG/0503716.

[WW05] Corran Webster and Adam Winchester. Boundaries of hyperbolic metric spaces. *Pacific J. Math.*, 221(1):147–158, 2005.

[WW06] Corran Webster and Adam Winchester. Busemann points of infinite graphs. *Trans. Amer. Math. Soc.*, 358(9):4209–4224, 2006.

[YK92] S. Yu. Yakovenko and L. A. Kontorer. Nonlinear semigroups and infinite horizon optimization. In *Idempotent analysis*, volume 13 of *Adv. Soviet Math.*, pages 167–210. Amer. Math. Soc., Providence, RI, 1992.

Marianne Akian
INRIA Saclay – Île-de-France and CMAP
CMAP, École Polytechnique
91128 Palaiseau Cedex
France
marianne.akian@inria.fr

Stéphane Gaubert
INRIA Saclay – Île-de-France and CMAP
CMAP, École Polytechnique
91128 Palaiseau Cedex
France
stephane.gaubert@inria.fr

Cormac Walsh
INRIA Saclay – Île-de-France and CMAP
CMAP, École Polytechnique
91128 Palaiseau Cedex
France
cormac.walsh@inria.fr

A note on the p-adic Galois representations attached to Hilbert modular forms

Christopher Skinner

Received: October 21, 2008

Communicated by Don Blasius

Abstract. We show that the p-adic Galois representations attached to Hilbert modular forms of motivic weight are potentially semistable at all places above p and are compatible with the local Langlands correspondence at these places, proving this for those forms not covered by the previous works of T. Saito and of D. Blasius and J. Rogawski.

2000 Mathematics Subject Classification: 11F80, 11F41

1 Introduction

Let F be a totally real extension of $\mathbf{Q}$ of degree d. Let $\overline{F}$ be an algebraic closure of F and let $G_F := \mathrm{Gal}(\overline{F}/F)$. Let $I := \mathrm{Hom}_{\mathbf{Q}}(F, \mathbf{C})$ be the set of embeddings of F into $\mathbf{C}$. The set I indexes the archimedean places of F. For each finite place v of F let $\overline{F}_v$ be an algebraic closure of F_v and fix an F-embedding $\overline{F} \hookrightarrow \overline{F}_v$. These determine a choice of a decomposition group $D_v \subset G_F$ for each v and an identification of D_v with $\mathrm{Gal}(\overline{F}_v/F_v)$. Let p be a rational prime and fix an algebraic closure $\overline{\mathbf{Q}}_p$ of $\mathbf{Q}_p$ and an isomorphism $\iota : \mathbf{C} \xrightarrow{\sim} \overline{\mathbf{Q}}_p$. Via composition with ι the set I is identified with the embeddings of F into $\overline{\mathbf{Q}}_p$.

Let π be a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_F)$. Then π is a restricted tensor product $\pi = \otimes' \pi_v$ with v running over all places of F. Assume that each π_i, $i \in I$, is a discrete series representation with Blattner parameter $k_i \geq 2$ and central character $x \mapsto \mathrm{sgn}(x)^{k_i} |x|_i^{-w}$ with w an integer independent of i. We say that π has infinity type $(\mathbf{k}, w)$, $\mathbf{k} := (k_i)_{i \in I}$. Assume also that each $k_i \equiv w \pmod 2$. In this case, π is an automorphic representation associated with a Hilbert modular eigenform of weight $\mathbf{k}$. We recall that attached to π (and ι) is a two-dimensional semisimple Galois representation

$$\rho_\pi : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$$

such that

$$\mathrm{WD}(\rho_\pi|_{D_v})^{\text{Fr-ss}} \cong \iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2}) \quad \forall v \nmid p\infty. \tag{1}$$

Here $\mathrm{WD}(\sigma)$ denotes the Weil-Deligne representation over $\overline{\mathbf{Q}}_p$ associated to a continuous representation $\sigma : D_v \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$ for a place $v \nmid p\infty$ (see [Ta, (4.2.1)]), and the superscript 'Fr-ss' denotes its Frobenius semi-simplification. Also, $\mathrm{Rec}_v(\tau)$ denotes the Frobenius semi-simple Weil-Deligne representation over $\mathbf{C}$ associated with an irreducible admissible representation τ of $\mathrm{GL}_n(F_v)$ by the local Langlands correspondence, and $\iota\mathrm{Rec}_v(\tau)$ is the Weil-Deligne representation over $\overline{\mathbf{Q}}_p$ obtained from $\mathrm{Rec}_v(\tau)$ by change of scalars via the isomorphism ι. We choose Rec_v so that when $n = 1$, Rec_v is the inverse of the Artin map of local class field theory normalized so that uniformizers correspond to geometric frobenius elements. The existence of a ρ_π satisfying (1) was established by Carayol [Ca2], Wiles [W], Blasius and Rogawski [BR], and Taylor [Tay1], following the work of Eichler, Shimura, Deligne, Langlands, and others on the Galois representations associated with elliptic modular eigenforms.

The purpose of this note is to complete the proof of the analog of (1) at places $v \mid p$:

THEOREM 1 *Let $v \mid p$ be a place of F. The representation $\rho_\pi|_{D_v}$ is potentially semistable with Hodge-Tate type $(\mathbf{k}, w)$ and satisfies*

$$\mathrm{WD}(\rho_\pi|_{D_v})^{\text{Fr-ss}} \cong \iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2}). \tag{2}$$

We recall that $\rho_v := \rho_\pi|_{D_v}$ is potentially semistable if

$$D_{pst}(\rho_v) := \bigcup_{L/F_v} (\rho_v \otimes_{\mathbf{Q}_p} B_{st})^{\mathrm{Gal}(\overline{F}_v/L)}$$

is a free $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}$-module of rank 2, where here L is ranging over all finite extensions of F_v, $F_{v,0}^{ur}$ is the union of all absolutely unramified subfields of $\overline{F}_v$, and B_{st} is Fontaine's ring of semistable p-adic periods (the latter has a continuous action of $D_v = \mathrm{Gal}(\overline{F}_v/F_v)$ with the property that $B_{st}^{\mathrm{Gal}(\overline{F}_v/L)} = L_0$, the maximal absolutely unramified subfield of L). We also recall that the module $D_{HT}(\rho_v) := (V \otimes_{\mathbf{Q}_p} B_{HT})^{D_v}$ is a graded $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_v$-module (recall that $B_{HT} := \oplus_{n\in\mathbf{Z}}\mathbf{C}_{F_v}(n)$, $\mathbf{C}_{F_v} := \widehat{\overline{F}}_v$, with the obvious action of D_v). By $\rho_\pi|_{D_v}$ having Hodge-Tate type $(\mathbf{k}, w)$, we mean that for $j \in \mathrm{Hom}_{\mathbf{Q}_p}(F_v, \overline{\mathbf{Q}}_p)$ the induced graded module $D_{HT}(\rho_v) \otimes_{\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F, j} \overline{\mathbf{Q}}_p$ is non-zero in degrees $(w - k_{i(j)})/2$ and $(w + k_{i(j)} - 2)/2$, where $i(j) \in I$ is the induced embedding of F into $\overline{\mathbf{Q}}_p$. To make sense of the left-hand side of (2) we recall that Fontaine has defined an action of the Weil-Deligne group on $D_{pst}(\rho_v)$. Given an embedding $\tau : F_{v,0}^{ur} \hookrightarrow \overline{\mathbf{Q}}_p$ we obtain a Weil-Deligne representation over $\overline{\mathbf{Q}}_p$ on $\mathrm{WD}(\rho_v)_\tau := D_{pst}(\rho_v) \otimes_{\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}, \tau} \overline{\mathbf{Q}}_p$. This representation is independent of τ up to equivalence, and we have denoted an element of its equivalence class by $\mathrm{WD}(\rho_v)$. The right-hand side of (2) has the same meaning as in (1).

Saito proved that Theorem 1 holds when either d is odd or there exists a finite place w such that π_w is square-integrable [Sa1, Sa2]; this builds on the aforementioned work of Carayol. Under the same hypotheses or when d is even and some k_i is strictly larger than 2, Blasius and Rogawski proved that $\rho|_{D_v}$ is potentially semistable of Hodge-Tate type $(\mathbf{k}, w)$, and when additionally $\pi_p = \otimes_{v|p}\pi_v$ is unramified they essentially showed that the full conclusion of the theorem holds [BR] (some additional, albeit minor, observations are required to extend their arguments to all such cases). The theorem is of course also known for those π that are the automorphic induction of a (necessarily) algebraic Hecke character of an imaginary quadratic extension of F (such representations are often called CM representations). In this case, Theorem 1 follows from the results in [Se]. These results account for the cases where ρ_π is known to arise from a motive; the conclusion of the theorem then follows from various deep comparison theorems between suitable cohomology theories.

It remains to deal with the cases where ρ_π is not known to arise from a motive, namely those cases where each $k_i = 2$, each π_v is a principal series representation, and π is not a CM representation. In [Tay2] it is shown that if ρ_π is residually irreducible and π_v, $v|p$, is unramified, then $\rho_\pi|_{D_v}$ is crystalline with the predicted Hodge-Tate weights. For $p > 2$ unramified in F, the same result is proved in [Br] without the hypothesis that ρ_π be residually irreducible. For those ρ_π that are residually irreducible, Kisin [Ki1] deduced Theorem 1 from his results on potentially semistable deformation rings, Taylor's construction of the representations ρ_π, and Saito's results. In this paper, we prove Theorem 1 by a different approach. A simple base change argument reduces the theorem, in the cases not covered by Saito's results, to that where d is even and each π_v, $v|p$, is unramified. From the automorphy of the symmetric square $\mathrm{Sym}^2\pi$ and the results of [Mo] it follows that $\mathrm{Sym}^2\rho_v$ is crystalline[1] and even that $\mathrm{WD}(\mathrm{Sym}^2\rho_v) \cong \iota\mathrm{Rec}_v(\mathrm{Sym}^2\pi_v \otimes |\cdot|_v^{-1})$. From results of Wintenberger [Win1, Win2] we then deduce that ρ_v is crystalline up to a (possibly trivial) quadratic twist and hence that $\mathrm{WD}(\rho_v)$ is isomorphic to a (possibly trivial) quadratic twist of $\iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})$. There exists a suitable p-adic analytic family of eigensystems of cuspidal representations of $\mathrm{GL}_2(\mathbf{A}_F)$ (essentially due to Buzzard [Bu1] in the cases needed) that contains an eigensystem attached to ρ_π. For members of this family with sufficiently regular weights Theorem 1 is known by the work of Blasius and Rogawski. An appeal to a result of Kisin then shows that $\mathrm{WD}(\rho_v)$ has at least one D_v-eigenspace predicted by (2), from which we then conclude that (2) holds.

After completing the first draft of this paper, the author learned that Tong Liu [L] has also proven Theorem 1, at least for $p > 2$, by an argument that is a generalization of that of Kisin [Ki1].

Acknowledgements. The author's work on this paper was inspired by a question

[1] As remarked at the end of 2.4.1, a similar use of the symmetric square yields a proof of the Ramanujan conjecture for π. This conjecture has previously been established in [B2].

about what was known regarding Theorem 1 asked by Henri Darmon at the summer school on the stable trace formula, automorphic forms, and Galois representations held at BIRS in August of 2008. The referee prodded the author to write a note with more details. The author's research is supported by grants DMS-0701231 and DMS-0803223 from the National Science Foundation and by a fellowship from the David and Lucile Packard Foundation.

2 The proof of Theorem 1

We keep to the notation from the introduction. We assume some familiarity on the part of the reader with p-adic Hodge theory, particularly the theory of Hodge-Tate weights and the notions of crystalline and semistable representations. A good reference is [Fo]. While p-adic Hodge theory is usually applied to continuous representations of $\mathrm{Gal}(\overline{F}_v/F_v)$, $v|p$, defined over a finite extension of $\mathbf{Q}_p$, we apply it to continuous representations over $\overline{\mathbf{Q}}_p$. This should cause no confusion as the latter are always defined over a finite extension of $\mathbf{Q}_p$. While this is well-known, references seem rare, so we provide a quick proof.

Let Γ be a compact group and $\rho : \Gamma \to \mathrm{GL}_n(\overline{\mathbf{Q}}_p)$ a continuous representation. The subfields L of $\overline{\mathbf{Q}}_p$ that are finite over $\mathbf{Q}_p$ form a countable set, and as each $\mathrm{GL}_n(L)$ is closed in $\mathrm{GL}_n(\overline{\mathbf{Q}}_p)$, the subgroups $\Gamma_L := \rho^{-1}(\mathrm{GL}_n(L))$ form a countable set of closed subgroups of Γ whose union is Γ. Since Γ is compact, it carries a Haar measure with total measure finite and non-zero. As the countable union of measurable sets each having measure zero also has measure zero, it follows that some Γ_L must have non-zero measure and hence have finite index in Γ. Write $\Gamma = \sqcup_{i=1}^m g_i\Gamma_L$. Then ρ takes values in $\mathrm{GL}_n(L')$ where L' is the finite extension of $\mathbf{Q}_p$ generated by L and the entries of the $\rho(g_i)$.

2.1 Weil-Deligne representations over $\overline{\mathbf{Q}}_p$ for $v|p$

Let $v|p$ be a place of F. Let $B_{HT} := \oplus_{n\in\hat{\mathbf{Z}}}\mathbf{C}_{F_v}(n)$ with the obvious action of D_v. Let $B_{cris} \subset B_{st}$ be Fontaine's rings of crystalline and semistable p-adic periods, respectively. Recall that the latter are naturally $F_{v,0}^{ur}$-algebras equipped with a continuous action of D_v such that $B_?^{\mathrm{Gal}(\overline{F}_v/L)} = L_0$ for any finite extension L/F_v, $? = cris, st$, and that furthermore they are both equipped with a compatible $F_{v,0}^{ur}$-semilinear Frobenius morphism $\varphi : B_? \to B_?$ (that is, $\varphi(ax) = \mathrm{frob}_p(a)\varphi(x)$ for all $a \in F_{v,0}^{ur}$, where $\mathrm{frob}_p \in \mathrm{Gal}(F_{v,0}^{ur}/\mathbf{Q}_p)$ is the absolute arithmetic Frobenius). Additionally, B_{st} is equipped with an $F_{v,0}^{ur}$-linear and D_v-equivariant monodromy operator $N : B_{st} \to B_{st}$ such that $B_{cris} = B_{st}^{N=0}$.

For a finite-dimensional $\overline{\mathbf{Q}}_p$-vector space V with a continuous $\overline{\mathbf{Q}}_p$-linear action of D_v we put

$$D_{HT}(V) := (V \otimes_{\mathbf{Q}_p} B_{HT})^{D_v}, \quad D_{cris}(V) := (V \otimes_{\mathbf{Q}_p} B_{cris})^{D_v},$$

and

$$D_{st}^L(V) := (V \otimes_{\mathbf{Q}_p} B_{st})^{\mathrm{Gal}(\overline{F}_v/L)}, \quad D_{pst}(V) := \bigcup_{L/F_v} D_{st}^L(V),$$

where L/F_v is a finite extension. Then $D_{HT}(V)$ is a finite, graded $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_v$-module. Also, $D_{cris}(V)$ is a finite $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}$-module, $D_{st}^L(V)$ is a finite $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} L_0$-module, and $D_{pst}(V)$ is a finite $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}$-module, each of rank at most $\dim_{\overline{\mathbf{Q}}_p}(V)$. The action of φ induces a $\overline{\mathbf{Q}}_p$-linear, $F_{v,0}$-semilinear (resp. L_0-semilinear) Frobenius operator on $D_{cris}(V)$ (resp. $D_{st}^L(V)$) that we also denote by φ. The action of the monodromy operator N on B_{st} induces a $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} L_0$-linear nilpotent operator on $D_{st}^L(V)$ that we also denote by N and which satisfies $N \circ \varphi = p\varphi \circ N$. These are compatible with varying L, so φ and N are defined on $D_{pst}(V)$ as well. Note that $D_{cris}(V) = D_{st}^{F_v}(V)^{N=0}$.

Let $W_v \subset D_v$ be the Weil group of F_v. The action of D_v on V and B_{st} induces a $\overline{\mathbf{Q}}_p$-linear, $F_{v,0}^{ur}$-semilinear action r_{sl} of W_v on $D_{pst}(V)$. We define another action r of W_v on $D_{pst}(V)$: for $w \in W_K$ we let $r(w) = r_{sl}(w) \circ \varphi^{\nu(w)}$ with $\nu(w) \in \mathbf{Z}$ such that w acts on $F_{v,0}^{ur}$ as $\mathrm{frob}_p^{-\nu(w)}$. This also defines an action on $D_{cris}(V)$. The action r is $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}$-linear, and we have

$$N \circ r(w) = N \circ r_{sl}(w) \circ \varphi^{\nu(w)} \circ N = r_{sl}(w) \circ N \circ \varphi^{\nu(w)} = p^{\nu(w)} r(w) \circ N.$$

It follows that the pair (r, N) defines an action of the Weil-Deligne group W_v' of F_v on $D_{pst}(V)$. Moreover, if $\tau : F_{v,0}^{ur} \hookrightarrow \overline{\mathbf{Q}}_p$ is any embedding, then it also follows that the induced action on

$$\mathrm{WD}(V)_\tau := D_{pst}(V) \otimes_{\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}, \tau} \overline{\mathbf{Q}}_p$$

is a Weil-Deligne representation over $\overline{\mathbf{Q}}_p$ (the subscript τ on the tensor sign means that we consider $\overline{\mathbf{Q}}_p$ as a $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}$-algebra via the homomorphism $id \otimes \tau$). Furthermore, $d \otimes x \mapsto \varphi(d) \otimes x$ defines an isomorphism $\mathrm{WD}(V)_{\tau \circ \mathrm{frob}_p} \xrightarrow{\sim} \mathrm{WD}(V)_\tau$ of Weil-Deligne representations over $\overline{\mathbf{Q}}_p$, hence the equivalence class of $\mathrm{WD}(V)_\tau$ is independent of the choice of τ. We let $\mathrm{WD}(V)$ be any member of this equivalence class.

We recall that V is potentially semistable if $D_{pst}(V)$ is a free $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}^{ur}$-module of rank equal to $\dim_{\overline{\mathbf{Q}}_p} V$ or, equivalently, $\dim_{\overline{\mathbf{Q}}_p} \mathrm{WD}(V) = \dim_{\overline{\mathbf{Q}}_p} V$. Similarly, V is crystalline if $D_{cris}(V)$ is a free $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}$-module of rank $\dim_{\overline{\mathbf{Q}}_p} V$. This is equivalent to $(V \otimes_{\mathbf{Q}_p} B_{cris})^{I_v}$ being a free $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{0,v}^{ur}$-module of rank equal to $\dim_{\overline{\mathbf{Q}}_p} V$, where $I_v \subset D_v$ is the inertia subgroup. Thus, V is crystalline if and only if V is potentially semistable and both N and I_v act trivially on $D_{pst}(V)$. In particular, V is crystalline if and only if $\dim_{\overline{\mathbf{Q}}_p} \mathrm{WD}(V) = \dim_{\overline{\mathbf{Q}}_p}(V)$, $\mathrm{WD}(V)$ is unramified (i.e., $N = 0$ and the inertia group I_v acts trivially). Consequently, for V crystalline the eigenvalues of $w \in W_v$ on $\mathrm{WD}(V)^{\text{Fr-ss}}$ are just the roots of the characteristic polynomial of the $\overline{\mathbf{Q}}_p$-endomorphism induced by $\varphi^{\nu(w)}$. We also recall that for a crystalline representation V there is $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_v$-filtration on $D_{cris}(V) \otimes_{F_{v,0}} F_v$ whose associated graded module is just $D_{HT}(V)$.

Suppose now that π_v is unramified. From the preceding paragraph it follows that (2) holds if $\rho_v = \rho_\pi|_{D_v}$ is crystalline and if for all $w \in W_v$

$$\det(1 - T\varphi^{\nu(w)}|D_{cris}(V) \otimes_{\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0},\tau} \overline{\mathbf{Q}}_p) = \det(1 - Tw|\iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})) \tag{3}$$

for some (equivalently, each) embedding $\tau : F_{v,0} \hookrightarrow \overline{\mathbf{Q}}_p$.

2.2 Reduction to d even and π_v unramified

As mentioned in the introduction, Saito has proven Theorem 1 when the degree d of F is odd or some π_v is square-integrable [Sa1],[Sa2]. We may therefore assume that d is even and that π_v is a principal series representation for finite places v. Theorem 1 then asserts that each ρ_v is potentially crystalline with predicted Hodge-Tate weights. Clearly, this is true for $\rho_v = \rho_\pi|_{D_v}$ if and only if there is a finite extension F'/F such that it is true for $\rho_\pi|_{D_{v'}}$, $v'|v$ the place of F' determined by the fixed embedding $\overline{F} \hookrightarrow \overline{F}_v$. Additionally, if ρ_v is potentially crystalline with the predicted Hodge-Tate weights, then to establish (2) it is enough to show that

$$\mathrm{trace}(w|\mathrm{WD}(\rho_v)) = \mathrm{trace}(w|\iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})) \tag{4}$$

for all $w \in W_v$ with $\nu(w) > 0$.

Let $v|p$. For a given $w \in W_v$ such that $\nu(w) > 0$ there exists an abelian extension F'/F such that (a) the base change π' of π to $\mathrm{GL}_2(\mathbf{A}_{F'})$ is cuspidal and unramified at each place over p and (b) $w \in W_{v'} \subseteq W_v$ for $v'|v$ the place of F' determined by the fixed embedding $\overline{F} \hookrightarrow \overline{F}_v$. That (a) can be satisfied is a consequence of each local constituent of π being a principal series representation (we are, of course, using that base change is known for GL_2 for abelian extensions). That (b) can be simultaneously satisfied with (a) is a simple consequence of $\nu(w) > 0$. Note that the extension F'/F may depend on w. As $\rho_{\pi'} \cong \rho_\pi|_{G_{F'}}$, it follows that $\mathrm{WD}(\rho_{\pi'}|_{D_{v'}}) \cong \mathrm{WD}(\rho_\pi|_{D_v})|_{W'_{v'}}$. Similarly, $\mathrm{Rec}_{v'}(\pi_{v'} \otimes |\cdot|_{v'}^{-1/2}) \cong \mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})|_{W'_{v'}}$. Therefore if Theorem 1 holds for π', then ρ_v is potentially crystalline with the predicted Hodge-Tate weights and (4) holds for the given w. This shows that if Theorem 1 holds whenever the representation is unramifed at all primes above p then it also holds for π. Consequently, it suffices to prove Theorem 1 under the assumption that each π_v, $v|p$, is unramified.

2.3 Galois representations in the cohomology of certain Shimura varieties

As mentioned in the introduction, Blasius and Rogawski have essentially proved Theorem 1 in the case where some $k_i > 2$ and each π_v, $v|p$, is unramified [BR].

We explain this here, giving the necessary modifications required to make their argument cover all such cases. We also record some additional consequences for Galois representations associated with essentially self-dual representations of $\mathrm{GL}_3(\mathbf{A}_F)$.

2.3.1 The Shimura varieties

Let $E_0 \subseteq \overline{F}$ be an imaginary quadratic extension of $\mathbf{Q}$ in which p splits and set $E = FE_0$. Fix a place v_0 of E_0 above p. For convenience we assume that for each place $v|p$ of F the fixed embedding $\overline{F} \hookrightarrow \overline{F}_v$ induces the valuation v_0 on E_0. Fix an embedding $E_0 \hookrightarrow \mathbf{C}$ such that - again for convenience - composition with ι also induces the valuation v_0. Let ϕ be the CM type of E consisting of those embeddings $E \hookrightarrow \mathbf{C}$ extending the fixed embedding of E_0. For $\tau \in \phi$ we write $\bar{\tau}$ for the composition of τ with complex conjugation. Restriction to F determines a bijection between ϕ and I, and we write τ_i for the element of ϕ extending $i \in I$. Via composition with ι, ϕ determines a place of E above each place $v|p$ of F; the fixed decomposition group D_v is also a decomposition group for the place of E above p so determined, hence we also denote this place by v, writing $\bar{v}$ for its conjugate (note that each place $v|p$ of F splits in E). If M is an $\mathcal{O}_E$-module, then $M_\infty := M \otimes \mathbf{C}$ decomposes as $M_\infty \cong \prod_{\tau\in\phi} M_\tau \oplus M_{\bar{\tau}}$ with $M_\sigma := M \otimes_{\mathcal{O}_E,\sigma} \mathbf{C}$ for any embedding $\sigma : E \hookrightarrow \mathbf{C}$. Similarly, $M_p := M \otimes \mathbf{Z}_p$ decomposes as $M_p \cong \prod_v M_v \oplus M_{\bar{v}}$ with $M_w := M \otimes_{\mathcal{O}_E} \mathcal{O}_{E,w}$ for a place $w|p$ of E.

Fix $i_0 \in I$. Let Φ be the Hermitian E-pairing on $V := E^3$ (viewed as column vectors) defined by the diagonal matrix $J := \mathrm{diag}(\alpha, 1, 1)$ with $\alpha \in F^\times$ such that $\tau_{i_0}(\alpha) < 0$ and $\tau_i(\alpha) > 0$ for $i \neq i_0$: $\Phi(x, y) = {}^t\bar{x}Jy$. Then Φ has signature $(2, 1)$ with respect to τ_{i_0} and signature $(3, 0)$ with respect to all other τ_i. Let $U(\Phi)_{/\mathbf{Q}}$ be the unitary group of Φ and $G := GU(\Phi)_{/\mathbf{Q}}$ its similitude group. We note that $G(\mathbf{C}) \cong \mathbf{C}^\times \times \prod_{\tau\in\phi} \mathrm{GL}_{\mathbf{C}}(V_\tau)$, where the projection to the $\mathbf{C}^\times$-factor is the similitude character, and the projection to the second factor is via the corresponding projection of $\mathrm{GL}_{E\otimes\mathbf{C}}(V_\infty)$. Similarly, $G(\mathbf{Q}_p) \cong \mathbf{Q}_p^\times \times \prod_v \mathrm{GL}_{E_v}(V_v)$, where v runs over the place of F dividing p (or the fixed places of E over these). Let $\psi := \mathrm{trace}_{E/\mathbf{Q}}\beta\Phi$ with β a totally imaginary element of E_0. Then there exists an $\mathcal{O}_E$-lattice $\Lambda \subset V$ such that ψ identifies Λ_p with its $\mathbf{Z}_p$-dual.

Let $\mathbf{S} := \mathrm{Res}_{\mathbf{C}/\mathbf{R}}\mathbf{G}_m$, so $\mathbf{S}(R) = (\mathbf{C} \otimes_{\mathbf{R}} R)^\times$ for any $\mathbf{R}$-algebra R. We identify $\mathbf{S}(\mathbf{C}) = (\mathbf{C} \otimes \mathbf{C})^\times$ with $\mathbf{C}^\times \times \mathbf{C}^\times$ via $z \otimes w \mapsto (zw, \bar{z}w)$. Let $h : \mathbf{S} \to G_{/\mathbf{R}}$ be the homomorphism such that for $(z, w) \in \mathbf{S}(\mathbf{C})$

$$h(z, w) = (zw) \times \prod_{\tau\in\phi} \begin{cases} \mathrm{diag}(z, w, w) & \tau = \tau_{i_0} \\ \mathrm{diag}(w, w, w) & \tau \neq \tau_{i_0}. \end{cases}$$

Let $h(z) = h(z, \bar{z})$. We assume that β is such that $\psi(x, h(i)x)$ is positive definite for $x \in V \otimes \mathbf{R}$. As explained in [Ko], associated with E, V, ψ, and h

is a family of PEL moduli spaces S_K over[2] E, $K \subset G(\mathbf{A}_f)$ being a neat open compact subgroup: in the notation of [Ko, §5] we take[3] $B = E$ with $*$ the non-trivial automorphism fixing F and $(V, (-,-)) = (V, \psi)$; then $C = \mathrm{End}_E(V)$ and the G of *loc. cit.* is the group G defined above, and we take for the $*$-homomorphism $\mathbf{C} \to C \otimes \mathbf{R}$ the $\mathbf{R}$-linear extension of $z \mapsto h(z)$. The varieties S_K are smooth over E and, being solutions to PEL moduli problems, are equipped with 'universal' abelian varieties A_K/S_K. As explained in [Ko, §8], S_K is naturally identified with a disjoint union of a finite number of copies of the canonical model Sh_K over E of the Shimura variety associated with G, h^{-1}, and K, indexed by the isomorphism classes of Hermitian E-spaces (V', ψ') that are everywhere locally isomorphic to (V, ψ). We identify Sh_K with the copy corresponding to the class of (V, ψ) and let A_K/Sh_K be the restriction of the universal abelian variety.

Suppose $K = K_pK^p$ with $K^p \subseteq G(\mathbf{A}_f^p)$ and $K_p \subset G(\mathbf{Q}_p)$ identified with a subgroup $\mathbf{Z}_p^\times \times \prod_{v|p} K_v \subseteq \mathbf{Z}_p^\times \times \prod_v \mathrm{GL}_{\mathcal{O}_{E,\bar v}}(\Lambda_{\bar v})$. Let $v|p$ be a fixed place. If $K_v = \mathrm{GL}_{\mathcal{O}_{E,v}}(\Lambda_v)$, then an argument of Carayol [Ca1, §5] shows that A_K and S_K have good reduction at v. A model of S_K over $\mathcal{O}_{F,v} = \mathcal{O}_{E,v}$ is obtained by considering a moduli problem as in [Ca1, 5.2.2]. To be be precise, one considers the functor from the category of locally Noetherian $\mathcal{O}_{F,v}$-schemes to the category of sets that sends an $\mathcal{O}_{F,v}$-scheme R to the set of isomorphism classes of quadruples $(A, i, \theta, \bar k^v)$ where (a) A is an abelian scheme over R of relative dimension $3d$ and $i : \mathcal{O}_E \hookrightarrow \mathrm{End}_R(A)$ is an embedding such that $\mathrm{Lie}(A)_v$ is a locally free $\mathcal{O}_R$-module of rank one on which $\mathcal{O}_{F,v} = \mathcal{O}_{E,v}$ acts via the structure map $\mathcal{O}_{F,v} \to \mathcal{O}_R$ and such that $\mathrm{Lie}(A)_{v'} = 0$ for all $v'|p$, $v' \neq v$; (b) θ is a prime-to-p polarization of A satisfying $\theta \circ i(x) = i(\bar x)^\vee \circ \theta$ for all $x \in \mathcal{O}_E$; (c) $\bar k^v$ is a K-level structure as[4] in [Ca1, 5.2.2(c)] but with $V_{\mathbf{Z}}$ in the definition of W there replaced by Λ. That this functor is isomorphic over $F_v = E_v$ to that in [Ko, §5] defining S_{K/E_v} follows from the arguments in [Ca1, 2.4-2.6,5.2.2]. That it is representable by a smooth, projective scheme $\mathcal{S}_K$ over $\mathcal{O}_{F,v}$ follows from the arguments in [Ca1, 5.3-5.5]. The p-divisible group A_p of A decomposes under the action of $\mathcal{O}_{E,p} = \mathcal{O}_E \otimes \mathbf{Z}_p$ as $A_p = \prod_{v'|p} A_{v'} \times A_{\bar v'}$. The condition on $\mathrm{Lie}(A)_{v'}$ in (a) then implies that $A_{v'}$ is ind-étale if $v' \neq v$, and part of the level structure $\bar k^v$ is a class modulo $\prod_{v' \neq v} K_{v'}$ of $\mathcal{O}_{E,p}$-linear R-isomorphisms $k_p^v : \prod_{v' \neq v} A[p^n]_{v'} \xrightarrow{\sim} \prod_{v' \neq v} (p^{-n}\Lambda/\Lambda)_{v'}$ with n any integer so large that $K_{v'}$ contains the kernel of the reduction map $\mathrm{GL}_{\mathcal{O}_{E,v'}}(\Lambda_{v'}) \to \mathrm{GL}_{\mathcal{O}_{E,v'}}(\Lambda_{v'}/p^n\Lambda_{v'})$ (see [Ca1, 5.2.3(ii)]). The condition that Λ_p is self-dual ensures that over F_v this moduli problem is equivalent to one with a usual

[2]The reflex field in this case is $\tau_{i_0}(E) \subset \mathbf{C}$ which we identify with E via τ_{i_0}.

[3]As we are only defining the moduli spaces over E at this point, the conditions at p in [Ko, §5] are superfluous.

[4]When adapting the arguments of [Ca1] to the setting of this paper, the roles of the superscripts 1 and 2 in *loc. cit.* are switched. This is a result of our choice of the homomorphism h and the identification of E with the reflex field. A homomorphism $\mathbf{S} \to G_{/R}$ more naturally generalizing that in *loc. cit.* would be $(z, w) \mapsto h(w, z)$. We have chosen h here so that Sh_K is the Shimura variety in [BR].

K-level structure. The representability of this moduli problem by a scheme $\mathcal{S}_K$ over $\mathcal{O}_{F,v}$ follows from the arguments in [Ca1, 5.3] and the properness from those in [Ca1, 5.5]. The smoothness of this scheme follows exactly as in [Ca1, 5.4]. The key point is that for R a local artinian $\mathcal{O}_{F,v}$-module, the conditions on the dimension of A and on $\mathrm{Lie}(A)_v$ in (a) imply that A_v is a divisible $\mathcal{O}_{F,v}$-module of height 3 whose formal (or connected) part has height 1 (we are keeping to the terminology in the Appendix of [Ca1]). The smoothness then follows by the deformation argument given in *loc. cit.* Over E_v, $\mathcal{S}_K$ is just S_K, and A_K is the base change of the universal abelian scheme $\mathcal{A}_K/\mathcal{S}_K$. Hence S_K, Sh_K, and A_K have good reduction at v.

2.3.2 Theorem 1 when some $k_i > 2$ and each π_v unramified

We can now explain how the arguments in [BR] yield Theorem 1 when $d > 1$, some $k_i > 2$, and each π_v, $v|p$, is unramified. Without loss of generality we may assume that $w = \max_{i\in I} k_i$; choosing a different w amounts to replacing ρ_π by a Tate-twist. We may assume that E_0 has been chosen so that the base change π_E of π to $\mathrm{GL}_2(\mathbf{A}_E)$ is cuspidal (equivalently, π is not a CM representation associated to a Hecke character of E). Fix an algebraic Hecke character μ of $\mathbf{A}_E^\times$ satisfying $\mu|_{\mathbf{A}_F^\times} = \omega_{E/F}$, the quadratic character of the extension E/F, and such that μ is unramified at each place over p. As explained[5] in [BR, Prop. 4.1.2], there exists a global L-packet τ on the quasi-split unitary group $U(2)_{/F}$ such that its non-standard base change to $\mathrm{GL}_2(\mathbf{A}_E)$ (with respect to μ) is $\pi_E \otimes \eta|\cdot|_E^{1/2}$ with η an algebraic Hecke character of $\mathbf{A}_E^\times$ that is unramified at each place above p. It follows from [BR, Lem. 4.2.1] that there exists a global character θ of $U(1)_{/F}$ unramified at all places above p for which the L-packet $\rho = \tau \otimes \theta$ of $U(2) \times U(1)$ is such that the endoscopic L-packet $\Pi(\rho_f)$ for $U(\Phi)$ contains an element σ_f with $d(\sigma_f) := \#\{\sigma_\infty \in \Pi(\rho_\infty) \ : \ \epsilon(\sigma_\infty)\epsilon(\sigma_f) = 1\} = 2$. Let χ be an algebraic character of the center of G extending the central character of $\Pi(\rho)$ and unramified at all places above p (cf. [BR, §1.2]). The pair (σ_f, χ) defines an admissible representation $\pi(\sigma_f, \chi)$ of $G(\mathbf{A}_{\mathbf{Q}}^\infty)$. From the definition of σ_f it follows that σ_p is an unramified representation of $U(\mathbf{Q}_p) \cong \prod_v \mathrm{GL}_{E_{\bar{v}}}(V_{\bar{v}})$ in the sense that it is a tensor product of unramified principal series representations of each factor. In particular, as χ is unramified at each place above p, $\pi(\sigma_f, \chi)^K \neq 0$ for $K = K_pK^p$ with K_p identified with $\mathbf{Z}_p^\times \times \prod_{v|p} \mathrm{GL}_{\mathcal{O}_{E,v}}(\Lambda_v)$ and K^p sufficiently small.

As explained in the proof of [BR, Thm. 3.3.1], associated with $\pi(\sigma_f, \chi)$ is a motive $M = (A_K^n, e)$ with coefficients in a number field $T \subset \mathbf{C}$ (this motive is denoted M_0 in *loc. cit.*; n is some integer depending on the weights of π, μ, θ, and χ, A_K^n is the n-fold self-product over the Shimura variety Sh_K, and e is an idempotent in $Z_h(A_K^n \times A_K^n)$) such that for any prime ℓ and any isomorphism

[5] The representations of $\mathrm{GL}_2(\mathbf{A}_F)$ in [BR] are normalized so that what is denoted by π there equals $\pi \otimes |\cdot|_F^{1/2}$ in this paper.

$\iota' : \mathbf{C} \xrightarrow{\sim} \overline{\mathbf{Q}}_\ell$, the ℓ-adic realization M_ℓ of M satisfies

$$M_{\ell,\iota'} := M_\ell \otimes_{T\otimes\mathbf{Q}_\ell,\iota'} \overline{\mathbf{Q}}_\ell \cong \rho_{\pi,\iota'}|_{G_E} \otimes \rho_{\eta\psi,\iota'}, \tag{5}$$

where the subscript ι' denotes that the objects on the right-hand side are the ℓ-adic Galois representations[6] associated with the embedding ι'. Here ψ is the Hecke character $z \mapsto \chi(N_{E/E_0}(\bar{z}))$ of $\mathbf{A}_E^\times$. Equivalently,

$$\mathrm{WD}(M_{\ell,\iota'}|_{D_w}) \cong \iota'\mathrm{Rec}_w(\pi_{E,w} \otimes \eta_w\psi_w|\cdot|_w^{-1/2}) \tag{6}$$

for all places $w \nmid \ell$ of E, D_w being any decomposition group for w. More precisely, (6) is only shown in [BR, Thm. 3.3.1] for those $w \nmid \ell$ coprime to the conductor of π and the absolute discriminant of E. But this together with the existence of the ℓ-adic representations associated with π, $\eta\psi$, and ι' implies (5), from which (6) follows for all places $w \nmid \ell$. This relies on more than is proved in *loc. cit.*; it also requires the work of Carayol and Taylor on the existence of the ℓ-adic representations.

As A_K has good reduction at $v|p$, it follows - from the theorems of Faltings and of Katz and Messing cited in [BR, §5] together with (5) and (6) - that for a place $v|p$ of F the representation $M_{p,\iota}$ is crystalline at v and for all $w \in W_v$

$$\begin{aligned}\det(1 - X\varphi^{\nu(w)}|D_{cris}(M_{p,\iota}|_{D_v}) \otimes_{\overline{\mathbf{Q}}_p\otimes_{\mathbf{Q}_p}F_{v,0},\tau} \overline{\mathbf{Q}}_p)&\\ = \det(1 - Xw|\iota\mathrm{Rec}_v(\pi_v \otimes \eta_v\psi_v|\cdot|_v^{-1/2}).&\end{aligned} \tag{7}$$

As η and ψ are both unramified at all places above p, $\rho_{\eta\psi|\cdot|_E}$ is crystalline at v. It then follows that $\rho_v \cong (M_{p,\iota} \otimes \rho_{\eta\psi}^{-1})|_{D_v}$ is crystalline, and so (3) follows from (7). That ρ_v has Hodge-Tate type $(\mathbf{k}, w)$ is immediate from [BR, Thm. 2.5.1(ii)] and Faltings' proof of the deRham conjecture.

2.3.3 Essentially self-dual representations of $\mathrm{GL}_3(\mathbf{A}_F)$

Let $\Pi = \otimes'\Pi_v$ be a cuspidal automorphic representation of $\mathrm{GL}_3(\mathbf{A}_F)$ for which each Π_i, $i \in I$, is such that its corresponding representation $\mathrm{Rec}_i(\Pi_i)$ of the Weil group of F_i satisfies

$$\mathrm{Rec}_i(\Pi_i)|_{\mathbf{C}^\times} \cong z^{a_i}\bar{z}^{b_i} \oplus z^{b_i}\bar{z}^{a_i} \oplus (z\bar{z})^{(a_i+b_i)/2}, \quad a_i \neq b_i \in \mathbf{Z}, a_i + b_i \in 2\mathbf{Z}. \tag{8}$$

Suppose also that $\Pi^\vee \cong \Pi\otimes\psi$ for some Hecke character ψ (then ψ is necessarily algebraic). As explained in [B1, 4.1-4.6], it is a consequence of the results in [Mo] that for each prime ℓ and each isomorphism $\iota' : \mathbf{C} \xrightarrow{\sim} \overline{\mathbf{Q}}_\ell$ there is an ℓ-adic Galois representation $\rho_{\Pi,\iota'} : G_F \to \mathrm{GL}_3(\overline{\mathbf{Q}}_\ell)$ satisfying $\mathrm{WD}(\rho_{\Pi,\iota'}|_{D_v}) \cong$

[6] For an algebraic Hecke character ψ of a number field, we denote by $\rho_{\psi,\iota'}$ the ℓ-adic Galois representation associated with ψ and ι', normalized so that the restriction of the Galois character to the decomposition group at a place $w \nmid \ell$ is just the image of the local character ψ_w under the inverse of the Artin map, composed with ι'.

$\iota'\mathrm{Rec}_v(\Pi_v)$ for all places $v \nmid \ell$ that are prime to the conductor of Π and the absolute discriminant of F.

The proof of the existence of $\rho_{\Pi,\iota'}$ follows the arguments in [BR]. In particular, letting E be as in 2.3.2, if the base change of Π to E is still cuspidal then, as explained in the proof of [B1, Thm. 4.2], there is a motive $M = (A_K^m, e)$, K small enough, such that the ℓ-adic realizations of M yield $\rho_\Pi|_{G_E}$ twisted by a representation associated with an algebraic Hecke character of $\mathbf{A}_E^\times$. If Π is unramified at each $v|p$ then one can take $K = K_pK^p$ with K_p identified with $\mathbf{Z}_p^\times \times \prod_{v|p} \mathrm{GL}_{\mathcal{O}_{E,v}}(\Lambda_v)$ and the Hecke character can be taken unramified at each $v|p$. Then arguing as in 2.3.2 shows that $\rho_\Pi := \rho_{\Pi,\iota}$ is crystalline at each $v|p$ and such that $D_{HT}(\rho_\Pi|_{D_v}) \otimes_{\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_v, j} \overline{\mathbf{Q}}_p$, $j \in \mathrm{Hom}_{\mathbf{Q}_p}(F_v, \overline{\mathbf{Q}}_p)$, is non-zero in degrees $-a_{i(j)}, -b_{i(j)}$, and $-(a_{i(j)} + b_{i(j)})/2$, $i(j) \in I$ being the induced embedding of F. Furthermore, if $\mathrm{WD}(\rho_{\Pi,\iota'}|_{D_v}) \cong \iota'\mathrm{Rec}_v(\Pi_v)$ for some $\ell \neq p$ (only an additional condition if p is not prime to the absolute discriminant of F), then these arguments also show that $\mathrm{WD}(\rho_\Pi|_{D_v}) \cong \iota\mathrm{Rec}_v(\Pi_v)$.

Remark. Suppose Π_v is unramified at each $v|p$. From the good reduction of the Shimura variety Sh_K with K_p as in 2.3.2 or 2.3.3, it follows easily from the Weil conjectures that the Frobenius-at-v eigenvalues of any ℓ-adic representation $\rho_{\Pi,\iota'}$, $\ell \neq p$, have absolute value as predicted by the Ramanujan conjecture for Π_v when considered as elements of $\mathbf{C}$ via ι'. Therefore, if $\mathrm{WD}(\rho_{\Pi,\iota'}|_{D_v}) \cong \iota'\mathrm{Rec}_v(\Pi_v)$, then the Ramanujan conjecture is true for Π_v. This argument shows (at least) that if q is a prime such that Π_w is unramified for all $w|q$, then the Ramanujan conjecture is true for Π_w, $w|q$, provided there is some prime $\ell \neq q$ such that the ℓ-adic representation $\rho_{\Pi,\iota'}$ satisfies $\mathrm{WD}(\rho_{\Pi,\iota'}|_{D_w}) \cong \iota'\mathrm{Rec}_w(\Pi_w)$.

2.4 Theorem 1 for the remaining cases

As a consequence of the work of Saito [Sa1, Sa2], the remarks in 2.2, and the results of [BR] as described in 2.3.2, to complete the proof of Theorem 1 it remains to consider the case where d is even, each $k_i = 2$, each π_v, $v|p$, is unramified, and π is not a CM representation. Replacing π by a twist by an integral power of $|\cdot|_F$ if necessary (which corresponds to twisting ρ_π by a power of the cyclotomic character), we may also assume that $w = 2$. Hereon we assume we are in this case.

2.4.1 An application of the symmetric square

Let $\Pi := \mathrm{Sym}^2\pi \otimes |\cdot|_F^{-1}$, with $\mathrm{Sym}^2\pi$ the symmetric square lift of π to $\mathrm{GL}_3(\mathbf{A}_F)$ (cf. [GJ]). As π is not a CM representation, Π is cuspidal. Since $\mathrm{Rec}_i(\pi_i)|_{\mathbf{C}^\times} \cong (\bar{z}/z)^{1/2} \bar{\oplus} (z/\bar{z})^{1/2}$, $\mathrm{Rec}_i(\Pi_i)|_{\mathbf{C}^\times} \cong \mathrm{Sym}^2\mathrm{Rec}_i(\pi_i|\cdot|_i^{-1/2})|_{\mathbf{C}^\times}$ satisfies (8) with $a_i = -2$ and $b_i = 0$. Furthermore, as $\pi^\vee \cong \pi \otimes \omega^{-1}$, ω the central character of π, it follows that $\Pi^\vee \cong \Pi \otimes \omega^{-2}|\cdot|_F^2$. Therefore, Π satisfies all the hypotheses in 2.3.3. In particular, there exist associated ℓ-adic representations

$\rho_{\Pi,\iota'}$. Clearly $\rho_{\Pi,\iota'} \cong \mathrm{Sym}^2\rho_{\pi,\iota'}$, so $\mathrm{WD}(\rho_{\Pi,\iota'}|_{D_w}) \cong \iota'\mathrm{Rec}_w(\mathrm{Sym}^2\pi_w \otimes |\cdot|_w^{-1})$ for all $w \nmid \ell$. Since π_v, and therefore Π_v, is unramified at each $v|p$, as explained in 2.3.3 we can conclude from this that for each $v|p$: (i) $\mathrm{Sym}^2\rho_\pi|_{D_v}$ is crystalline for $v|p$, (ii) $\mathrm{WD}(\mathrm{Sym}^2\rho_\pi|_{D_v}) \cong \iota\mathrm{Rec}_v(\mathrm{Sym}^2\pi_v \otimes |\cdot|_v^{-1})$, and (iii) $D_{HT}(\mathrm{Sym}^2\rho_\pi|_{D_v}) \otimes_{\overline{\mathbf{Q}}_p\otimes_{\mathbf{Q}_p}F_v,j} \overline{\mathbf{Q}}_p$, $j \in \mathrm{Hom}_{\mathbf{Q}_p}(F_v, \overline{\mathbf{Q}}_p)$, is non-zero in degrees 2, 1, and 0.

Let $v|p$. By conclusion (iii) of the preceding paragraph, the graded module $D_{HT}(\mathrm{Sym}^2\rho_v)$ is the symmetric square of the expected graded module for ρ_v. It then follows from results of Wintenberger[7] - Thm. 1.1.3, Prop. 1.2, and Remarks 1.1.4 of [Win1] or Thm. 2.2.2 of [Win2], applied to the isogeny $\mathrm{GL}_2 \to \mathrm{GL}_2/\pm 1$ - that there is a crystalline representation $\rho : D_v \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ such that $\mathrm{Sym}^2\rho = \mathrm{Sym}^2\rho_v$. From this it follows that ρ_v is isomorphic to a (possibly trivial) quadratic twist of ρ. In particular, ρ_v is potentially crystalline. Therefore $\mathrm{WD}(\mathrm{Sym}^2\rho_v) \cong \mathrm{Sym}^2\mathrm{WD}(\rho_v)$, and it then follows from conclusion (ii) of the preceding paragraph that $\mathrm{Sym}^2\mathrm{WD}(\rho_v) \cong \mathrm{Sym}^2\iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})$. From this it follows that $\mathrm{WD}(\rho_v)$ is isomorphic to a (possibly trivial) quadratic twist of $\iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})$. It also follows that ρ_v is of Hodge-Tate type $(\mathbf{k}, w)$ $(= ((2)_{i\in I}, 2)$ in this case).

Remark. We can also use $\mathrm{Sym}^2\pi$ to show that the Ramanujan conjecture holds for π. We may assume that π is not a CM representation. Let q be a prime. It then follows from the remark at the end of 2.3.3 that if π_w is unramified at each $w|q$, then the Ramanujan conjecture holds for each $\mathrm{Sym}^2\pi_w$ and hence for π_w. A simple base change argument like that in 2.2 then shows that the Ramanujan conjecture holds at all places where π is a principal series. In particular, this establishes the Ramanujan conjecture for those π for which there is no finite place v with π_v square-integrable. That the Ramanujan conjecture is known when such a v exists follows from Carayol's work [Ca2]. The Ramanujan conjecture has already been established for π by Blasius [B2].

2.4.2 The existence of a crystalline period

Recall that we are assuming that for each $v|p$, $\pi_v \cong \pi(\alpha_v, \beta_v)$ is an unramified principal series[8]. As $\mathrm{WD}(\rho_v)$ is isomorphic to a (possibly trivial) quadratic twist of $\iota\mathrm{Rec}_v(\pi_v \otimes |\cdot|_v^{-1/2})$, to prove (2) it suffices to show that $\mathrm{WD}(\rho_v)^{\mathrm{frob}_v=\alpha_v(\varpi_v)q_v^{1/2}} \neq 0$, where frob_v is a geometric frobenius at v, ϖ_v is a uniformizer at v, and q_v is the order of the residue field at v. This is equivalent to showing that $D_{cris}(\rho_v|_{D_v})^{\varphi^{f_v}=q_v^{1/2}\alpha_v(\varpi_v)}$ is a $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}$-module of rank

[7]Note that 'weakly admissible = admissible' has been proved by Colmez and Fontaine, and 'de Rham = potentially semistable' has been proved (independently in some cases) by André, Berger, Kedlaya, and Mebkhout, and so the hypotheses on which these results depend are known to hold.

[8]By $\pi(\alpha, \beta)$ we mean the usual principal series representation that is the induction to $\mathrm{GL}_2(F_v)$ of the character $\left(\begin{smallmatrix} a & * \\ 0 & d \end{smallmatrix}\right) \mapsto \alpha(a)\beta(d)|a/d|_v^{1/2}$ of the upper-triangular Borel.

at least one. To establish a lower bound on this rank, we make use of p-adic analytic families of cuspidal representations.

Let $\mathcal{O}$ denote the integer ring of F and let $\mathcal{O}_p := \mathcal{O} \otimes \mathbf{Z}_p \xrightarrow{\sim} \prod_{v|p} \mathcal{O}_v$. Let $S_p := \{v|p\}$ be the set of places of F over p and let S_π be the set of finite places of F at which π is ramified. Let $S := S_\pi \cup S_p$ and $K^S := \prod_{v \notin S, v\nmid\infty} \mathrm{GL}_2(\mathcal{O}_v)$. Let $\mathcal{H}^S$ be the abelian Hecke algebra

$$\mathcal{H}^S := C_c(\mathrm{GL}_2(\mathbf{A}_{F,f}^S)//K^S).$$

For each $v \in S_p$ let

$$I_v := \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathcal{O}_v) \ : \ \varpi_v | c\}, \quad I_p := \prod_{v|p} I_v,$$

and let $\mathcal{U}_v \subset C_c(\mathrm{GL}_2(\mathcal{O}_v)//I_v)$ be the abelian subalgebra generated by the characteristic functions

$$U_v := \mathrm{char}(I_v \mathrm{diag}(\varpi_v, 1) I_v).$$

Put

$$\mathcal{U}_p := \otimes_{v|p} \mathcal{U}_p \quad \text{and} \quad \mathbf{T}^S := \mathcal{U}_p \otimes \mathcal{H}^S.$$

Then there exists an $f_\pi \in \pi^{K^S I_p}$ that is an eigenvector for the (usual) action of the Hecke ring $\mathbf{T}^S$ such that $\mathrm{char}(I_v \mathrm{diag}(\varpi_v, 1) I_v)$ acts with eigenvalue $q_v^{1/2} \alpha_v(\varpi_v)$.

Let $K \subset \overline{\mathbf{Q}}_p$ be a finite extension of $\mathbf{Q}_p$ containing each $i(F)$, $i \in I$, and the eigenvalues for the action of $\mathbf{T}^S$ on f_π. Let $|K^\times| = \{|x|_p \ : \ x \in K^\times\}$. For $r \in |K^\times|$, we denote by B_r the usual closed rigid ball over K of radius r (so $B_r(\mathbf{C}_p) = \{x \in \mathbf{C}_p \ : \ |x|_p \leq r\}$, where $\mathbf{C}_p := \widehat{\overline{\mathbf{Q}}}_p$). Then $\mathcal{O}(B_1) = K < T >$. Let $A_r := \mathcal{O}(B_r)$; this is an affinoid K-algebra. From the work of Buzzard [Bu1, Bu2] one can deduce that if $r_0 \in |K^\times|$ is sufficiently small, then there exists a reduced finite torsion-free A_{r_0}-algebra $\mathcal{R}$ (so also an affinoid K-algebra) and a homomorphism $\phi : \mathbf{T}^S \to \mathcal{R}$ satisfying (i)-(iii) below. For $x \in \mathrm{Hom}_K(\mathcal{R}, \overline{\mathbf{Q}}_p)$ put $\phi_x := x \circ \phi$. Then:

(i) if x is such that $x(1+T) = (1+q)^{n_x}$, $n_x \in p(p-1)\mathbf{Z}_{>0}$ ($q = p$ if p odd and $q = 4$ if $p = 2$), then there exists a cuspidal representation π_x of $\mathrm{GL}_2(\mathbf{A}_F)$ with infinity type $(\mathbf{k}_x, w_x) = ((n_x+2)_{i\in I}, n_x+2))$ and which is unramified at all $v|p$ and such that $\phi_x : \mathbf{T}^S \to \overline{\mathbf{Q}}_p$ gives the eigenvalues of the action of $\mathbf{T}^S$ on an eigenvector $f_x \in \pi_x^{K^S I_p}$;

(ii) there exists $x_0 \in \mathcal{X}(K)$ with $x_0(1+T) = 1$ such that ϕ_{x_0} gives the eigenvalues of the action of $\mathbf{T}^S$ on f_π;

(iii) if $\phi_v := \phi(U_v) \in \mathcal{R}^\times$, then $|x(\phi_v)|_p$ is constant for all x;

(iv) there exists a continuous representation

$$\rho_{\mathcal{R}} : G_F \to \mathrm{GL}_2(\mathcal{R})$$

unramified away from S and such that for x as in (i) the representation $\rho_x : G_F \to \mathrm{GL}_2(\overline{\mathbf{Q}}_p)$ induced from $\rho_{\mathcal{R}}$ by x is equivalent to ρ_{π_x} and that induced by x_0 is equivalent to ρ_π.

Assuming the existence of $\mathcal{R}$ and ϕ, we can complete the proof of Theorem 1. Let $\Sigma \subset \mathrm{Hom}_K(\mathcal{R}, \overline{\mathbf{Q}}_p)$ be the set of x as in (i). Then Σ is Zariski dense by the finiteness of $\mathcal{R}$ over A_r. As explained in 2.3.2 we know that Theorem 1 holds for each π_x, $x \in \Sigma$. Let $v|p$ and $x \in \Sigma$. Then $\pi_{x,v} \cong \pi(\mu_x, \lambda_x)$, an unramified principal series with $x(\phi_v) = \mu_x(\varpi_v) q_v^{1/2}$. In particular, as Theorem 1 holds for $\rho_x \cong \rho_{\pi_x}$ we have that $D_{cris}(\rho_x|_{D_v})^{\varphi^{f_v} = x(\phi_v)}$ is a $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}$-module of rank at least one for all $x \in \Sigma$, where f_v is the residue class degree of F_v (so $q_v = p^{f_v}$). As the Hodge-Tate type of ρ_x, $x \in \Sigma$, is $(\mathbf{k}_x, w_x)$, each $D_{HT}(\rho_x|_{D_v}) \otimes_{\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_v, j} \overline{\mathbf{Q}}_p$ is non-zero in degrees 0 and $n_x + 1$. It then follows easily from [Ki2, (5.15)] that[9]

$$D_{cris}(\rho_\pi|_{D_v})^{\varphi^{f_v} = q_v^{1/2}\alpha_v(\varpi_v)} = D_{cris}(\rho_{x_0}|_{D_v})^{\varphi^{f_v} = x_0(\phi_v)}$$

is also a $\overline{\mathbf{Q}}_p \otimes_{\mathbf{Q}_p} F_{v,0}$-module of rank at least one.

While the existence of $\mathcal{R}$ and ϕ is essentially proved in the work of Buzzard, there is no convenient reference in [Bu1]. So we conclude by explaining how their existence follows from this work. Let D be the quaternion algebra over F that is split at all finite places and compact modulo the center at all archimedean places. Fix a maximal order $\mathcal{O}_D$ of D, and for each finite place v of F fix an isomorphism $\mathcal{O}_{D,v} \cong M_2(\mathcal{O}_v)$. This identifies $\mathrm{GL}_2(\mathbf{A}_{F,f})$ with $(D \otimes_F \mathbf{A}_{F,f})^\times$. Let $\mathfrak{n}$ be the conductor of π and let $U_0 \subseteq \mathrm{GL}_2(\mathcal{O} \otimes \widehat{\mathbf{Z}})$ be the subgroup of matrices with lower left entries in $\mathfrak{n} \otimes \widehat{\mathbf{Z}}$, and let $U = U_0 \cap I_p$. Let $J := \{v|p\}$. For $\mathbf{a} \in \mathbf{Z}^J_{>0}$ let $U_{\mathbf{a}} := \prod_{v \in J} U_v^{a_v}$. For $v \in J$ let $\sigma_v := \mathrm{ord}_v(\alpha_v(\varpi_v) q_v^{1/2})$, and let $\sigma_{\mathbf{a}} := \sum_{v \in J} a_v \sigma_v$.

For $r \in |K^\times|$ with $r \leq 1$ we define a homomorphism $\kappa : \mathcal{O}_p^\times \times \mathcal{O}_p^\times \to A_r^\times$ by

$$\kappa((x_v),(y_v)) = \prod_{v \in S_p} \prod_{j \in \mathrm{Hom}_{\mathbf{Q}_p}(F_v, \overline{\mathbf{Q}}_p)} j(y_v)(1+T)^{\log_p \mathrm{Nm}_{F_v/\mathbf{Q}_p}(x_v)}.$$

[9] Proposition (5.14) and Corollary (5.15) of [Ki2] are only stated for representations of $G_{\mathbf{Q}_p} = \mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. But it is easily checked that the arguments extend to the case of the representations of $D_v = \mathrm{Gal}(\overline{F}_v/F_v)$ under consideration here; the necessary results with φ replaced by φ^{f_v} (e.g., Corollary (3.7)) are easily deduced from those for φ. A key point is that our hypotheses on the weights in the family $\mathcal{X}$ ensure that the polynomial $P(X) \in (\mathcal{O}(\mathcal{X}) \otimes_{\mathbf{Q}_p} F_v)[X]$ provided by Sen's theory as in [Ki2, (2.2)] is of the form $P(X) = XQ(X)$ with the constant coefficient of Q not a zero-divisor.

Let $\mathcal{W}$ be the rigid analytic weight space over K defined in §8 of [Bu1]. Then B_r is identified with a reduced affinoid subspace of $\mathcal{W}$ such that κ is the induced weight in the sense of *loc. cit.* Let $m \in |K^\times|$ be so small that the A_r-Banach module $\mathbf{S}^D_\kappa(U;m)$ of overconvergent automorphic forms is defined (notation as in [Bu1, §9]). This is equipped with an A_r-linear action of $\mathbf{T}^S$ such that each U_v is a completely continuous operator. For $b \in B_r(\overline{\mathbf{Q}}_p)$ such that the induced map $e_b : A_r \to \overline{\mathbf{Q}}_p$ sends $1+T$ to $(1+q)^{n_b}$ with $n_b \in p(p-1)\mathbf{Z}_{\geq 0}$ we have a $\mathbf{T}^S$-equivariant inclusion of the classical forms of weight $(\mathbf{k}_b, \mathbf{w}_b)$: $S^D_{\mathbf{k}_b,\mathbf{w}_b}(U) \subseteq \mathbf{S}^D_\kappa(U;m) \otimes_{\mathcal{O}(B_r),e_b} \overline{\mathbf{Q}}_p$, $\mathbf{w}_b := (n_b+2)_{i\in I} \in \mathbf{Z}^I$ (see [Bu1, §11]). By the Jacquet-Langlands correspondence, there exists $f_0 \in S_{2,2}(U)$ having the same $\mathbf{T}^S$-eigenvalues as f_π. Recall that by the theory of Fredholm series and orthonormalizable Banach modules as developed by Coleman, Ash and Stevens, and Buzzard, if r is small enough then there is a finite A_r-direct summand $\mathcal{N} \subset \mathbf{S}^D_\kappa(U;m)$ that is stable under $\mathbf{T}^S$ and such that for each $\mathbf{a} \in \mathbf{Z}^J_{>0}$ the Fredholm series for $U_\mathbf{a}$ on $\mathcal{N}$ is a factor of the slope $\sigma_\mathbf{a}$ part of the Fredholm series $P_\mathbf{a}(X) \in A_r\{X\}$ associated to the completely continuous operator $U_\mathbf{a}$ on $\mathbf{S}^D_\kappa(U;m)$ (the latter is well-defined for r small enough), and furthermore is such that $f_0 \in \mathcal{N} \otimes_{A_r,e_0} K$. If r is sufficiently small then for any $b \in A_r$ with $n_b \in p(p-1)\mathbf{Z}_{>0}$ it follows from the arguments in [Bu2, §7] (see also the comment at the end of §11 of [Bu1]) that $N_b := \mathcal{N} \otimes_{A_r,e_b} \overline{\mathbf{Q}}_p$ is comprised of classical forms in $S^D_{n_b+2,n_b+2}(U)$ (n_b is divisible by a high power of p; the smaller r is, the larger the power of p). By the definition of $\mathcal{N}$, any $\mathbf{T}^S$-eigenform in N_b is such that the eigenvalue of U_v has slope σ_v, and so if r is small enough relative to σ_v then it is easily seen that the v-constituent of the irreducible representation of $\mathrm{GL}_2(\mathbf{A}_{F,f})$ generated by f is not special and therefore must be an unramified principal series.

Let R be the A_r-algebra generated by the image of $\mathbf{T}^S$ in $\mathrm{End}_{A_r}(\mathcal{N})$; this is a finite torsion-free A_r-algebra and so an affinoid K-algebra. Note that there exists a K-homomorphism $\phi_0 : R \to K$ giving the eigenvalues of the $\mathbf{T}^S$-action on f_0. Let A be the normalization of the quotient of R by a minimal prime containing the kernel of ϕ_0. This is a reduced finite torsion-free A_r-algebra and so also an affinoid K-algebra. Let $\phi : \mathbf{T}^S \to A$ be the canonical homomorphism. It follows from the definitions that (i), (ii), and (iii) hold with $\mathcal{R}$ replaced by A. For each $x \in \mathrm{Hom}_K(A, \overline{\mathbf{Q}}_p)$ as in (i), let $T_x : G_F \to \overline{\mathbf{Q}}_p$ be the continuous pseudo-representation associated with ρ_{π_x} (so $T_x = \mathrm{trace}\,\rho_{\pi_x}$). Since for a place $w \nmid \mathfrak{n}p$, $T_x(\mathrm{frob}_w) = x \circ \phi(\mathrm{char}(\mathrm{GL}_2(\mathcal{O}_w)\mathrm{diag}(\varpi_w,1)\mathrm{GL}_2(\mathcal{O}_w))$, $\varpi_w \in \mathcal{O}_w$ a uniformizer, it follows easily from the Cebotarev density theorem and the Zariski density of the set Σ_A of $x \in \mathrm{Hom}_K(A,\overline{\mathbf{Q}}_p)$ as in (i) that there is a continuous pseudo-representation $T : G_F \to A$ such that $T_x = x \circ T$. From the general theory of pseudo-representations (cf. [Tay3]) there is a semisimple Galois representation $\rho_A : G_F \to \mathrm{GL}_2(F_A)$, F_A the field of fractions of A, such that $T = \mathrm{trace}\,\rho_A$. It is easy to see that there is a finite A-module $M \subset F_A^2$ on which G_F acts continuously and such that $V_x := M_x \otimes_{A_x,x} \overline{\mathbf{Q}}_p$ is isomorphic to the representation ρ_{π_x}, $x \in \Sigma_A$ or x any extension of ϕ_0 to A (here the subscript x on M and A denotes the localization at the kernel of x). Such a module M is given explicitly as follows. Fix a basis of ρ_A such that for some $i \in I$ the corresponding complex conjugation in G_F is diagonalized (with eigenvalues 1 and -1). Writing $\rho_A(\sigma) = \left(\begin{smallmatrix} a_\sigma & b_\sigma \\ c_\sigma & d_\sigma \end{smallmatrix}\right)$, we have $a_\sigma, d_\sigma, b_\sigma c_{\sigma'} \in A$ for all $\sigma, \sigma' \in G_F$ and that these define continuous functions of σ and σ'. It follows that the A-submodules $\mathcal{B}$ and $\mathcal{C}$ of F_A generated by $\{b_\sigma \ : \sigma \in G_F\}$ and $\{c_\sigma \ : \sigma \in G_F\}$, respectively, are fractional ideals of A satisfying $\mathcal{CB} \subseteq A$ (note that by the semisimplicity of ρ_A and the diagonalization of the chosen complex conjugation, $\mathcal{B} = 0$ if and only of $\mathcal{C} = 0$). We can then take $M = A \oplus A$ if $\mathcal{C} = 0$ and $M = A \oplus \mathcal{C}$ otherwise. Being a finite A-module, M is a Banach A-module and the continuity of the action of G_F on M is clear from the continuity of the functions a_σ, d_σ, and $b_\sigma c_{\sigma'}$. As A is normal, for any $x \in \mathrm{Hom}_K(A, \overline{\mathbf{Q}}_p)$ the localization A_x is a DVR, and so M_x is a free A_x-module of rank two. The representation V_x is then two-dimensional and its associated pseudo-representation is $x \circ T$. Therefore if $x \in \Sigma_A$ or x any extension of ϕ_0 to A, the pseudo-representation associated with V_x equals that associated with ρ_{π_x}. As the latter representation is irreducible (this irreducibility is well-known, but see also the remark below) it follows that $V_x \cong \rho_{\pi_x}$. As A is normal and finite over A_r, there is an $f \notin TA_r$ (in fact one can pick f not to be zero on any given finite set of points of B_r) such that M_f is free over A_f. Let $r_0 \leq r$ be so small that $f \in A_{r_0}^\times$. Then (i)-(iv) hold with $\mathcal{R}$ the quotient of $A \otimes_{A_r} A_{r_0}$ by any minimal prime (a finite A_{r_0}-algebra and so an affinoid K-algebra) and with $\rho_{\mathcal{R}}$ the representation of G_F on the free $\mathcal{R}$-module $M \otimes_A \mathcal{R}$.

Remark. We recall that there is a quick proof of the irreducibility of ρ_π using that it is potentially semistable (really only that it is Hodge-Tate), which was established in 2.4.1. If $\rho_\pi \cong \chi_1 \oplus \chi_2$, then each χ_i is potentially semistable and hence is the Galois representation associated to an algebraic Hecke character ψ_i of F (cf. [Se], esp. III,2.3-2.4). It then follows that $L(\pi \otimes \psi_2^{-1}, s - 1/2) = L(\psi_1/\psi_2, s)\zeta_F(s)$. As $\psi_i = \psi_i' | \cdot |_F^{a_i}$ with $a_i \in \mathbf{Z}$ and ψ_i' finite and since we may assume $a_1 \geq a_2$, $L(\psi_1/\psi_2, 1) = L(\psi_1'/\psi_2', a_1 - a_2 + 1) \neq 0$. But this implies that $L(\pi \otimes \psi_2^{-1}, s)$ has a pole at $s = 1$, contradicting the cuspidality of π.

References

[B1] D. Blasius, *Automorphic forms and Galois representations: some examples*, Automorphic forms, Shimura varieties, and L-functions, Vol. II (Ann Arbor, MI, 1988), Perspect. Math., vol 11, Academic Press, Boston, MA, 1990, pp. 1–13.

[B2] D. Blasius, *Hilbert modular forms and the Ramanujan conjecture*, Noncommutative geometry and number theory, Aspects Math., E37, Vieweg, Wiesbaden, 2006, pp. 35–56.

[BR] D. Blasius and J. Rogawski, *Motives for Hilbert modular forms*, Invent. Math. 114 (1993), no. 1, 55–87.

[Br] C. Breuil, *Une remarque sur les représentations locales p-adiques et les congruences entre formes modulaires de Hilbert*, Bull. Soc. Math. France 127 (1999), no. 3, 459–472.

[Bu1] K. Buzzard, *Eigenvarieties*, L-functions and Galois representations, London Math. Soc. Lecture Note Ser., vol. 320, Cambridge Univ. Press, Cambridge, 2007, pp. 59–120.

[Bu2] K. Buzzard,*On p-adic families of automorphic forms*, Modular curves and abelian varieties, Progr. Math., vol. 224, Birkhuser, Basel, 2004, pp. 23–44.

[Ca1] H. Carayol, *Sur la mauvaise rduction des courbes de Shimura*, Compositio Math. 59 (1986), no. 2, 151–230.

[Ca2] H. Carayol, *Sur les représentations l-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. Ecole Norm. Sup. (4) 19 (1986), no. 3, 409–468.

[Fo] J.-M. Fontaine, *Arithmtique des représentations galoisiennes p-adiques*, Cohomologies p-adiques et applications arithmétiques. III. AstÉrisque No. 295 (2004), xi, 1–115.

[GJ] S. Gelbart and H. Jacquet, *A relation between automorphic representations of* GL(2) *and* GL(3), Ann. Sci. cole Norm. Sup. (4) 11 (1978), no. 4, 471–542.

[Ki1] M. Kisin, *Potentially semi-stable deformation rings*, J. Amer. Math. Soc. 21 (2008), no. 2, 513–546.

[Ki2] M. Kisin, *Overconvergent modular forms and the Fontaine-Mazur conjecture*, Invent. Math. 153 (2003), no. 2, 373–454.

[Ko] R. Kottwitz, *Points on some Shimura varieties over finite fields*, J. Amer. Math. Soc. 5 (1992) no. 2, 373–444.

[L] T. Liu, *Lattices in filtered (φ, N)-modules*, preprint 2008.

[Mo] *The zeta functions of Picard modular surfaces*, edited by Robert P. Langlands and Dinakar Ramakrishnan, Universit de Montrèal, Centre de Recherches Mathèmatiques, Montreal, QC, 1992.

[Sa1] T. Saito, *Modular forms and p-adic Hodge theory*, Invent. Math. 129 (1997), no. 3, 607–620.

[Sa2] T. Saito, *Hilbert modular forms and p-adic Hodge theory*, preprint (2006).

[Se] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, W. A. Benjamin, Inc., New York-Amsterdam 1968 xvi+177

[Ta] J. Tate, *Number theoretic background*, Automorphic forms, representations and L-functions, Part 2, Proc. Sympos. Pure Math., vol. XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 3–26.

[Tay1] R. Taylor, *On Galois representations associated to Hilbert modular forms*, Invent. Math. 98 (1989), no. 2, 265–280.

[Tay2] R. Taylor, *On Galois representations associated to Hilbert modular forms. II*, Elliptic curves, modular forms, and Fermat's last theorem (Hong Kong, 1993), Ser. Number Theory, I, Int. Press, Cambridge, MA, 1995, pp. 185–191.

[Tay3] R Taylor, *Galois representations associated to siegel modular forms of low weight*, Duke Math. J. 63 (1991), no. 2, 281–332.

[W] A. Wiles, *On ordinary λ-adic representations associated to modular forms*, Invent. Math. 94 (1988), no. 3, 529–573.

[Win1] J.-P. Wintenberger, *Relëvement selon une isogènie de systëmes l-adiques de reprsentations galoisiennes associès aux motifs*, Invent. Math. 120 (1995), no. 2, 215–240.

[Win2] J.-P. Wintenberger, *Propriètès du groupe tannakien des structures de Hodge p-adiques et torseur entre cohomologies cristalline et tale*, Ann. Inst. Fourier (Grenoble) 47 (1997), no. 5, 1289–1334.

Christopher Skinner
Department of Mathematics
Princeton University
Fine Hall, Washington Road
Princeton, NJ 08544-1000
USA

Equivariant Counts of Points of the Moduli Spaces of Pointed Hyperelliptic Curves

Jonas Bergström

Received: June 18, 2008
Revised: March 27, 2009

Communicated by Thomas Peternell

Abstract. We consider the moduli space $\mathcal{H}_{g,n}$ of n-pointed smooth hyperelliptic curves of genus g. In order to get cohomological information we wish to make $\mathbb{S}_n$-equivariant counts of the numbers of points defined over finite fields of this moduli space. We find recurrence relations in the genus that these numbers fulfill. Thus, if we can make $\mathbb{S}_n$-equivariant counts of $\mathcal{H}_{g,n}$ for low genus, then we can do this for every genus. Information about curves of genus 0 and 1 is then found to be sufficient to compute the answers for $\mathcal{H}_{g,n}$ for all g and for $n \leq 7$. These results are applied to the moduli spaces of stable curves of genus 2 with up to 7 points, and this gives us the $\mathbb{S}_n$-equivariant Galois (resp. Hodge) structure of their ℓ-adic (resp. Betti) cohomology.

2000 Mathematics Subject Classification: 14H10, 11G20
Keywords and Phrases: Cohomology of moduli spaces of curves, curves over finite fields.

1. Introduction

By virtue of the Lefschetz trace formula, counting points defined over finite fields of a space gives a way of finding information on its cohomology. In this article we wish to count points of the moduli space $\mathcal{H}_{g,n}$ of n-pointed smooth hyperelliptic curves of genus g. On this space we have an action of the symmetric group $\mathbb{S}_n$ by permuting the marked points of the curves. To take this action into account we will make $\mathbb{S}_n$-*equivariant* counts of the numbers of points of $\mathcal{H}_{g,n}$ defined over finite fields.
For every n we will find simple recurrence relations in the genus, for the equivariant number of points of $\mathcal{H}_{g,n}$ defined over a finite field. Thus, if we can count these numbers for low genus, we will know the answer for every genus. The hyperelliptic curves will need to be separated according to whether the

characteristic is odd or even and the respective recurrence relations will in some cases be different.

When the number of marked points is at most 7 we use the fact that the base cases of the recurrence relations only involve the genus 0 case, which is easily computed, and previously known $\mathbb{S}_n$-equivariant counts of points of $\mathcal{M}_{1,n}$, to get equivariant counts for every genus. If we consider the odd and even cases separately, then all these counts are polynomials when considered as functions of the number of elements of the finite field. For up to five points these polynomials do not depend upon the characteristic. But for six-pointed hyperelliptic curves there is a dependence, which appears for the first time for genus 3.

By the Lefschetz trace formula, the $\mathbb{S}_n$-equivariant count of points of $\mathcal{H}_{g,n}$ is equivalent to the trace of Frobenius on the ℓ-adic $\mathbb{S}_n$-equivariant Euler characteristic of $\mathcal{H}_{g,n}$. But this information can also be formulated as traces of Frobenius on the Euler characteristic of some natural local systems $\mathbb{V}_\lambda$ on $\mathcal{H}_g$. By Theorem 3.2 in [1] we can use this connection to determine the Euler characteristic, evaluated in the Grothendieck group of absolute Galois modules, of all $\mathbb{V}_\lambda$ on $\mathcal{H}_g \otimes \overline{\mathbb{Q}}$ of weight at most 7. These result are in agreement with the results on the ordinary Euler characteristic and the conjectures on the motivic Euler characteristic of $\mathbb{V}_\lambda$ on $\mathcal{H}_3$ by Bini-van der Geer in [5], the ordinary Euler characteristic of $\mathbb{V}_\lambda$ on $\mathcal{H}_2$ by Getzler in [16], and the $\mathbb{S}_2$-equivariant cohomology of $\mathcal{H}_{g,2}$ for all $g \geq 2$ by Tommasi in [20].

The moduli stack $\overline{\mathcal{M}}_{g,n}$ of stable n-pointed curves of genus g is smooth and proper, which implies purity of the cohomology. If the $\mathbb{S}_n$-equivariant count of points of this space, when considered as a function of the number of elements of the finite field, gives a polynomial, then using the purity we can determine the $\mathbb{S}_n$-equivariant Galois (resp. Hodge) structure of its individual ℓ-adic (resp. Betti) cohomology groups (see Theorem 3.4 in [2] which is based on a result of van den Bogaart-Edixhoven in [6]). All curves of genus 2 are hyperelliptic and hence we can apply this theorem to $\overline{\mathcal{M}}_{2,n}$ for all $n \leq 7$. These results on genus 2 curves are all in agreement with the ones of Faber-van der Geer in [9] and [10]. Moreover, for $n \leq 3$ they were previously known by the work of Getzler in [14, Section 8].

Acknowledgements

The method I shall use to count points of the moduli space of pointed hyperelliptic curves follows a suggestion by Nicholas M. Katz. I thank Bradley Brock for letting me read an early version of the article [7], and Institut Mittag-Leffler for support during the preparation of this article. I would also like to thank my thesis advisor Carel Faber.

Outline

Let us give an outline of the paper, where $\star$. denotes the section.

- $\star_2$ In this section we define $\mathbb{S}_n$-equivariant counts of points of $\mathcal{H}_{g,n}$ over a finite field k, and we formulate the counts in terms of numbers $a_\lambda|_g$, which are connected to the H^1's of the hyperelliptic curves.
- $\star_3$ The hyperelliptic curves of genus g, in odd characteristic, are realized as degree 2 covers of $\mathbf{P}^1$ given by square-free polynomials of degree $2g+2$ or $2g+1$. The numbers $a_\lambda|_g$ are then expressed in terms of these polynomials in equation (3.2). The expression for $a_\lambda|_g$ is decomposed into parts denoted u_g, which are indexed by pairs of tuples of numbers $(\mathbf{n};\mathbf{r})$. The special cases of genus 0 and 1 are discussed in Section 3.1.
- $\star_4$ A recurrence relation is found for the numbers u_g (Theorem 4.12). The first step is to use the fact that any polynomial can be written uniquely as a monic square times a square-free one. This results in an equation which gives U_g in terms of u_h for h less than or equal to g, where U_g denotes the expression corresponding to u_g, but in terms of all polynomials instead of only the square-free ones. The second step is to use that, if g is large enough, U_g can be computed using a simple interpolation argument.
- $\star_5$ The recurrence relations for the u_g's are put together to form a linear recurrence relation for $a_\lambda|_g$, whose characteristic polynomial is given in Theorem 5.2.
- $\star_6$ It is shown how to compute u_0 for any pair $(\mathbf{n};\mathbf{r})$.
- $\star_7$ Information on the cases of genus 0 and 1 is used to compute, for all g, u_g for tuples $(\mathbf{n};\mathbf{r})$ of degree at most 5, and $a_\lambda|_g$ of weight at most 7.
- $\star_8$ The hyperelliptic curves are realized, in even characteristic, as pairs (h,f) of polynomials fulfilling three conditions. The numbers u_g and U_g are then defined to correspond to the case of odd characteristic.
- $\star_9$ In even characteristic, a recurrence relation is found for the numbers u_g (Theorem 9.11). Lemmas 9.6 and 9.7 show that one can do something in even characteristic corresponding to uniquely writing a polynomial as a monic square times a square-free one in odd characteristic. This results in a relation between U_g and u_h for h less than or equal to g. Then, as in odd characteristic, a simple interpolation argument is used to compute U_g for g large enough.
- $\star_{10}$ The same amount of information as in Section 7 is obtained in the case of even characteristic. It is noted that $a_\lambda|_g$ is independent of the characteristic for weight at most 5 (Theorem 10.3). This does not continue to hold for weight 6 where there is dependency for genus at least 3 (see Example 10.6).
- $\star_{11}$ The counts of points of the previous sections are used to get cohomological information. This is, in particular, applied to $\overline{\mathcal{M}}_{2,n}$ for $n \leq 7$.
- $\star_{12}$ In the first appendix, a more geometric interpretation is given of the information contained in all the numbers u_g of at most a certain degree (see Lemma 12.8).
- $\star_{13}$ In the second appendix, we find that for sufficiently large g we can compute the Euler characteristic, with $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-structure, of the part

of the cohomology of sufficiently high weight, of some local systems $\mathbb{V}_\lambda$ on $\mathcal{H}_g$. We will also see that these results are, in a sense, *stable* in g.

2. Equivariant counts

Let k be a finite field with q elements and denote by k_m a degree m extension. Define $H_{g,n}$ to be the coarse moduli space of $\mathcal{H}_{g,n} \otimes \bar{k}$ and let F be the geometric Frobenius morphism.

The purpose of this article is to make $\mathbb{S}_n$-equivariant counts of the number of points defined over k of $H_{g,n}$. With this we mean a count, for each element $\sigma \in \mathbb{S}_n$, of the number of fixed points of $F\sigma$ acting on $H_{g,n}$. Note that these numbers only depend upon the cycle type $c(\sigma)$ of the permutation σ.

Define $\mathcal{R}_\sigma$ to be the category of hyperelliptic curves of genus g that are defined over k together with marked points $(p_1, \dots, p_n)$ defined over $\bar{k}$ such that $(F\sigma)(p_i) = p_i$ for all i. Points of $H_{g,n}$ are isomorphism classes of n-pointed hyperelliptic curves of genus g defined over $\bar{k}$. For any pointed curve X that is a representative of a point in $H_{g,n}^{F\sigma}$, the set of fixed points of $F\sigma$ acting on $H_{g,n}$, there is an isomorphism from X to the pointed curve $(F\sigma)X$. Using this isomorphism we can descend to an element of $\mathcal{R}_\sigma$ (see [17, Lem. 10.7.5]). Therefore, the number of $\bar{k}$-isomorphism classes of the category $\mathcal{R}_\sigma$ is equal to $|H_{g,n}^{F\sigma}|$.

Fix an element $Y = (C, p_1, \dots, p_n)$ in $\mathcal{R}_\sigma$. We then have the following equality (see [12] or [17]):

$$\sum_{\substack{[X] \in \mathcal{R}_\sigma/\cong k \\ X \cong_{\bar{k}} Y}} \frac{1}{|\mathrm{Aut}_k(X)|} = 1.$$

This enables us to go from $\bar{k}$-isomorphism classes to k-isomorphism classes:

$$|H_{g,n}^{F\sigma}| = \sum_{[Y] \in \mathcal{R}_\sigma/\cong_{\bar{k}}} 1 = \sum_{[Y] \in \mathcal{R}_\sigma/\cong_{\bar{k}}} \sum_{\substack{[X] \in \mathcal{R}_\sigma/\cong k \\ X \cong_{\bar{k}} Y}} \frac{1}{|\mathrm{Aut}_k(X)|} = \sum_{[X] \in \mathcal{R}_\sigma/\cong_k} \frac{1}{|\mathrm{Aut}_k(X)|}.$$

For any curve C over k, define $C(\sigma)$ to be the set of n-tuples of distinct points $(p_1, \dots, p_n)$ in $C(\bar{k})$ that fulfill $(F\sigma)(p_i) = p_i$.

Notation 2.1. A partition λ of an integer m consists of a sequence of non-negative integers $\lambda_1, \dots, \lambda_\nu$ such that $|\lambda| := \sum_{i=1}^{\nu} i\lambda_i = m$. We will write $\lambda = [1^{\lambda_1}, \dots, \nu^{\lambda_\nu}]$.

Say that $\tau \in \mathbb{S}_n$ consists of one n-cycle. The elements of $C(\tau)$ are then given by the choice of $p_1 \in C(k_n)$ such that $p_1 \notin C(k_i)$ for every $i < n$. By an inclusion-exclusion argument it is then straightforward to show that

$$|C(\tau)| = \sum_{d|n} \mu(n/d)\, |C(k_d)|,$$

where μ is the Möbius function. Say that λ is any partition and that $\sigma \in \mathbb{S}_{|\lambda|}$ has the property $c(\sigma) = \lambda$. Since $C(\sigma)$ consists of tuples of distinct points it directly follows that

$$(2.1) \qquad |C(\sigma)| = \prod_{i=1}^{\nu} \prod_{j=0}^{\lambda_i - 1} \left(\sum_{d|i} \Big(\mu(i/d)\, |C(k_d)| - ji \Big) \right).$$

Fix a curve C over k and let $X_1, \dots, X_m$ be representatives of the distinct k-isomorphism classes of the subcategory of $\mathcal{R}_\sigma$ of elements $(D, q_1, \dots, q_n)$ where $D \cong_k C$. For each X_i we can act with $\mathrm{Aut}_k(C)$ which gives an orbit lying in $\mathcal{R}_\sigma$ and where the stabilizer of X_i is equal to $\mathrm{Aut}_k(X_i)$. Together the orbits of $X_1, \dots, X_m$ will contain $|C(\sigma)|$ elements and hence we obtain

$$(2.2) \qquad |H_{g,n}^{F\sigma}| = \sum_{[X] \in \mathcal{R}_\sigma / \cong_k} \frac{1}{|\mathrm{Aut}_k(X)|} = \sum_{[C] \in \mathcal{H}_g(k)/\cong_k} \frac{|C(\sigma)|}{|\mathrm{Aut}_k(C)|}.$$

We will compute slightly different numbers than $|H_{g,n}^{F\sigma}|$, but which contain equivalent information. Let C be a curve defined over k. The Lefschetz trace formula tells us that for all $m \geq 1$,

$$(2.3) \qquad |C(k_m)| = |C_{\bar{k}}^{F^m}| = 1 + q^m - a_m(C) \text{ where } a_m(C) = \mathrm{Tr}\big(F^m, H^1(C_{\bar{k}}, \mathbb{Q}_\ell)\big).$$

If we consider equations (2.1) and (2.2) in view of equation (2.3) we find that

$$|H_{g,n}^{F\sigma}| = \sum_{[C] \in \mathcal{H}_g(k)/\cong_k} \frac{1}{|\mathrm{Aut}_k(C)|} f_\sigma(q, a_1(C), \dots, a_n(C)),$$

where $f_\sigma(x_0, \dots, x_n)$ is a polynomial with coefficients in $\mathbb{Z}$. Give the variable x_i degree i. Then there is a unique monomial in f_σ of highest degree, namely $x_1^{\lambda_1} \cdots x_\nu^{\lambda_\nu}$. The numbers which we will pursue will be the following.

DEFINITION 2.2. For $g \geq 2$ and any partition λ define

$$(2.4) \qquad a_\lambda|_g := \sum_{[C] \in \mathcal{H}_g(k)/\cong_k} \frac{1}{|\mathrm{Aut}_k(C)|} \prod_{i=1}^{\nu} a_i(C)^{\lambda_i}.$$

This expression will be said to have *weight* $|\lambda|$. Let us also define

$$a_0|_g := \sum_{[C] \in \mathcal{H}_g(k)/\cong_k} \frac{1}{|\mathrm{Aut}_k(C)|},$$

an expression of weight 0.

3. REPRESENTATIVES OF HYPERELLIPTIC CURVES IN ODD CHARACTERISTIC

Assume that the finite field k has an odd number of elements. The hyperelliptic curves of genus $g \geq 2$ are the ones endowed with a degree 2 morphism to $\mathbf{P}^1$. This morphism induces a degree 2 extension of the function field of $\mathbf{P}^1$. If we consider hyperelliptic curves defined over the finite field k and choose an affine coordinate x on $\mathbf{P}^1$, then we can write this extension in the form $y^2 = f(x)$, where f is a square-free polynomial with coefficients in k of degree $2g+1$ or $2g+2$. At infinity, we can describe the curve given by the polynomial f in the coordinate $t = 1/x$ by $y^2 = t^{2g+2} f(1/t)$. We will therefore let $f(\infty)$, which corresponds to $t = 0$, be the coefficient of f of degree $2g + 2$.

DEFINITION 3.1. Let P_g denote the set of square-free polynomials with coefficients in k and of degree $2g+1$ or $2g+2$, and let $P'_g \subset P_g$ consist of the monic polynomials. Write C_f for the curve corresponding to the element f in P_g.

By construction, there exists for each k-isomorphism class of objects in $\mathcal{H}_g(k)$ an f in P_g such that C_f is a representative. Moreover, the k-isomorphisms between curves corresponding to elements of P_g are given by k-isomorphisms of their function fields. By the uniqueness of the linear system g^1_2 on a hyperelliptic curve, these isomorphisms must respect the inclusion of the function field of $\mathbf{P}^1$. The k-isomorphisms are therefore precisely (see [16, p. 126]) the ones induced by elements of the group $G := \mathrm{GL}_2^{\mathrm{op}}(k) \times k^*/D$ where

$$D := \{(\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, a^{g+1}) : a \in k^*\} \subset \mathrm{GL}_2^{\mathrm{op}}(k) \times k^*$$

and where an element

$$\gamma = [(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, e)] \in G$$

induces the isomorphism

$$(x, y) \mapsto \left(\frac{ax + b}{cx + d}, \frac{ey}{(cx + d)^{g+1}} \right).$$

This defines a left group action of G on P_g, where $\gamma \in G$ takes $f \in P_g$ to $\tilde{f} \in P_g$, with

$$\tilde{f}(x) = \frac{(cx + d)^{2g+2}}{e^2} f\Big(\frac{ax + b}{cx + d}\Big). \tag{3.1}$$

NOTATION 3.2. Let us put $I := 1/|G| = (q^3 - q)^{-1}(q - 1)^{-1}$.

DEFINITION 3.3. Let $\chi_{2,m}$ be the quadratic character on k_m. Recall that it is the function that takes $\alpha \in k_m$ to 1 if it is a square, to -1 if it is a nonsquare and to 0 if it is 0. With a square or a nonsquare we will always mean a nonzero element.

LEMMA 3.4. *If C_f is the hyperelliptic curve corresponding to $f \in P_g$ then*

$$a_m(C_f) = - \sum_{\alpha \in \mathbf{P}^1(k_m)} \chi_{2,m}\big(f(\alpha)\big).$$

Proof: The fiber of $C_f \to \mathbf{P}^1$ over $\alpha \in \mathbf{A}^1(k_m)$ will consist of two points defined over k_m if $f(\alpha)$ is a square in k_m, no point if $f(\alpha)$ is a nonsquare in k_m, and one point if $f(\alpha) = 0$. By the above description of f in terms of the coordinate $t = 1/x$, the same holds for $\alpha = \infty$. The lemma now follows from equation (2.3). □

We will now rephrase equation (2.4) in terms of the elements of P_g. By what was said above, the stabilizer of an element f in P_g under the action of G is equal to $\mathrm{Aut}_k(C_f)$ and hence

$$\begin{aligned}(3.2)\quad a_\lambda|_g &= \sum_{[f]\in P_g/G} \frac{1}{|\mathrm{Stab}_G(f)|} \prod_{i=1}^{\nu} a_i(C_f)^{\lambda_i} = \\ &= \frac{1}{|G|}\sum_{f\in P_g}\prod_{i=1}^{\nu} a_i(C_f)^{\lambda_i} = I \sum_{f\in P_g}\prod_{i=1}^{\nu}\Big(-\sum_{\alpha\in \mathbf{P}^1(k_i)} \chi_{2,i}\big(f(\alpha)\big)\Big)^{\lambda_i}.\end{aligned}$$

This can up to sign be rewritten as

$$(3.3)\qquad I\sum_{f\in P_g}\sum_{(\alpha_{1,1},\ldots,\alpha_{\nu,\lambda_\nu})\in S}\prod_{i=1}^{\nu}\prod_{j=1}^{\lambda_i}\chi_{2,i}\big(f(\alpha_{i,j})\big),$$

where $S := \prod_{i=1}^{\nu} \mathbf{P}^1(k_i)^{\lambda_i}$, in other words, $\alpha_{i,j}\in \mathbf{P}^1(k_i)$ for each $1\le i\le \nu$ and $1\le j\le\lambda_i$. The sum (3.3) will be split into parts for which we, in Section 4, will find recurrence relations in g.

Definition 3.5. For any tuple $\mathbf{n}=(n_1,\ldots,n_m)\in\mathbb{N}^m_{\ge 1}$, let the set $A(\mathbf{n})$ consist of the tuples $\alpha=(\alpha_1,\ldots,\alpha_m)\in\prod_{i=1}^m \mathbf{P}^1(k_{n_i})$ such that for any $1\le i,j\le m$ and any $s\ge 0$,

$$F^s(\alpha_i)=\alpha_j \implies n_i|s \text{ and } i=j.$$

Let us also define $A'(\mathbf{n}) := A(\mathbf{n})\cap\prod_{i=1}^m \mathbf{A}^1(k_{n_i})$.

Definition 3.6. Let $\mathcal{N}_m$ denote the set of pairs $(\mathbf{n};\mathbf{r})$ such that $\mathbf{n} = (n_1,\ldots,n_m)\in\mathbb{N}^m_{\ge 1}$ and $\mathbf{r}=(r_1,\ldots,r_m)\in\{1,2\}^m$.

Definition 3.7. For any $g\ge -1$, $(\mathbf{n};\mathbf{r})\in\mathcal{N}_m$ and $\alpha=(\alpha_1,\ldots,\alpha_m)\in A(\mathbf{n})$ define

$$u_{g,\alpha}^{(\mathbf{n};\mathbf{r})} := I\sum_{f\in P_g}\prod_{i=1}^{m}\chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i}$$

and

$$u_g^{(\mathbf{n};\mathbf{r})} := \sum_{\alpha\in A(\mathbf{n})} u_{g,\alpha}^{(\mathbf{n};\mathbf{r})}.$$

Construction-Lemma 3.8. *For each λ, there are positive integers $c_1,\ldots,c_s$ and $m_1,\ldots,m_s$, and moreover pairs $(\mathbf{n}^{(i)};\mathbf{r}^{(i)})\in\mathcal{N}_{m_i}$ for each $1\le i\le s$, such that for any finite field k,*

$$a_\lambda|_g=\sum_{i=1}^{s} c_i\, u_g^{(\mathbf{n}^{(i)};\mathbf{r}^{(i)})}.$$

Proof: The lemma will be proved by writing the set S as a disjoint union of parts that only depend upon the partition λ, and which therefore are independent of the chosen finite field k.

For each positive integer i, let $i=d_{i,1}>\ldots>d_{i,\delta_i}=1$ be the divisors of i.

- ⋆ For each $1 \leq i \leq \nu$, let $T_{i,1}, \ldots, T_{i,\delta_i}$ be an ordered partition of the set $\{1, \ldots, \lambda_i\}$ into (possibly empty) subsets.
- ⋆ For each $1 \leq i \leq \nu$ and each $1 \leq j \leq \delta_i$, let $Q_{i,j,1}, \ldots, Q_{i,j,\kappa_{i,j}}$ be an unordered partition (where $\kappa_{i,j}$ is arbitrary) of the set $T_{i,j}$ into non-empty subsets.

From such a choice of partitions we define a subset $S' = S(\{T_{i,j}\}, \{Q_{i,j,k}\})$ of S consisting of the tuples $(\alpha_{1,1}, \ldots, \alpha_{\nu,\lambda_\nu}) \in S$ fulfilling the following two properties.

- ⋆ If $x \in T_{i,j}$ then: $\alpha_{i,x} \in k_j$ and $\forall s < j, \alpha_{i,x} \notin k_s$.
- ⋆ If $x \in Q_{i,j,k}$ and $x' \in Q_{i',j',k'}$ then:

$$\exists s : F^s(\alpha_{i,x}) = \alpha_{i',x'} \iff (i,j,k) = (i',j',k').$$

Define $\mathbf{n}$ to be equal to the tuple

$$(\overbrace{d_{1,1}, \ldots, d_{1,1}}^{\kappa_{1,1}}, \overbrace{d_{1,2}, \ldots, d_{1,2}}^{\kappa_{1,2}}, \ldots, \overbrace{d_{1,\delta_1}, \ldots, d_{1,\delta_1}}^{\kappa_{1,\delta_1}}, \overbrace{d_{2,1}, \ldots, d_{2,1}}^{\kappa_{2,1}}, \ldots, \overbrace{d_{\nu,\delta_\nu}, \ldots, d_{\nu,\delta_\nu}}^{\kappa_{\nu,\delta_\nu}}).$$

Let $\rho_{i,j,k}$ be equal to 2 if either $i/d_{i,j}$ or $|Q_{i,j,k}|$ is even, and 1 otherwise. Define $\mathbf{r}$ to be equal to

$$(\rho_{1,1,1}, \rho_{1,1,2}, \ldots, \rho_{1,1,\kappa_{1,1}}, \rho_{1,2,1}, \ldots, \rho_{1,\delta_1,\kappa_{1,\delta_1}}, \rho_{2,1,1}, \ldots, \rho_{\nu,\delta_\nu,\kappa_{\nu,\delta_\nu}}).$$

The equality

$$u_g^{(\mathbf{n};\mathbf{r})} = I \sum_{f \in P_g} \sum_{(\alpha_{1,1},\ldots,\alpha_{\nu,\lambda_\nu}) \in S'} \prod_{i=1}^{\nu} \prod_{j=1}^{\lambda_i} \chi_{2,i}\big(f(\alpha_{i,j})\big)$$

is clear in view of the following three simple properties of the quadratic character.

- ⋆ Say that $\alpha \in \mathbf{P}^1(k_s)$, then if $\tilde{s}/s$ is even we have $\chi_{2,\tilde{s}}\big(f(\alpha)\big) = \chi_{2,s}\big(f(\alpha)\big)^2$ and if $\tilde{s}/s$ is odd we have $\chi_{2,\tilde{s}}\big(f(\alpha)\big) = \chi_{2,s}\big(f(\alpha)\big)$.
- ⋆ If for any $\alpha, \beta \in \mathbf{P}^1$ we have $F^s(\alpha) = \beta$ for some s, then $\chi_{2,i}\big(f(\alpha)\big) = \chi_{2,i}\big(f(\beta)\big)$ for all i.
- ⋆ Finally, for any $\alpha \in \mathbf{P}^1$ and any s, we have $\chi_{2,s}\big(f(\alpha)\big)^r = \chi_{2,s}\big(f(\alpha)\big)^2$ if r is even and $\chi_{2,s}\big(f(\alpha)\big)^r = \chi_{2,s}\big(f(\alpha)\big)$ if r is odd.

The lemma now follows directly from the fact that the sets $S(\{T_{i,j}\}, \{Q_{i,j,k}\}) \subset S$ (for different choices of partitions $\{T_{i,j}\}$ and $\{Q_{i,j,k}\}$) are disjoint and cover S. □

The set of data $\{(c_i, (\mathbf{n}^{(i)}; \mathbf{r}^{(i)}))\}$ resulting from the procedure given in the proof of Construction-Lemma 3.8 is, after assuming the pairs $(\mathbf{n}^{(i)}; \mathbf{r}^{(i)})$ to be distinct, unique up to simultaneous reordering of the elements of $\mathbf{n}^{(i)}$ and $\mathbf{r}^{(i)}$ for each i, and it will be called *the decomposition of* $a_\lambda|_g$.

Definition 3.9. For a partition λ, the pair

$$(\mathbf{n}; \mathbf{r}) = \big((\overbrace{1, \ldots, 1}^{\lambda_1}, \overbrace{2, \ldots, 2}^{\lambda_2}, \ldots, \overbrace{\nu, \ldots, \nu}^{\lambda_\nu}); (1, \ldots, 1)\big)$$

will appear in the decomposition of $a_\lambda|_g$ (corresponding to the partitions $T_{i,1} = \{1,\dots,\lambda_i\}$ for $1 \le i \le \nu$, and $Q_{i,1,k} = \{k\}$ for $1 \le i,k \le \nu$) with coefficient equal to 1, and it will be called the *general case*. All other pairs $(\mathbf{n};\mathbf{r})$ appearing in the decomposition of $a_\lambda|_g$ will be refered to as *degenerations* of the general case.

Definition 3.10. For any $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$, the number $|\mathbf{n}| := \sum_{i=1}^m n_i$ will be called the *degree* of $(\mathbf{n};\mathbf{r})$.

Lemma 3.11. *The general case is the only case in the decomposition of $a_\lambda|_g$ which has degree equal to the weight of $a_\lambda|_g$.*

Proof: If $(\mathbf{n};\mathbf{r})$ appears in the decomposition of $a_\lambda|_g$ and is associated to the partitions $\{T_{i,j}\}$ and $\{Q_{i,j,k}\}$, then $|\mathbf{n}| = \sum_{i=1}^{\nu}\sum_{j=1}^{\delta_i}\kappa_{i,j}d_{i,j}$. Since $\lambda_i = \sum_{j=1}^{\delta_i}\kappa_{i,j}$ and $1 \le d_{i,j} \le i$, the equality $|\lambda| = |\mathbf{n}|$ implies that $\kappa_{i,1} = \lambda_i$ and $\kappa_{i,j} = 0$ if $j \neq 1$. $\square$

Lemma 3.12. *If $(\mathbf{n};\mathbf{r})$ appears in the decomposition of $a_\lambda|_g$ then $\sum_{i=1}^m r_i n_i \le |\lambda|$ and these two numbers have the same parity.*

Proof: If $(\mathbf{n};\mathbf{r})$ appears in the decomposition of $a_\lambda|_g$ and is associated to the partitions $\{T_{i,j}\}$ and $\{Q_{i,j,k}\}$, then $\sum_{i=1}^m r_i n_i = \sum_{i=1}^{\nu}\sum_{j=1}^{\delta_i}\sum_{k=1}^{\kappa_{i,j}}\rho_{i,j,k}d_{i,j}$.
Let us prove the lemma by induction on m, starting with the case that $m = \sum_{i=1}^{\nu}\lambda_i$. In this case we must have $|Q_{i,j,k}| = 1$ for all $1 \le i \le \nu$, $1 \le j \le \delta_i$ and $1 \le k \le \kappa_{i,j}$, and hence $\rho_{i,j,k}$ is only equal to two if $i/d_{i,j}$ is even. This directly tells us that $\rho_{i,j,k}d_{i,j} \le i$, and that these two numbers have the same parity. Since $\lambda_i = \sum_{j=1}^{\delta_i}\kappa_{i,j}$, it follows that $\sum_{i=1}^m r_i n_i \le |\lambda|$ and that these two numbers have the same parity.
Assume now that $m = k$ and that the lemma has been proved for all pairs $(\tilde{\mathbf{n}};\tilde{\mathbf{r}})$ with $\tilde{m} > k$. Since $m < \sum_{i=1}^{\nu}\lambda_i$ we know that there exists numbers i_0, j_0, k_0 such that $|Q_{i_0,j_0,k_0}| \ge 2$. Let us fix an element $x \in Q_{i_0,j_0,k_0}$ and define a new pair $(\mathbf{n}';\mathbf{r}')$ associated to the partitions $\{T'_{i,j}\}$ and $\{Q'_{i,j,k}\}$ by putting:

- ⋆ $T'_{i,j} = T_{i,j}$ for all $1 \le i \le \nu$ and $1 \le j \le \delta_i$,
- ⋆ $Q'_{i_0,j_0,k_0} = Q_{i_0,j_0,k_0} \setminus \{x\}$,
- ⋆ $\kappa'_{i_0,j_0} = \kappa_{i_0,j_0} + 1$ and $Q'_{i_0,j_0,\kappa'_{i_0,j_0}} = \{x\}$,
- ⋆ $Q'_{i,j,k} = Q_{i,j,k}$ in all other cases.

The pair $(\mathbf{n}',\mathbf{r}')$ thus appears in the decomposition of λ, and $m' = k+1$. Moreover, we directly find that $\sum_{i=1}^m r_i n_i \le \sum_{i=1}^{m'} r'_i n'_i$ and that these two numbers have the same parity. By the induction hypothesis the lemma is then also true for $(\mathbf{n};\mathbf{r})$. $\square$

Example 3.13. Let us decompose $a_{[2^2]}|_g$ starting with the general case:

$$a_{[2^2]}|_g = I\sum_{f\in P_g}\Big(-\sum_{\alpha\in\mathbf{P}^1(k_2)}\chi_{2,2}\big(f(\alpha)\big)\Big)^2 = I\sum_{f\in P_g}\sum_{\alpha,\beta\in\mathbf{P}^1(k_2)}\chi_{2,2}\big(f(\alpha)f(\beta)\big) =$$
$$= u_g^{((2,2);(1,1))} + 2u_g^{((2,1);(1,2))} + 2u_g^{((2);(2))} + u_g^{((1,1);(2,2))} + u_g^{((1);(2))}.$$

EXAMPLE 3.14. The decomposition of $a_{[1^4,2]}|_g$, starting with the general case:

$$a_{[1^4,2]}|_g = -u_g^{((2,1,1,1,1);(1,1,1,1,1))} - 6u_g^{((2,1,1,1);(1,2,1,1))} - 3u_g^{((2,1,1);(1,2,2))} - 4u_g^{((2,1,1);(1,1,1))} - u_g^{((2,1);(1,2))} - u_g^{((1,1,1,1,1);(2,1,1,1,1))} - 6u_g^{((1,1,1,1);(2,2,1,1))} - 4u_g^{((1,1,1,1),(1,1,1,1))} - 3u_g^{((1,1,1);(2,2,2))} - 22u_g^{((1,1,1);(2,1,1))} - 7u_g^{((1,1);(2,2))} - 8u_g^{((1,1);(1,1))} - u_g^{((1);(2))}.$$

3.1. THE CASES OF GENUS 0 AND 1. We would like to have an equality of the same kind as in equation (3.2), but for curves of genus 0 and 1. Every curve of genus 0 or 1 has a morphism to $\mathbf{P}^1$ of degree 2 and in the same way as for larger genera, it then follows that every k-isomorphism class of curves of genus 0 or 1 has a representative among the curves coming from polynomials in P_0 and P_1 respectively. But there is a difference, compared to the larger genera, in that for curves of genus 0 or 1 the g_2^1 is not unique. In fact, the group G induces (in the same way as for $g \geq 2$) all k-isomorphisms between curves corresponding to elements of P_0 and P_1 that respect their given morphisms to $\mathbf{P}^1$ (i.e a fixed g_2^1), but not all k-isomorphisms between curves of genus 0 or 1 are of this form. Let us, for all $r \geq 0$, define the category $\mathcal{A}_r$ consisting of tuples $(C, Q_0, \ldots, Q_r)$ where C is a curve of genus 1 defined over k and the Q_i are, not necessarily distinct, points on C defined over k. The morphisms of $\mathcal{A}_r$ are, as expected, isomorphisms of the underlying curves that fix the marked points. Note that $\mathcal{A}_0$ is isomorphic to the category $\mathcal{M}_{1,1}(k)$. We also define, for all $r \geq 0$, the category $\mathcal{B}_r$ consisting of tuples $(C, L, Q_1, \ldots, Q_r)$ of the same kind as above, but where L is a g_2^1. A morphism of $\mathcal{B}_r$ is an isomorphism ϕ of the underlying curves that fixes the marked points, and such that there is an isomorphism τ making the following diagram commute:

$$\begin{tikzcd} C \arrow[r, "\phi"] \arrow[d, "L"'] & C' \arrow[d, "L'"] \\ \mathbf{P}^1 \arrow[r, "\tau"] & \mathbf{P}^1. \end{tikzcd}$$

Consider P_1 as a category where the morphisms are given by the elements of G. To every element of P_1 there corresponds, precisely as for $g \geq 2$, a curve C_f together with a g_2^1 given by the morphism to $\mathbf{P}^1$, thus an element of $\mathcal{B}_0$. Since every morphism in $\mathcal{B}_0$ between objects corresponding to elements of P_1 is induced by an element of G, and since for every k-isomorphism class of an element in $\mathcal{B}_0$ there is a representative in P_1, the two categories P_1 and $\mathcal{B}_0$ are equivalent.

For all $r \geq 1$ there are equivalences of the categories $\mathcal{A}_r$ and $\mathcal{B}_r$ given by

$$(C, Q_0, \ldots, Q_r) \mapsto (C, |Q_0 + Q_1|, Q_1, \ldots, Q_r),$$

with inverse

$$(C, L, Q_1, \ldots, Q_r) \mapsto (C, |L - Q_1|, Q_1, \ldots, Q_r).$$

We therefore have the equality

$$\sum_{[X]\in\mathcal{A}_r/\cong_k}\frac{1}{|\mathrm{Aut}_k(X)|}\prod_{i=1}^{\nu}a_i(C)^{\lambda_i}=\sum_{[Y]\in\mathcal{B}_r/\cong_k}\frac{1}{|\mathrm{Aut}_k(Y)|}\prod_{i=1}^{\nu}a_i(C)^{\lambda_i}.$$

The Riemann hypothesis tells us that $|a_r(C)|\leq 2g\sqrt{q^r}$, for any finite field k with q elements and for any curve C defined over k of genus g. For genus 1 this implies that $|C(k)|\geq q+1-2\sqrt{q}>0$, and thus every genus 1 curve has a point defined over k. There is therefore a number s such that $1\leq|C(k)|\leq s$ for all genus 1 curves C. As in the argument preceding equation (2.2) we can take a representative $(C,Q_0,\ldots,Q_r)$ for each element of $\mathcal{A}_r/\cong_k$ and act with $\mathrm{Aut}_k(C,Q_0)$, respectively for each representative $(C,L,Q_1,\ldots,Q_r)$ of $\mathcal{B}_0/\cong_k$ act with $\mathrm{Aut}_k(C,L)$, and by considering the orbits and stabilizers we get

$$\sum_{j=1}^{s}j^r\sum_{\substack{[X]\in\mathcal{A}_0/\cong_k\\|C(k)|=j}}\frac{1}{|\mathrm{Aut}_k(X)|}\prod_{i=1}^{\nu}a_i(C)^{\lambda_i}=\sum_{j=1}^{s}j^r\sum_{\substack{[Y]\in\mathcal{B}_0/\cong_k\\|C(k)|=j}}\frac{1}{|\mathrm{Aut}_k(Y)|}\prod_{i=1}^{\nu}a_i(C)^{\lambda_i}.$$

Since this holds for all $r\geq 1$ we can, by a Vandermonde argument, conclude that we have an equality as above for each fixed j. We can therefore extend Definition 2.2 to genus 1 in the following way:

$$(3.4)\quad a_\lambda|_1:=\sum_{\substack{[(C,Q_0)]\in\\\mathcal{M}_{1,1}(k)/\cong_k}}\frac{1}{|\mathrm{Aut}_k(C,Q_0)|}\prod_{i=1}^{\nu}a_i(C)^{\lambda_i}=$$

$$=\sum_{[f]\in P_1/G}\frac{1}{|\mathrm{Stab}_G(f)|}\prod_{i=1}^{\nu}a_i(C_f)^{\lambda_i}=I\sum_{f\in P_1}\prod_{i=1}^{\nu}a_i(C_f)^{\lambda_i},$$

which gives an agreement with equation (3.2).
All curves of genus 0 are isomorphic to $\mathbf{P}^1$ and $a_r(\mathbf{P}^1)=0$ for all $r\geq 1$. In this trivial case we just let equation (3.2) be the definition of $a_\lambda|_0$.

4. Recurrence relations for u_g in odd characteristic

This section will be devoted to finding, for a fixed finite field k with an odd number of elements and for a fixed pair $(\mathbf{n};\mathbf{r})\in\mathcal{N}_m$, a recurrence relation for u_g. Notice that we will often suppress the pair $(\mathbf{n};\mathbf{r})$ in our notation and for instance write u_g instead of $u_g^{(\mathbf{n};\mathbf{r})}$.
Fix a nonsquare t in k and an $\alpha=(\alpha_1,\ldots,\alpha_m)\in A(\mathbf{n})$. Multiplying with the element t gives a fixed point free action on the set P_g and therefore

$$(4.1)\quad u_{g,\alpha}=I\sum_{f\in P_g}\prod_{i=1}^{m}\chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i}=I\sum_{f\in P_g}\prod_{i=1}^{m}\chi_{2,n_i}\big(t\,f(\alpha_i)\big)^{r_i}=$$

$$=I\sum_{f\in P_g}\prod_{i=1}^{m}\chi_{2,n_i}(t)^{r_i}\,\chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i}=(-1)^{\sum_{i=1}^{m}r_in_i}\,u_{g,\alpha}.$$

This computation and Lemmas 3.8 and 3.12 proves the following lemma.

LEMMA 4.1. *For any $g \geq -1$, $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ and $\alpha \in A(\mathbf{n})$, if $\sum_{i=1}^m r_i n_i$ is odd then $u_{g,\alpha} = 0$. Consequently, $a_\lambda|_g$ is equal to 0 if it has odd weight.*

Thus, the only interesting cases are those for which $\sum_{i=1}^m r_i n_i$ is even.

REMARK 4.2. The last statement of Lemma 4.1 can also be found as a consequence of the existence of the hyperelliptic involution.

We also see from equation (4.1) that

$$u_{g,\alpha} = I\,(q-1) \sum_{f \in P'_g} \prod_{i=1}^m \chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i} \quad \text{if } \sum_{i=1}^m r_i n_i \text{ is even.} \tag{4.2}$$

DEFINITION 4.3. Let Q_g denote the set of all polynomials (that is, not necessarily square-free) with coefficients in k and of degree $2g+1$ or $2g+2$, and let $Q'_g \subset Q_g$ consist of the monic polynomials. For a polynomial $h \in Q_g$ we let $h(\infty)$ be the coefficient of the term of degree $2g+2$ (which extends the earlier definition for elements in P_g). For any $g \geq -1$, $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ and $\alpha \in A(\mathbf{n})$, define

$$U_{g,\alpha}^{(\mathbf{n};\mathbf{r})} := I \sum_{h \in Q_g} \prod_{i=1}^m \chi_{2,n_i}\big(h(\alpha_i)\big)^{r_i},$$

$$U_g^{(\mathbf{n};\mathbf{r})} := \sum_{\alpha \in A(\mathbf{n})} U_{g,\alpha}^{(\mathbf{n};\mathbf{r})} \quad \text{and} \quad \hat{U}_g^{(\mathbf{n};\mathbf{r})} := \sum_{i=-1}^{g} U_i^{(\mathbf{n};\mathbf{r})}.$$

We will find an equation relating U_g to u_i for all $-1 \leq i \leq g$. Moreover, for g large enough we will be able to compute U_g. Together, this will give us our recurrence relation for u_g.
With the same arguments as was used to prove equation (4.2) one shows that

$$U_{g,\alpha} = I\,(q-1) \sum_{h \in Q'_g} \prod_{i=1}^m \chi_{2,n_i}\big(h(\alpha_i)\big)^{r_i} \quad \text{if } \sum_{i=1}^m r_i n_i \text{ is even.} \tag{4.3}$$

DEFINITION 4.4. For any $\alpha = (\alpha_1, \ldots, \alpha_m) \in A'(\mathbf{n})$, let $b_j = b_j^{\mathbf{n}}$ be the number of monic polynomials l of degree j such that $l(\alpha_i)$ is nonzero for all i. Let us also put $\hat{b}_j = \hat{b}_j^{\mathbf{n}} := \sum_{i=0}^j b_i^{\mathbf{n}}$.

LEMMA 4.5. *For each $j \geq 0$ and $\mathbf{n} \in \mathbb{N}_{\geq 1}^m$, we have the equality*

$$b_j = q^j + \sum_{i=1}^{j} (-1)^i \sum_{\substack{1 \leq m_1 < \ldots < m_i \leq m \\ \sum_{l=1}^i n_{m_l} \leq j}} q^{j - \sum_{l=1}^i n_{m_l}} \tag{4.4}$$

from which it follows that b_j does not depend upon the choice of $\alpha \in A'(\mathbf{n})$.

Proof: The numbers b_j can be computed by inclusion-exclusion, where the choice of $1 \leq m_1 < \ldots < m_i \leq m$ corresponds to demanding the polynomial to be 0 in the points $\alpha_{m_1}, \ldots, \alpha_{m_i}$. □

Notation 4.6. For any $\alpha \in A'(\mathbf{n})$, let p_{α_i} denote the minimal polynomial of α_i and put $p_\alpha := \prod_{i=1}^m p_{\alpha_i}$.

Lemma 4.7. *For any $\alpha \in A'(\mathbf{n})$ there is a one-to-one correspondence between polynomials f defined over k with $\deg(f) \leq |\mathbf{n}| - 1$, and tuples $(f(\alpha_1), \ldots, f(\alpha_m)) \in \prod_{i=1}^m k_{n_i}$.*

Proof: For any $\alpha \in A'(\mathbf{n})$ we have $\deg(p_{\alpha_i}) = n_i$ and $\gcd(p_{\alpha_i}, p_{\alpha_j}) = 1$ if $i \neq j$. The lemma now follows from the Chinese remainder theorem, which tells us that the morphism $k[x]/p_\alpha \to \prod_{i=1}^m k[x]/p_{\alpha_i} \cong \prod_{i=1}^m k_{n_i}$ given by $f(x) \mapsto (f(\alpha_1), \ldots, f(\alpha_m))$ is an isomorphism. $\square$

Notation 4.8. Let R_j denote the set of polynomials of degree j and let R'_j be the subset containing the monic polynomials.

We will divide into two cases.

4.1. The case $\alpha \in A'(\mathbf{n})$. Fix an element $\alpha \in A'(\mathbf{n})$. Any nonzero polynomial h can be written uniquely in the form $h = f\,l^2$ where f is a square-free polynomial and l is a monic polynomial. This statement translates directly into the equality

$$U_{s,\alpha} = I \sum_{j+k=s} \sum_{l \in R'_j} \sum_{f \in P_k} \prod_{i=1}^m \chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i} \chi_{2,n_i}\big(l(\alpha_i)\big)^{2r_i} = \sum_{j=0}^{s+1} b_j u_{s-j,\alpha},$$

because for any $\beta \in \mathbf{A}^1(k_s)$, $\chi_{2,s}\big((fl^2)(\beta)\big) = \chi_{2,s}\big(f(\beta)\big)$ if $l(\beta) \neq 0$. Summing this equality over all s between -1 and g gives

$$\hat{U}_{g,\alpha} = \sum_{j=0}^{g+1} \hat{b}_j u_{g-j,\alpha}. \tag{4.5}$$

If $r_i = 2$ for all i, then it follows from equation (4.3) that

$$U_{s,\alpha} = I\,(q-1) \sum_{h \in Q'_s} \prod_{i=1}^m \chi_{2,n_i}\big(h(\alpha_i)\big)^2 = I\,(q-1)(b_{2s+2} + b_{2s+1}).$$

Summing this equality over all s between -1 and g gives

$$\hat{U}_{g,\alpha} = I\,(q-1)\hat{b}_{2g+2} \quad \text{for } g \geq -1 \text{ if } \forall i : r_i = 2. \tag{4.6}$$

In $\hat{U}_{g,\alpha}$ we are summing over all polynomials h of degree less than or equal to $2g+2$, and every h can uniquely be written on the form $h_1 + p_\alpha h_2$, with $\deg h_1 \leq |\mathbf{n}| - 1$ and $\deg h_2 \leq 2g + 2 - |\mathbf{n}|$. Hence if $2g + 2 \geq |\mathbf{n}| - 1$ we find that

$$\hat{U}_{g,\alpha} = I\,q^{2g+3-|\mathbf{n}|} \sum_{s=1}^{|\mathbf{n}|-1} \sum_{h_1 \in R_s} \prod_{i=1}^m \chi_{2,n_i}\big(h_1(\alpha_i)\big)^{r_i}.$$

Using Lemma 4.7 we can reformulate this equality as

$$\hat{U}_{g,\alpha} = I\,q^{2g+3-|\mathbf{n}|} \sum_{(\beta_1,\ldots,\beta_m) \in \prod_{i=1}^m k_{n_i}} \prod_{i=1}^m \chi_{2,n_i}(\beta_i)^{r_i}.$$

For any j, half of the nonzero elements in k_j are squares and half are nonsquares, and thus if $r_i = 1$ for some i, we can conclude from this equality that

$$\hat{U}_{g,\alpha} = 0 \quad \text{for } g \geq (|\mathbf{n}| - 3)/2 \text{ if } \exists i : r_i = 1. \tag{4.7}$$

4.2. THE CASE $\alpha \in A(\mathbf{n}) \setminus A'(\mathbf{n})$. Fix an element $\alpha \in A(\mathbf{n}) \setminus A'(\mathbf{n})$. We can assume that $\alpha_1 = \infty$, and then $\tilde{\alpha} := (\alpha_2, \ldots, \alpha_m) \in A'(\tilde{\mathbf{n}})$ where $\tilde{\mathbf{n}} := (n_2, \ldots, n_m)$.
If $h \in Q_g$ and $f \in P_j$ such that $h = f\, l^2$ for some monic polynomial l (which is then unique), then $h(\infty) = f(\infty)$, because the coefficient of h of degree $2g+2$ must equal the coefficient of f of degree $2j+2$. As in Section 4.1 we get
(4.8)

$$U_{g,\alpha} = I \sum_{j+k=g} \sum_{l \in R'_j} \sum_{f \in P_k} f(\infty) \prod_{i=2}^{m} \chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i} \chi_{2,n_i}\big(l(\alpha_i)\big)^{2r_i} = \sum_{j=0}^{g+1} b_j^{\tilde{\mathbf{n}}} u_{g-j,\alpha}.$$

If $\sum_{i=1}^m r_i n_i$ is even, equation (4.3) and the definition of $h(\infty)$ shows that

$$U_{g,\alpha} = I\,(q-1) \sum_{h \in R'_{2g+2}} \prod_{i=2}^{m} \chi_{2,n_i}\big(h(\alpha_i)\big)^{r_i}. \tag{4.9}$$

If $r_i = 2$ for all i, then equation (4.9) tells us that

$$U_{g,\alpha} = I\,(q-1) b_{2g+2}^{\tilde{\mathbf{n}}} \quad \text{for } g \geq -1, \forall i : r_i = 2. \tag{4.10}$$

If $2g+2 \geq |\mathbf{n}| - 1$, an element $h \in R'_{2g+2}$ can be written uniquely as $h = h_1 + p_{\tilde{\alpha}} h_2$, where $\deg(h_1) \leq |\mathbf{n}| - 2$, $\deg(h_2) \geq 0$ and h_2 monic. In the same way as in Section 4.1 we can (if $\sum_{i=1}^m r_i n_i$ is even) use this together with equation (4.9) and Lemma 4.7 to conclude that

$$U_{g,\alpha} = 0 \quad \text{for } g \geq (|\mathbf{n}| - 3)/2, \exists i : r_i = 1, \tag{4.11}$$

which of course also holds if $\sum_{i=1}^m r_i n_i$ is odd by Lemma 4.1 and equation (4.8).

REMARK 4.9. Fix an $\alpha \in A(\mathbf{n})$. If there is an element $\beta \in \mathbf{A}^1(k)$ such that $\beta \notin \{\alpha_1, \ldots, \alpha_n\}$, then $T(\alpha) := (T(\alpha_1), \ldots, T(\alpha_n))$ is in $A'(\mathbf{n})$, where T is the projective transformation of $\mathbf{P}_k^1$ defined by $x \mapsto \beta x/(x - \beta)$.
In the notation of equation (3.1), $\chi_{2,n_i}\big(f(T(\alpha_i))\big) = \chi_{2,n_i}\big(\tilde{f}(\alpha_i)\big)$ (with $e = 1$). Since this induces a permutation of P_g, we find that $u_{g,\alpha} = u_{g,T(\alpha)}$ and similarly that $U_{g,\alpha} = U_{g,T(\alpha)}$. So, if $q \geq |\mathbf{n}|$, then equations (4.5), (4.6) and (4.7) will also hold for $\alpha \in A(\mathbf{n}) \setminus A'(\mathbf{n})$. By Lemma 4.10 in the next section, we will see that this is true even if $q < |\mathbf{n}|$.

4.3. THE TWO CASES JOINED. In this section we will put the results of the two previous sections together using the following lemma.

LEMMA 4.10. *For any* $\tilde{\mathbf{n}} = (n_2, \ldots, n_m)$, *if* $\mathbf{n} = (1, n_2, \ldots, n_m)$ *then* $\hat{b}_j^{\mathbf{n}} = b_j^{\tilde{\mathbf{n}}}$.

Proof: Fix any tuple $\mathbf{n} = (n_1, \ldots, n_m)$ and put $n := |\mathbf{n}|$. If we let $t_i = q^{n_i}$ in the formula

$$\prod_{i=1}^{m}(t_i - 1) = t_1 \cdots t_m + \sum_{i=1}^{m}(-1)^i \sum_{1 \le m_1 < \ldots < m_i \le m} t_1 \cdots t_m \frac{1}{t_{m_1}} \cdots \frac{1}{t_{m_i}},$$

then the right hand side is equal to the right hand side of equation (4.4), and hence

$$\prod_{i=1}^{m}(q^{n_i} - 1) = b_n^{\mathbf{n}}. \tag{4.12}$$

Say that $b_j^{\mathbf{n}} = \sum_{i=0}^{j} c_{j,i}^{\mathbf{n}} q^i$ and $\hat{b}_j^{\mathbf{n}} = \sum_{i=0}^{j} \hat{c}_{j,i}^{\mathbf{n}} q^i$. If $i \le j$ then equation (4.4) implies that $c_{j,i}^{\mathbf{n}} = c_{n,n+i-j}^{\mathbf{n}}$ and hence $\hat{c}_{j,i}^{\mathbf{n}} = \sum_{s=0}^{j} c_{n,n+i-s}^{\mathbf{n}}$. By equation (4.12) we know that $q-1$ divides $b_n^{\mathbf{n}}$, and if $b_n^{\mathbf{n}}/(q-1) = \sum_{i=0}^{n-1} d_i q^i$ then $\hat{c}_{j,i}^{\mathbf{n}} = d_{n-1+i-j}$.
So, if $n_1 = 1$ and $\tilde{\mathbf{n}} = (n_2, \ldots, n_m)$ then $b_n^{\mathbf{n}}/(q-1) = b_{n-1}^{\tilde{\mathbf{n}}}$ and thus $\hat{c}_{j,i}^{\mathbf{n}} = c_{n-1,n-1+i-j}^{\tilde{\mathbf{n}}} = c_{j,i}^{\tilde{\mathbf{n}}}$. □

Notation 4.11. Let us write $J := I\,(q-1)\,|A(\mathbf{n})|$.

Theorem 4.12. *For any pair* $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$,

$$\sum_{j=0}^{g+1} \hat{b}_j u_{g-j} = \begin{cases} J\,\hat{b}_{2g+2} & \text{if } \forall i : r_i = 2,\ g \ge -1; \\ 0 & \text{if } \exists i : r_i = 1,\ g \ge \frac{|\mathbf{n}|-3}{2}. \end{cases}$$

Proof: The theorem follows from combining equations (4.5), (4.6), (4.7) and equations (4.8), (4.10), (4.11), using Lemma 4.10. □

Note that with this theorem we can, for any $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ such that $r_i = 2$ for all i, compute u_g for any g. Moreover, for any pair $(\mathbf{n};\mathbf{r})$ we can compute u_g for any g, if we already know u_g for all $g < (|\mathbf{n}|-3)/2$.

Lemma 4.13. *For any* $\mathbf{n}$, $q-1$ *divides* $b_{|\mathbf{n}|}^{\mathbf{n}}$, *and if we write* $b_{|\mathbf{n}|}^{\mathbf{n}}/(q-1) = \sum_{i=0}^{|\mathbf{n}|-1} d_i q^i$ *then* $\hat{b}_j - q\hat{b}_{j-1} = d_{|\mathbf{n}|-1-j}$.

Proof: The first claim is shown in the proof of Lemma 4.10. Using the notation of that proof we find that $\hat{b}_j - q\hat{b}_{j-1} = \sum_{i=0}^{j} d_{n-1+i-j} q^i - \sum_{i=0}^{j-1} d_{n+i-j} q^{i+1} = d_{n-1-j}$. Note that d_{n-1-j} only depends upon $\mathbf{n}$ and not on q. □

Theorem 4.14. *For any pair* $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$,

$$\sum_{j=0}^{\min(|\mathbf{n}|-1,g+1)} (\hat{b}_j - q\hat{b}_{j-1}) u_{g-j} = \begin{cases} J\,(\hat{b}_{2g+2} - q\hat{b}_{2g}) & \text{if } \forall i : r_i = 2,\ g \ge 0; \\ 0 & \text{if } \exists i : r_i = 1,\ g \ge \frac{|\mathbf{n}|-1}{2}. \end{cases}$$

Proof: Let us temporarily put $F(s) := \sum_{j=0}^{s+1} \hat{b}_j u_{s-j}$. From Lemma 4.13 we find that $\hat{b}_j - q\hat{b}_{j-1} = 0$ if $j > |\mathbf{n}| - 1$. The theorem then follows from applying Theorem 4.12 to the expression $F(g) - qF(g-1)$. □

For $g \geq (|\mathbf{n}|-1)/2$, Theorem 4.14 presents us with a linear recurrence relation for u_g which has coefficients that are independent of the finite field k.

EXAMPLE 4.15. If $(\mathbf{n};\mathbf{r}) = ((2,1,1,1);(1,2,1,1))$ then $b_5^{\mathbf{n}}/(q-1) = (q^2-1)(q-1)^2 = q^4 - 2q^3 + 2q - 1$. Applying Lemma 4.13 and then Theorem 4.14 we get

$$u_g - 2u_{g-1} + 2u_{g-3} - u_{g-4} = 0 \quad \text{for } g \geq 3.$$

EXAMPLE 4.16. Let us compute u_g, for all $g \geq -1$, when $(\mathbf{n};\mathbf{r}) = ((1,1,1),(2,2,2))$. We have that $u_{-1} = J = 1$ and since $r_i = 2$ for all i, Theorem 4.14 gives the equality $u_0 = 2u_{-1} + J(q^2 - 3q + 1) = q^2 - 3q + 3$. Applying Theorem 4.14 again we get

$$u_g - 2u_{g-1} + u_{g-2} = q^{2g-1}(q-1)^3 \quad \text{for } g \geq 1.$$

Solving this recurrence relation gives

$$u_g^{((1,1,1);(2,2,2))} = \frac{q^{2g+3}(q-1) - (2g+2)(q^2-1) + 3q + 1}{(q+1)^2} \quad \text{for } g \geq -1.$$

5. LINEAR RECURRENCE RELATIONS FOR $a_\lambda|_g$

REMARK 5.1. From a sequence v_n that fulfills a linear recurrence relation with characteristic polynomial C we can, for any polynomial D, in the obvious way construct a linear recurrence relation for v_n with characteristic polynomial CD. Thus, from two sequences v_n and w_n that each fulfill linear recurence relation with characteristic polynomial C and D respectively, we can construct a linear recurence relation for the sequence $v_n + w_n$ with characteristic polynomial $\mathrm{lcm}(C, D)$.

THEOREM 5.2. *By applying Theorem 4.14 to each pair $(\mathbf{n};\mathbf{r})$ appearing in the decomposition (given by Lemma 3.8) of $a_\lambda|_g$, we get a linear recurrence relation for $a_\lambda|_g$. The characteristic polynomial $C(X)$ of this linear recurrence relation equals*

$$\frac{1}{X-1}\prod_{i=1}^{\nu}(X^i-1)^{\lambda_i}. \tag{5.1}$$

Proof: Fix any pair $(\mathbf{n};\mathbf{r})$ in the decomposition of $a_\lambda|_g$ and put $n = |\mathbf{n}|$. Lemma 4.13 tells us that $\hat{b}_j - q\hat{b}_{j-1}$ is equal to the coefficient of q^{n-1-j} in $b_n/(q-1)$. If $g \geq n-1$, then these numbers are also the coefficients in the recurrence relation given by Theorem 4.14. By equation (4.12), the characteristic polynomial $C_{(\mathbf{n};\mathbf{r})}$ of this linear recurrence relation is equal to $(\prod_{i=1}^m (X^{n_i}-1))/(X-1)$.
We find that the linear recurrence relation in the general case (see Definition 3.9) will have characteristic polynomial equal to C. Moreover, we find (by their construction in the proof of Lemma 3.8) that if $(\mathbf{n};\mathbf{r})$ is a degenerate case then $C_{(\mathbf{n};\mathbf{r})}|C$. The theorem now follows from Remark 5.1. ☐

Theorem 5.2 tells us that if we can compute $a_\lambda|_g$ for $g < |\lambda| - 1$ then we can compute it for every g. But note that by considering the individual cases in the decomposition of $a_\lambda|_g$ we will do much better in Section 7, in the sense that we will be able to use information from curves of only genus 0 and 1 to compute $a_\lambda|_g$ for any λ such that $|\lambda| \le 6$.

Example 5.3. For $\lambda = [1^4, 2]$ the characteristic polynomial equals $(X-1)^4(X+1)$, so if V_g is a particular solution to the linear recurrence relation for $a_{[1^4,2]}|_g$ then

$$a_{[1^4,2]}|_g = V_g + A_3 g^3 + A_2 g^2 + A_1 g + A_0 + B_0(-1)^g,$$

where A_0, A_1, A_2, A_3 and B_0 do not depend upon g.

6. Computing u_0

In this section we will see that we can compute u_0 for any choice of a pair $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$. This is due to the fact that if C is a curve of genus 0 then, for all r, $|C(k_r)| = 1 + q^r$ or equivalently $a_r(C) = 0$.

Construction-Lemma 6.1. *For each* $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$, *there are numbers* $c_1, \dots, c_s$ *and pairs* $(\mathbf{n}^{(1)};\mathbf{r}^{(1)}), \dots, (\mathbf{n}^{(s)};\mathbf{r}^{(s)})$, *where* $\mathbf{r}^{(i)} = (2, \dots, 2)$ *for all* i, *such that for any finite field* k,

$$u_0^{(\mathbf{n};\mathbf{r})} = \sum_{i=1}^{s} c_i u_0^{(\mathbf{n}^{(i)};\mathbf{r}^{(i)})}.$$

Proof: Fix a pair $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$. We will use induction over the number $n := |\mathbf{n}|$, where the base case $n = 0$ is trivial.
Let us put $(\tilde{\mathbf{n}};\tilde{\mathbf{r}}) = ((n_2, \dots, n_m);(r_2, \dots, r_m))$. For an $\tilde{\alpha} = (\alpha_2, \dots, \alpha_m) \in A(\tilde{\mathbf{n}})$ let $\hat{\mathbf{P}}^1_{\tilde{\alpha}}(k_i)$ be the set of all points in $\mathbf{P}^1(k_i) \setminus \{\alpha_2, \dots, \alpha_m\}$ that are not defined over a proper subfield of k_i. The set of $\alpha_1 \in \mathbf{P}^1(k_{n_1})$ such that $(\alpha_1, \dots, \alpha_m) \in A(\mathbf{n})$ then equals

$$\mathbf{P}^1(k_{n_1}) \setminus \Big(\bigcup_{i|n_1} \hat{\mathbf{P}}^1_{\tilde{\alpha}}(k_i) \bigcup_{n_i|n_1} \{\alpha_i, \dots, F^{n_i-1}\alpha_i\} \Big). \tag{6.1}$$

Assume now that the lemma has been proved for all pairs of degree strictly less than n. By reordering the elements of the pair $(\mathbf{n};\mathbf{r})$ we can assume that $r_1 = 1$, because otherwise $\mathbf{r} = (2, \dots, 2)$ and we are done. By applying equation (6.1) we get

$$\begin{aligned} I \sum_{\alpha \in A(\mathbf{n})} \prod_{i=1}^{m} \chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i} &= I \sum_{\tilde{\alpha} \in A(\tilde{\mathbf{n}})} \prod_{i=2}^{m} \chi_{2,n_i}\big(f(\alpha_i)\big)^{r_i} \cdot \\ &\cdot \Big(-a_{n_1}(C_f) - \sum_{i|n_1} \sum_{\beta \in \hat{\mathbf{P}}^1_{\tilde{\alpha}}(k_i)} \chi_{2,n_1}\big(f(\beta)\big) - \sum_{n_i|n_1} n_i\, \chi_{2,n_1}\big(f(\alpha_i)\big) \Big). \end{aligned} \tag{6.2}$$

Let us put $(\mathbf{n}^{(i)};\mathbf{r}^{(i)}) = ((i, n_2, \dots, n_m);(n_i/i, r_2, \dots, r_m))$ for all i that divides n_i and $\tilde{\mathbf{r}}^{(i)} = (r_2, \dots, r_{i-1}, r_i\, n_1/n_i, r_{i+1}, \dots, r_m)$ for all n_i that divides n_1.

Summing both sides of equation (6.2) over polynomials $f \in P_0$ and using that $a_{n_1}(C_f) = 0$ then gives

$$u_0^{(\mathbf{n};\mathbf{r})} = -\sum_{i|n_1} u_0^{(\mathbf{n}^{(i)};\mathbf{r}^{(i)})} - \sum_{n_i|n_1} n_i u_0^{(\tilde{\mathbf{n}};\tilde{\mathbf{r}}^{(i)})}. \tag{6.3}$$

Since $|\tilde{\mathbf{n}}| < n$ and $|\mathbf{n}^{(i)}| < n$, the lemma follows by induction from equation (6.3). $\square$

EXAMPLE 6.2. In the case $(\mathbf{n};\mathbf{r}) = ((6,6,3,1,1);(1,1,2,2,2))$, the first step in the procedure in the proof of Lemma 6.1 equals

$$u_0^{(\mathbf{n};\mathbf{r})} = -u_0^{((6,3,3,1,1);(1,2,2,2,2))} - u_0^{((6,3,2,1,1);(1,2,1,2,2))} - u_0^{((6,3,1,1,1);(1,2,2,2,2))} \\ - 5u_0^{((6,3,1,1);(1,2,2,2))} - 6u_0^{((6,3,1,1);(2,2,2,2))}.$$

EXAMPLE 6.3. In the case $(\mathbf{n};\mathbf{r}) = ((4,1,1,1);(1,2,1,1))$, the procedure in the proof of Lemma 6.1 gives

$$u_0^{(\mathbf{n};\mathbf{r})} = u_0^{((2,1,1);(2,2,2))} + u_0^{((1,1,1);(2,2,2))} + u_0^{((1,1);(2,2))} - u_0^{((2,1);(2,2))} - u_0^{((1);(2))}$$

7. RESULTS FOR WEIGHT UP TO 7 IN ODD CHARACTERISTIC

We will in this section show that we, for any number g and any finite field k of odd characteristic, can compute all $a_\lambda|_g$ of weight at most 7. This is achieved by decomposing $a_\lambda|_g$ using Lemma 3.8 and employing the recurrence relation of Theorem 4.12 on the different parts. This involves finding the necessary base cases for the recurrence relations and that will be possible with the help of results on genus 0 curves obtained in Section 6, and on genus 1 curves obtained in the article [1].
We will write $a_\lambda|_{g,odd}$ and $u_{g,odd}$ to stress that all results are in the case of odd characteristic. See Section 10 for results in the case of even characteristic.

EXAMPLE 7.1. Theorem 4.12 is applicable even if the degree is 0 (if considered as a case when $r_i = 2$ for all i) and with $\hat{b}_j = \sum_{i=0}^{j} q^i$. From Theorem 4.12 we find that $a_0|_{0,odd} = Jq^2 = q/(q^2-1)$ and again from Theorem 4.12 that

$$a_0|_{g,odd} = J(q^{2g+2} - q^{2g}) = q^{2g-1} \quad \text{for } g \geq 1.$$

This result can also be found in [7, Proposition 7.1].

7.1. DEGREE AT MOST 3.

When the degree of the pair $(\mathbf{n};\mathbf{r})$ is at most 3 we find using Theorem 4.12 that we do not need any base cases to compute u_g for every g.

EXAMPLE 7.2. Let us consider $(\mathbf{n};\mathbf{r}) = ((2);(1))$. We have $u_{-1} = J = 1/(q+1)$ and applying Theorem 4.12 we get $u_0 = -(q+1)u_{-1} = -1$. Theorem 4.14 tells us that $u_g = -u_{g-1}$ for $g \geq 1$ and thus

$$u_{g,odd}^{((2);(1))} = (-1)^{g+1} \quad \text{for } g \geq 0.$$

Example 7.3. The result for $a_{[2]}|_{g,odd}$ is

$$a_{[2]}|_{g,odd} = -u_g^{((2);(1))} - u_g^{((1);(2))} = (-1)^g - q^{2g} \quad \text{for } g \geq 0.$$

Example 7.4. The result for $a_{[1^2]}|_{g,odd}$ is

$$a_{[1^2]}|_{g,odd} = u_g^{((1,1);(1,1))} + u_g^{((1);(2))} = -1 + q^{2g} \quad \text{for } g \geq 0.$$

Remark 7.5. The result for $(q^2+1)\, a_0|_{g,odd} - a_{[2]}|_{g,odd}$ can be found in lecture notes by Bradley Brock and Andrew Granville from 28 July 2003.

Example 7.6. Consider the case $(\mathbf{n};\mathbf{r}) = ((1,1,1);(2,1,1))$. We have $u_{-1} = J = 1$ and from Theorem 4.12 we get $u_0 = -(q-2)u_{-1} = -q+2$. Theorem 4.14 gives the recurrence relation $u_g = 2u_{g-1} - u_{g-2}$ for $g \geq 1$ and hence

$$u_{g,odd}^{((1,1,1);(2,1,1))} = g(-q+1) - q + 2.$$

7.2. Degree 4 or 5. From Theorem 4.12 we find that when the degree of the pair $(\mathbf{n};\mathbf{r})$ is 4 or 5 we need the base case of genus 0. But the genus 0 case is always computable using Lemma 6.1 and then Theorem 4.12, and hence the same is true for u_g for all g.

Example 7.7. For $(\mathbf{n};\mathbf{r}) = ((2,1,1);(1,1,1))$ we have $u_{-1} = q$ and from Lemma 6.1 it follows that

$$u_0^{((2,1,1);(1,1,1))} = -u_0^{((2,1);(1,2))} = u_0^{((1,1);(2,2))} + u_0^{((1);(2))} = q.$$

Using Theorem 4.12 we get $u_1 = -(q-1)u_0 - (q^2-q-1)u_{-1} = -q^3 + 2q$. Solving the recurrence relation $u_g = u_{g-1} - u_{g-2} - u_{g-3}$ for $g \geq 2$, coming from Theorem 4.14, gives

$$u_{g,odd}^{((2,1,1);(1,1,1))} = \frac{1}{4}\,(q^3-q)(-2g + (-1)^g - 1) + q.$$

Example 7.8. The result for $a_{[1^2,2]}|_{g,odd}$ is

$$\begin{aligned} a_{[1^2,2]}|_{g,odd} = &-u_g^{((2,1,1);(1,1,1))} - u_g^{((2,1);(1,2))} - u_g^{((1,1,1);(2,1,1))} \\ &- u_g^{((1,1);(2,2))} - 2u_g^{((1,1);(1,1))} - u_g^{((1);(2))} = \\ = -\frac{q^{2g+2}-1}{q+1} &- q^{2g} + \frac{1}{2}\,g(q^3+q-2) + \frac{1}{2}\begin{cases} 2q & \text{if } g \equiv 0 \bmod 2 \\ q^3-q-2 & \text{if } g \equiv 1 \bmod 2 \end{cases}\end{aligned}$$

7.3. Weight 6. We will not be able to compute u_g for all pairs $(\mathbf{n};\mathbf{r})$ of degree 6. But we will be able to compute u_g for all pairs $(\mathbf{n};\mathbf{r})$ that are general cases in the decomposition of $a_\lambda|_g$ for λ's of weight 6. This will be sufficient to compute all $a_\lambda|_g$ of weight 6, because we saw in Lemma 3.11 that only the general case will have degree 6 and therefore all degenerate cases are covered in Sections 7.1 and 7.2.

Let u_g be the general case in the decomposition of $a_\lambda|_g$. When the degree is equal to 6 we see from Theorem 4.12 that we need the base cases of genus 0 and 1 to compute u_g for all g. As we know, we can always compute u_0 using Lemma 6.1. For genus 1, the numbers $a_\lambda|_1$ have been computed for weight up to 6 by the author. This was done by embedding every genus 1 curve with a given point as a plane cubic curve, see [1, Section 15]. Since we know all the degenerate cases in the decomposition of $a_\lambda|_1$ we can then compute the general case u_1.

EXAMPLE 7.9. Let us deal with $(\mathbf{n};\mathbf{r}) = ((6);(1))$ which is the generic case in the decomposition of $a_{[6]}|_{g,odd}$ and for which we have $u_{-1} = J = q^3+q-1$. Using Lemma 6.1 we get

$$u_0^{((6);(1))} = -u_0^{((3);(2))} - u_0^{((2);(1))} - u_0^{((1);(2))} = -u_0^{((3);(2))} = -q^2.$$

Using the results of [1, Section 15] we find that $a_{[6]}|_1 = q-1$. Decomposing $a_{[6]}|_g$ gives $a_{[6]}|_1 = -u_1^{((6);(1))} - u_1^{((3);(2))} - u_1^{((2);(1))} - u_1^{((1);(2))}$. Thus, using Example 7.2, we get $u_1 = -(q-1)-(q^4-q^2-q-1)-1-q^2 = -q^4+1$. We can now apply Theorem 4.12 which gives $u_2 = -(q+1)u_1-(q^2+q+1)u_0-(q^3+q^2+q+1)u_{-1} = -q^6+q^2-q$, $u_3 = -u_2-u_1-u_0-u_{-1} = q^6+q^4-q^3$ and $u_4 = -u_3-u_2-u_1-u_0-u_{-1} = 0$. If we then multiply the characteristic polynomial for the linear recurrence relation of u_g by $X-1$ we get $u_g = u_{g-6}$ for all $g \geq 5$.

EXAMPLE 7.10. The result for $a_{[6]}|_{g,odd}$ is

$$a_{[6]}|_{g,odd} = -u_g^{((6);(1))} - u_g^{((3);(2))} - u_g^{((2);(1))} - u_g^{((1);(2))} = -q^{2g} - \frac{q^{2g+3}(q-1)}{q^2-q+1} +$$
$$+\frac{1}{q^2-q+1}\begin{cases} q^2 & \text{if } g \equiv 0 \bmod 3 \\ -q^2-1 & \text{if } g \equiv 1 \bmod 3 \\ 1 & \text{if } g \equiv 2 \bmod 3 \end{cases} + \begin{cases} q^2+1 & \text{if } g \equiv 0 \bmod 6 \\ q^4-2 & \text{if } g \equiv 1 \bmod 6 \\ q^6-q^2+q+1 & \text{if } g \equiv 2 \bmod 6 \\ -q^6-q^4+q^3-1 & \text{if } g \equiv 3 \bmod 6 \\ 1 & \text{if } g \equiv 4 \bmod 6 \\ -q^3-q & \text{if } g \equiv 5 \bmod 6 \end{cases}$$

REMARK 7.11. For any choice of λ and g, consider $a_\lambda|_{g,odd}$ as a function of the number q of elements of the finite field k of odd characteristic. If λ is of weight at most 7 it follows from our computations that this function is a polynomial in the variable q.
This will not continue to hold when considering for instance $a_{[1^6]}|_3$, that is, also including finite fields of *even* characteristic, see Example 10.6. But it will also not hold for instance for $a_{[1^{10}]}|_{1,odd}$, which for prime fields will be a polynomial function minus the Ramanujan τ-function, compare [15, Corollary 5.4].

8. REPRESENTATIVES OF HYPERELLIPTIC CURVES IN EVEN CHARACTERISTIC

Let k be a finite field with an even number of elements. We will again describe the hyperelliptic curves of genus $g \geq 2$ defined over k by their degree 2 morphism to $\mathbf{P}^1$. If we choose an affine coordinate x on $\mathbf{P}^1$ we can write the induced degree 2 extension of the function field of $\mathbf{P}^1$ in the form $y^2+h(x)y+f(x)=0$, where h and f are polynomials defined over k that fulfill the following conditions:

$$2g+1 \leq \max\big(2\deg(h), \deg(f)\big) \leq 2g+2; \tag{8.1}$$

$$\gcd(h, f'^2 + fh'^2) = 1; \tag{8.2}$$

$$t \nmid \gcd(h_\infty, f_\infty'^2 + f_\infty h_\infty'^2). \tag{8.3}$$

The last condition comes from the nonsingularity of the point(s) in infinity, around which the curve can be described in the variable $t = 1/x$ as $y^2 + h_\infty(t)y + f_\infty(t) = 0$, where $h_\infty := t^{g+1}h(1/t)$ and $f_\infty := t^{2g+2}f(1/t)$. We therefore define $h(\infty)$ and $f(\infty)$ to be equal to the degree $g+1$ and $2g+2$ coefficient respectively. For a reference see for instance [19, p. 294].

Definition 8.1. Let P_g denote the set of pairs (h, f) of polynomials defined over k, where h is nonzero, that fulfill all three conditions (8.1), (8.2) and (8.3). Write $C_{(h,f)}$ for the curve corresponding to the element (h, f) in P_g.

To each k-isomorphism class of objects in $\mathcal{H}_g(k)$ there is a pair (h, f) in P_g such that $C_{(h,f)}$ is a representative. All k-isomorphisms between the curves represented by elements of P_g are given by k-isomorphisms of their function fields, and since the g^1_2 of a hyperelliptic curve is unique the k-isomorphisms must respect the inclusion of the function field of $\mathbf{P}^1$.
Identify the set of polynomials $l(x)$ defined over k and of degree at most $g+1$ with k^{g+2}, and define the group homomorphism

$$\phi_g : \mathrm{GL}_2^{\mathrm{op}}(k) \times k^* \to \mathrm{Aut}(k^{g+2}),\ \phi_g(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, e)\big(l(x)\big) := e^{-1}(cx+d)^{g+1} l\Big(\frac{ax+b}{cx+d}\Big).$$

Now define the group $G_g := \big(k^{g+2} \rtimes_{\phi_g} (\mathrm{GL}_2^{\mathrm{op}}(k) \times k^*)\big)/D$ where

$$D := \{(0, \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, a^{g+1}) : a \in k^*\} \subset k^{g+2} \rtimes_{\phi_g} (\mathrm{GL}_2^{\mathrm{op}}(k) \times k^*).$$

The k-isomorphisms between curves corresponding to elements of P_g are then precisely the ones induced by elements of G_g by letting

$$\gamma = [(l(x), \begin{pmatrix} a & b \\ c & d \end{pmatrix}, e)] \in G_g$$

induce the isomorphism

$$(x, y) \mapsto \left(\frac{ax+b}{cx+d}, \frac{e\big(y + l(x)\big)}{(cx+d)^{g+1}} \right).$$

This defines a left group action of G_g on P_g, where $\gamma = [(l, \Lambda, e)] \in G_g$ takes $(h, f) \in P_g$ to $(\tilde{h}, \tilde{f}) \in P_g$, with

$$(\tilde{h}, \tilde{f}) = (\phi_g(\Lambda, e)(h), e^{-1}\phi_{2g}(\Lambda, e)(f) + l\,\phi_g(\Lambda, e)(h) + l^2).$$

Definition 8.2. Let τ_m be the function that takes $(a,b) \in k_m^2$ to 1 if the equation y^2+ay+b has two roots defined over k_m, 0 if it has one root and -1 if it has none.

Lemma 8.3. *If $C_{(h,f)}$ is the hyperelliptic curve corresponding to $(h,f) \in P_g$ then*

$$a_m(C_{(h,f)}) = -\sum_{\alpha \in \mathbf{P}^1(k_m)} \tau_m\big(h(\alpha), f(\alpha)\big).$$

Proof: Follows in the same way as Lemma 3.4. □

Notation 8.4. Let us put $I_g := 1/|G_g| = q^{-(g+2)}(q^3-q)^{-1}(q-1)^{-1}$.

In the same way as in the case of odd characteristic we get the equality

$$a_\lambda|_g = I_g \sum_{(h,f)\in P_g} \prod_{i=1}^{\nu} \Big(-\sum_{\alpha \in \mathbf{P}^1(k_i)} \tau_i\big(h(\alpha), f(\alpha)\big)\Big)^{\lambda_i}.$$

All results of Section 3.1 are independent of the characteristic and hence we extend the definition of $a_\lambda|_g$ to genus 0 and 1 in the same way as in that section.

Definition 8.5. For any $g \geq -1$, $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ and $\alpha \in A(\mathbf{n})$ define

$$u_{g,\alpha}^{(\mathbf{n};\mathbf{r})} := I_g \sum_{(h,f)\in P_g} \prod_{i=1}^{m} \tau_{n_i}\big(h(\alpha_i), f(\alpha_i)\big)^{r_i}$$

and

$$u_g^{(\mathbf{n};\mathbf{r})} := \sum_{\alpha \in A(\mathbf{n})} u_{g,\alpha}^{(\mathbf{n};\mathbf{r})}.$$

Construction-Lemma 8.6. *For each λ we have (in even characteristic) the same decomposition of $a_\lambda|_g$ as given by Construction-Lemma 3.8.*

Proof: The following properties of τ_m for $(h,f) \in P_g$ correspond precisely to the ones for the quadratic character.

- ⋆ Say that $\alpha \in \mathbf{P}^1(k_s)$, then $\tau_{\tilde{s}}\big(h(\alpha), f(\alpha)\big) = \tau_s\big(h(\alpha), f(\alpha)\big)^2$ if $\tilde{s}/s$ is even, and $\tau_{\tilde{s}}\big(h(\alpha), f(\alpha)\big) = \tau_s\big(h(\alpha), f(\alpha)\big)$ if $\tilde{s}/s$ is odd.
- ⋆ If for any $\alpha, \beta \in \mathbf{P}^1$ we have $F^s(\alpha) = \beta$ for some s, then $\tau_i\big(h(\alpha), f(\alpha)\big) = \tau_i\big(h(\alpha), f(\beta)\big)$ for all i.
- ⋆ Finally, for any $\alpha \in \mathbf{P}^1$ and any s, $\tau_{,s}\big(h(\alpha), f(\alpha)\big)^r = \tau_s\big(h(\alpha), f(\alpha)\big)^2$ if r is even and $\tau_s\big(h(\alpha), f(\alpha)\big)^r = \tau_s\big(h(\alpha), f(\alpha)\big)$ if r is odd.

With this established we can use the same proof as for Construction-Lemma 3.8. □

Since the decompositions are the same, Lemmas 3.11 and 3.12 also hold in even characteristic.

9. Recurrence relations for u_g in even characteristic

Analogously to Section 4, this section will be devoted to finding for a fixed pair $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$, a recurrence relation for u_g. Fix an $s \in k$ which does not lie in the set $\{r^2+r : r \in k\}$, that is, such that $\tau_1(1,s) = -1$. We define an involution on P_g sending (h,f) to $(h, f + s\,h^2)$. This involution is fixed point free and hence

$$u_{g,\alpha} = I_g \sum_{(h,f)\in P_g} \prod_{i=1}^{m} \tau_{n_i}\big(h(\alpha_i), f(\alpha_i)\big)^{r_i} =$$

$$= I_g \sum_{(h,f)\in P_g} \prod_{i=1}^{m} \tau_{n_i}\big(h(\alpha_i), f(\alpha_i) + s\,h^2(\alpha_i)\big)^{r_i} = (-1)^{\sum_{i=1}^{m} r_i n_i} u_{g,\alpha}.$$

Thus, Lemma 4.1 also holds in the case of even characteristic.

Definition 9.1. Let Q_g denote the set of pairs (h,f) of polynomials over k, where h is nonzero and h, f are of degree at most $g+1$, $2g+2$ respectively. Extending the definition for P_g above to a pair $(h,f) \in Q_g$, let $h(\infty)$ and $f(\infty)$ be equal to the degree $g+1$ and $2g+2$ coefficient of h and f respectively. For any $g \geq -1$, $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ and $\alpha \in A(\mathbf{n})$ define

$$\hat{U}^{(\mathbf{n};\mathbf{r})}_{g,\alpha} := I_g \sum_{(h,f)\in Q_g} \prod_{i=1}^{m} \tau_{n_i}\big(h(\alpha_i), f(\alpha_i)\big)^{r_i}$$

and

$$\hat{U}^{(\mathbf{n};\mathbf{r})}_{g} := \sum_{\alpha \in A(\mathbf{n})} \hat{U}^{(\mathbf{n};\mathbf{r})}_{g,\alpha}.$$

Remark 9.2. The connection between the sets Q_g and P_g which we will present below is due to Brock and Granville and can be found in an early version of [7]. There the connection is used to count the number of hyperelliptic curves in even characteristic, which is $a_0|_{g,even}$ in our terminology.

Lemma 9.3. *Let h and f be polynomials over k. For any irreducible polynomial m over k, the following two statements are equivalent:*

- $\star$ $m|\gcd(h, {f'}^2 + f{h'}^2)$;
- $\star$ *there is a polynomial l over k, such that $m|h$ and $m^2|f + hl + l^2$.*

Proof: Say that $\alpha \in k_n$ is a root of an irreducible polynomial m and of the polynomial $\gcd(h, {f'}^2 + f{h'}^2)$. Let l be equal to $f^{q^n/2}$. Working modulo $(x-\alpha)^2$ we then get

$$\begin{aligned} f + hl + l^2 &= f + hf^{q^n/2} + f^{q^n} \\ &\equiv f(\alpha) + f'(\alpha)(x-\alpha) + h'(\alpha)f(\alpha)^{q^n/2}(x-\alpha) + f(\alpha)^{q^n} \\ &\equiv (x-\alpha)(f'(\alpha) + h'(\alpha)f(\alpha)^{q^n/2}) \equiv (x-\alpha)(f'(\alpha)^2 + h'(\alpha)^2 f(\alpha))^{1/2} = 0, \end{aligned}$$

which tells us that $m^2|f+hl+l^2$. For the other direction, assume that we have an irreducible polynomial m and a polynomial l such that $m|h$ and $m^2|f+hl+l^2$.

Differentiating the polynomial $f + hl + l^2$ gives $m^2 | f' + h'l + hl'$, and thus $m | f' + h'l$. Taking squares we get $m^2 | f'^2 + h'^2 l^2$ and then it follows that $m^2 | f'^2 + h'^2(f + hl)$ and hence $m | f'^2 + h'^2 f$. □

Let (h, f) be an element of Q_g. In the first part of the proof of Lemma 9.3, we may take for l any representative of $f^{q^n/2}$ modulo h, because for these l we have $f + hl + l^2 \equiv f + hf^{q^n/2} + f^{q^n}$ modulo $(x - \alpha)^2$. In the second part it does not matter which degree l has. We conclude from this that Lemma 9.3 also holds if we assume that l is of degree at most $g + 1$.
Choose $g \geq -1$ and let $(h, f) \in Q_g$. Lemma 9.3 gives the following alternative formulation of the conditions (8.1), (8.2) and (8.3). For all polynomials l of degree at most $g + 1$:

$$m | h,\ m^2 | f + hl + l^2 \implies \deg(m) = 0; \tag{9.1}$$

$$\deg(h) = g + 1 \quad \text{or} \quad \deg(f + hl + l^2) \geq 2g + 1. \tag{9.2}$$

Here we used that $t | \gcd(h_\infty, f'^2_\infty + f_\infty h'^2_\infty)$ if and only if $t | h_\infty$ and there exists a polynomial l_∞ such that $\deg(l_\infty) \leq g + 1$ and $t^2 | f_\infty + h_\infty l_\infty + l^2_\infty$. In turn, this happens if and only if $\deg(h) \leq g$ and there exists a polynomial l of degree at most $g+1$ such that $\deg(f + hl + l^2) \leq 2g$, where we connect l and l_∞ using the definitions $l := x^{g+1} l_\infty(1/x)$ and $l_\infty := t^{g+1} l(1/t)$.
This reformulation leads us to making the following definition.

DEFINITION 9.4. Let $\sim_g$ be the relation on Q_g given by $(h, f) \sim_g (h, f + hl + l^2)$ if l is a polynomial of degree at most $g + 1$. This is an equivalence relation and since $(h, f) = (h, f + hl + l^2)$ if and only if $l = 0$ or $l = h$, the number of elements of each equivalence class $[(h, f)]_g$ is $q^{g+2}/2$. If $(h, f) \in P_g \subset Q_g$ then $[(h, f)]_g \subset P_g$ and we get an induced equivalence relation on P_g which we also denote $\sim_g$.

We will now construct all $\sim_g$ equivalence classes of elements of Q_g in terms of the $\sim_i$ equivalence classes of the elements in P_i, where i is between -1 and g. This is the counterpart of factoring a polynomial into a square-free part and a squared part in the case of odd characteristic.

DEFINITION 9.5. For $z := [(h, f)]_i \in P_i / \sim_i$ let V_z be the set of all equivalence classes $[(mh, m^2 f)]_g$ in Q_g for all monic polynomials m of degree at most $g - i$. This is well defined since if $(h_1, f_1) \sim_i (h_2, f_2)$ then $(mh_1, m^2 f_1) \sim_g (mh_2, m^2 f_2)$.

LEMMA 9.6. *The sets V_z for all $z \in P_i / \sim_i$ where $-1 \leq i \leq g$ are disjoint.*

Proof: Say that for some z_1 and z_2 the intersection $V_{z_1} \cap V_{z_2}$ is nonempty. That is, there exist $(h_1, f_1) \in P_{i_1}$, $(h_2, f_2) \in P_{i_2}$ and monic polynomials m_1, m_2 such that $m_1 h_1 = m_2 h_2$ and $m_1^2 f_1 = m_2^2 f_2 + m_2 h_2 l + l^2$. If for some irreducible polynomial r we have $r | m_1$ but $r \nmid m_2$, it follows that $r | h_2$ and $r^2 | m_2^2 f_2 + m_2 h_2 l + l^2$. By the equivalence of conditions (8.2) and (9.1), this implies that $r | (m_2^2 f_2)'^2 + m_2^2 f_2 (m_2 h_2)'^2$ which in turn implies that $r | f_2'^2 + f_2 h_2'^2$. Since $(h_2, f_2) \in P_{i_2}$ we see that r must be constant. Hence every irreducible

factor of m_1 is a factor of m_2. The situation is symmetric and therefore the converse also holds.
So far we have not ruled out the possibility that a factor in m_1 appears with higher multiplicity than in m_2, or vice versa. Let m be the product of all irreducible factors of m_1 and put $\tilde{m}_1 := m_1/m$, $\tilde{m}_2 := m_2/m$ and $\tilde{l} := l/m$. We are then in the same situation as above, that is $\tilde{m}_1 h_1 = \tilde{m}_2 h_2$ and $\tilde{m}_1^2 f_1 = \tilde{m}_2^2 f_2 + \tilde{m}_2 h_2 \tilde{l} + \tilde{l}^2$. Thus, if r is an irreducible polynomial such that $r|\tilde{m}_1$ but $r \nmid \tilde{m}_2$ we can argue as above to conclude that r is constant. By a repeated application of this line of reasoning we can conclude that m_1 and m_2 must be equal.
It now follows that $h_1 = h_2$ and that $m_2|l$, thus $(h_1, f_1) \sim_{i_1} (h_2, f_2)$. This tells us that $V_{z_1} \cap V_{z_2}$ is only nonempty when $z_1 = z_2$. $\square$

Lemma 9.7. *The sets V_z for all $z \in P_i/\sim_i$ where $-1 \leq i \leq g$ cover $Q_g/\sim_g$.*

Proof: Pick any element $(h_1, f_1) \in Q_g$ and put $g_1 := g$. We define a procedure, where at the ith step we ask if there are any polynomials m_i and l_i such that $\deg(m_i) > 0$, $\deg(l_i) \leq g_i + 1$, $m_i|h_i$ and $m_i^2|f_i + h_i l_i + l_i^2$. If so, take any such polynomials m_i, l_i and define $h_{i+1} := h_i/m_i$, $f_{i+1} := (f_i + h_i l_i + l_i^2)/m_i^2$ and $g_{i+1} := g_i - \deg(m_i)$. This procedure will certainly stop. Assume that the procedure has been carried out in some way and that it has stopped at the jth step, leaving us with some pair of polynomials (h_j, f_j).
Next, we take (h_j, f_{j+1}) to be any element of the set $[(h_j, f_j)]_{g_j}$ for which $\deg(f_{j+1})$ is minimal. Say that $f_{j+1} = f_j + h_j l_j + l_j^2$ where $\deg(l_j) \leq g_j + 1$ and let us define g_{j+1} to be the number such that $2g_{j+1} + 1 \leq \max\bigl(2\deg(h_j), \deg(f_{j+1})\bigr) \leq 2g_{j+1} + 2$. The claim is now that $(h_j, f_{j+1}) \in P_{g_{j+1}}$. By definition, condition (8.1) holds for (h_j, f_{j+1}). If there were polynomials m_{j+1} and l_{j+1} such that $m_{j+1}|h_j$ and $m_{j+1}^2|f_{j+1} + h_j l_{j+1} + l_{j+1}^2$ then the pair of polynomials m_{j+1} and $l_j + l_{j+1}$ would contradict that the process above stopped at the jth step. Hence condition (9.1) is fulfilled for (h_j, f_{j+1}). Condition (9.2) is fulfilled if $2\deg(h_j) \geq \deg(f_{j+1})$ because then $\deg(h_j) = g_{j+1} + 1$. On the other hand, if $2\deg(h_j) < \deg(f_{j+1})$ and there were a polynomial l_{j+1} such that $\deg(l_{j+1}) \leq g_{j+1} + 1$ and $\deg(f_{j+1} + h_j l_{j+1} + l_{j+1}^2) \leq 2g_{j+1}$ then this would contradict the minimality of $\deg(f_{j+1})$. We conclude that $(h_j, f_{j+1}) \in P_{g_{j+1}}$.
Finally we see that if we put $\hat{m}_r := \prod_{i=1}^{r-1} m_i$ and $l := \sum_{i=1}^{j} \hat{m}_i l_i$, then $\deg(l) \leq g + 1$, $h_1 = \hat{m}_j h_j$ and $f_1 = \hat{m}_j^2 f_{j+1} + h_1 l + l^2$. This shows that V_z contains $[(h_1, f_1)]_g$ where $z := [(h_j, f_{j+1})]_{g_{j+1}} \in P_{g_{j+1}}/\sim_{g_{j+1}}$. $\square$

Using the lemmas above we will be able to write $\hat{U}_g$ in terms of u_i for i between -1 and g. After this we will determine $\hat{U}_g$ for large enough values of g. We divide into two cases.

Notation 9.8. Let S_j denote all polynomials of degree at most j, and let $S'_j \subset S_j$ consist of the monic polynomials.

9.1. THE CASE $\alpha \in A'(\mathbf{n})$. Fix an element $\alpha \in A'(\mathbf{n})$. It follows from Lemma 9.6 and Lemma 9.7 that

$$(9.3)\quad \hat{U}_{g,\alpha} = \frac{I_g}{2} \sum_{l \in S_{g+1}} \sum_{j=-1}^{g} \sum_{z \in P_j/\sim_j} \sum_{[(h,f)]_j \in V_z} \prod_{i=1}^{m} \tau_{n_i}\big(h(\alpha_i), (f+hl+l^2)(\alpha_i)\big)^{r_i}.$$

LEMMA 9.9. *Choose any $s \geq 1$ and t_1, t_2 in k_s. We then have*

$$\tau_s(vt_1, v^2 t_2) = \tau_s(t_1, t_2) \quad \textit{for all } v \neq 0 \in k_s;$$
$$\tau_s(t_1, t_2 + vt_1 + v^2) = \tau_s(t_1, t_2) \quad \textit{for all } v \in k_s.$$

Proof: Clear. □

Fix elements $z = [(h_0, f_0)]_i \in P_i/\sim_i$ and $\beta \in \mathbf{A}^1(k_s)$ and define V_z' to be the subset of V_z of classes $[(\tilde{m}h_0, \tilde{m}^2 f_0)]_g$, where $\tilde{m}$ is a monic polynomial with $\tilde{m}(\beta) \neq 0$. Lemma 9.9 shows that $\tau_s(h(\beta), f(\beta))$ is constant for all s and (h,f) such that $[(h,f)]_g \in V_z'$. Applying this to equation (9.3) after recalling Definition 4.4 we find that

$$(9.4)\quad \hat{U}_{g,\alpha} = I_g \frac{q^{g+1}}{2} \sum_{j=-1}^{g} \sum_{z \in P_j/\sim_j} \sum_{\tilde{m} \in S'_{g-j}} \prod_{i=1}^{m} \tau_{n_i}\big((\tilde{m}h)(\alpha_i), (\tilde{m}^2 f)(\alpha_i)\big)^{r_i} =$$
$$= I_g \frac{q^{g+1}}{2} \sum_{j=-1}^{g} \hat{b}_{g-j} u_{j,\alpha} \frac{2}{q^{j+1} I_j} = \sum_{i=0}^{g+1} \hat{b}_i u_{g-i,\alpha},$$

where we have taken into account that the group of isomorphisms depends upon g and that the numbers of elements of the equivalence classes of the relations $\sim_{g-j}$ and $\sim_g$ differ by a factor q^j. From the definitions we see that $q^{g-j}\, I_g/I_j = 1$.

For any $g \geq -1$ and any $h_0 \in S_{g+1}$ it is clear that

(9.5)

$$\sum_{(h_0,f) \in Q_g} \prod_{i=1}^{m} \tau_{n_i}\big(h_0(\alpha_i), f(\alpha_i)\big)^{r_i} = \begin{cases} 0 & \text{if } \forall i: r_i = 2,\ \exists j: h_0(\alpha_j) = 0; \\ q^{2g+3} & \text{if } \forall i: r_i = 2,\ \forall j: h_0(\alpha_j) \neq 0. \end{cases}$$

For any g such that $2g+2 \geq |\mathbf{n}| - 1$, and any nonzero polynomial h_0 of degree at most $g+1$, Lemma 4.7 tells us that

$$(9.6)\quad \sum_{(h_0, f_1 + p_\alpha f_2) \in Q_g} \prod_{i=1}^{m} \tau_{n_i}\big(h_0(\alpha_i), (f_1 + p_\alpha f_2)(\alpha_i)\big)^{r_i} =$$
$$= q^{2g+3-|\mathbf{n}|} \sum_{f_1 \in S_{|\mathbf{n}|-1}} \prod_{i=1}^{m} \tau_{n_i}\big(h_0(\alpha_i), f_1(\alpha_i)\big)^{r_i} =$$
$$= q^{2g+3-|\mathbf{n}|} \sum_{(\beta_1,\ldots,\beta_m) \in \prod_{i=1}^{m} k_{n_i}} \prod_{i=1}^{m} \tau_{n_i}\big(h_0(\alpha_i), \beta_i\big)^{r_i} = 0 \quad \text{if } \exists i: r_i = 1,$$

because for all $a \in k_s$ there are as many $b \in k_s$ for which $\tau_s(a,b) = 1$ as there are $b \in k_s$ for which $\tau_s(a,b) = -1$.

Summing equations (9.5) and (9.6) over all $h_0 \in S_{g+1}$ and using that $q^{2g+3}I_g = Iq^{g+1}$ we get

$$\hat{U}_{g,\alpha} = \begin{cases} I\,(q-1)q^{g+1}\hat{b}_{g+1} & \text{if } \forall i: r_i = 2,\ g \geq -1; \\ 0 & \text{if } \exists i: r_i = 1,\ g \geq \frac{|\mathbf{n}|-3}{2}. \end{cases} \tag{9.7}$$

9.2. The case $\alpha \in A(\mathbf{n}) \setminus A'(\mathbf{n})$. Fix an $\alpha \in A(\mathbf{n}) \setminus A'(\mathbf{n})$. We can assume that $\alpha_1 = \infty$, and then $\tilde{\alpha} := (\alpha_2, \ldots, \alpha_m) \in A'(\tilde{\mathbf{n}})$ where $\tilde{\mathbf{n}} := (n_2, \ldots, n_m)$.

Lemma 9.10. *For any element $(h, f) \in P'_i$ and any monic polynomial m of degree $g - i$,*

$$\tau_s((mh)(\infty), (m^2 f)(\infty)) = \tau_s(h(\infty), f(\infty));$$
$$\tau_s((mh)(\infty), (f + lh + l^2)(\infty)) = \tau_s(h(\infty), f(\infty)).$$

Proof: Clear. □

For any $(h, f) \in Q_g$ it holds that if $\deg(h) < g+1$ then $\tau_s(h(\infty), f(\infty)) = 0$ for all s. Define therefore P'_g and Q'_g to be the subsets of P_g and Q_g respectively, that consist of pairs (h, f) such that $\deg(h) = g+1$. We get an induced relation $\sim_i$ on P'_i and Q'_i and we let V''_z be the set of all equivalence classes $[(mh, m^2 f)]_g$ in Q'_g for all monic polynomials m of degree $g-i$, where $z := [(h, f)]_i \in P'_i/\sim_i$. In the same way as in Lemma 9.6 and 9.7 we see that the sets V''_z for all $z \in P'_i/\sim_i$, where $-1 \leq i \leq g$, are disjoint and cover $Q'_g/\sim_g$. Using this together with Lemma 9.10 and the arguments showing equation (9.4) we find that

$$\hat{U}_{g,\alpha} = \frac{I_g}{2} \sum_{l \in S_{g+1}} \sum_{z \in Q'_g/\sim_g} \prod_{i=1}^{m} \tau_{n_i}\big(h(\alpha_i), (f + hl + l^2)(\alpha_i)\big)^{r_i} = \tag{9.8}$$
$$= I_g \frac{q^{g+1}}{2} \sum_{j=-1}^{g} \sum_{z \in P'_j/\sim_j} \sum_{\tilde{m} \in R'_{g-j}} \prod_{i=1}^{m} \tau_{n_i}\big((\tilde{m}h)(\alpha_i), (\tilde{m}^2 f)(\alpha_i)\big)^{r_i} = \sum_{i=0}^{g+1} b_i^{\tilde{\mathbf{n}}} u_{g-i,\alpha}.$$

If we choose g such that $2g+2 \geq |\mathbf{n}| - 1$, $h_0 \in R_{g+1}$ and we put $p_\alpha(x) := x\,p_{\tilde{\alpha}}$, then we find in the same way as for equation (9.6) that

$$\sum_{(h_0, f_1 + p_\alpha f_2) \in Q_g} \prod_{i=1}^{m} \tau_{n_i}\big(h_0(\alpha_i), (f_1 + p_\alpha f_2)(\alpha_i)\big)^{r_i} = \tag{9.9}$$
$$= q^{2g+3-|\mathbf{n}|} \sum_{(\beta_1, \ldots, \beta_m) \in \prod_{i=1}^{m} k_{n_i}} \prod_{i=1}^{m} \tau_{n_i}\big(h_0(\alpha_i), \beta_i\big)^{r_i} = 0 \quad \text{if } \exists i: r_i = 1.$$

Since equation (9.5) also hold for $\alpha \in A(\mathbf{n}) \setminus A'(\mathbf{n})$ we find, by summing over all polynomials $h_0 \in R_{g+1}$, that

$$\hat{U}_{g,\alpha} = \begin{cases} I\,(q-1)q^{g+1} b_{g+1}^{\tilde{\mathbf{n}}} & \text{if } \forall i: r_i = 2,\ g \geq -1; \\ 0 & \text{if } \exists i: r_i = 1,\ g \geq \frac{|\mathbf{n}|-3}{2}. \end{cases} \tag{9.10}$$

9.3. The two cases joined. Recall that $J = (q-1)\, I\, |A(\mathbf{n})|$.

Theorem 9.11. *For any pair* $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$,

$$\sum_{j=0}^{g+1} \hat{b}_j u_{g-j} = \begin{cases} J\, q^{g+1} \hat{b}_{g+1} & \text{if } \forall i : r_i = 2,\ g \geq -1; \\ 0 & \text{if } \exists i : r_i = 1,\ g \geq \frac{|\mathbf{n}|-3}{2}. \end{cases}$$

Proof: The theorem follows from combining equations (9.4), (9.7), (9.8) and (9.10), using Lemma 4.10. □

Theorem 9.12. *For any pair* $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$,

$$\sum_{j=0}^{\min(|\mathbf{n}|-1,g+1)} (\hat{b}_j - q\hat{b}_{j-1}) u_{g-j} = \begin{cases} J\, q^{g+1} (\hat{b}_{g+1} - \hat{b}_g) & \text{if } \forall i : r_i = 2,\ g \geq 0; \\ 0 & \text{if } \exists i : r_i = 1,\ g \geq \frac{|\mathbf{n}|-1}{2}. \end{cases}$$

Proof: In the notation of the proof of Theorem 4.14, the theorem follows from applying Theorem 9.11 to the expression $F(g) - qF(g-1)$. □

Theorem 9.13. *By applying Theorem 9.12 to each pair* $(\mathbf{n};\mathbf{r})$ *appearing in the decomposition (given by Lemma 8.6) of* $a_\lambda|_{g,even}$ *we get a linear recurrence relation for* $a_\lambda|_{g,even}$*. The characteristic polynomial of this linear recurrence relation equals* (5.1).

Proof: We know that the decomposition of $a_\lambda|_g$ is independent of characteristic, and since the left hand side of the equation in Theorem 9.12 is the same as the left hand side of the equation of Theorem 4.14 this theorem follows in the same way as Theorem 5.2. □

10. Results for weight up to 7 in even characteristic

In this section we compute, for any number g and any finite field k of even characteristic, all $a_\lambda|_{g,even}$ of weight at most 7. First we will exploit the similarities of Theorems 4.12 and 9.11.

Lemma 10.1. *If* $g \geq n-2$ *then* $\hat{b}_{2g+2} = q^{g+1}\hat{b}_{g+1}$.

Proof: Fix a pair $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$. Lemma 4.13 tells us that $\hat{b}_j = q\hat{b}_{j-1} + d_{|\mathbf{n}|-1-j}$, so if $j \geq |\mathbf{n}|$ then $\hat{b}_j = q\hat{b}_{j-1}$ and thus $\hat{b}_j = q^{j+1-|\mathbf{n}|}\hat{b}_{|\mathbf{n}|-1}$. □

Remark 10.2. If $r_i = 1$ for some i and $g \geq (|\mathbf{n}|-3)/2$, then the recursive relations of Theorems 9.11 and 4.12 are equal. On the other hand, if $r_i = 2$ for all i we see from Lemma 10.1 that the recursive relations of Theorems 9.11 and 4.12 are equal if $g \geq |\mathbf{n}|-2$.

Theorem 10.3. *For weight less than or equal to* 5, $a_\lambda|_{g,even} = a_\lambda|_{g,odd}$ *as functions (in this case polynomials) in* q.

Proof: Consider any $a_\lambda|_g$ with $|\lambda| \leq 5$. By Lemma 3.12 it suffices to show that u_g is independent of characteristic when $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ is such that $\sum_{i=1}^m n_i r_i \leq 5$. Clearly $u_{-1} = J$ is always independent of characteristic. Clearly, Lemma 6.1 also holds in even characteristic. We can therefore assume that $r_i = 2$ for all i in the case of genus 0. But if $r_i = 2$ for all i then $|\mathbf{n}| \leq 2$ and hence, by Remark 10.2, u_0 will be independent of characteristic.
This takes care of the base cases of the recurrence relations for u_g when $g \geq 1$, given by Theorems 4.12 and 9.11. Again by Remark 10.2 we see that (both in the case when $r_i = 2$ for all i, and when $r_i = 1$ for some i) when $g \geq 1$ these recurrence relations are the same. We can therefore conclude that u_g is independent of characteristic for all g. $\square$

We will now compute $a_\lambda|_{g,even}$ for weight 6 in the same way as in Section 7.3. To compute u_g of degree at most 5 using Theorem 9.11 we need to find the base case u_0. But when the genus is 0 we can use Lemma 6.1 (which also holds in even characteristic) to reduce to the case that $r_i = 2$ for all i, which is always computable using Theorem 9.11.
What is left is the general case of the decomposition of $a_\lambda|_{g,even}$. We then need the base cases of genus 0 and 1. Again, the genus 0 part is no problem. The computation of $a_\lambda|_1$ in [1] is independent of characteristic. We can therefore compute the genus 1 part (compare Section 7.3).

Remark 10.4. As in the case of odd characteristic, for all g and all λ such that $|\lambda| \leq 7$, $a_\lambda|_{g,even}$ is a polynomial when considered as a function in the number q (compare Remark 7.11) of elements of the finite field k of even characteristic. In Theorem 10.3 we saw that the polynomial functions $a_\lambda|_{g,odd}$ and $a_\lambda|_{g,even}$ are equal (for a fixed g), if $|\lambda| \leq 5$. But for weight 6 there are λ such that the two polynomials are different, this occurs for the first time for genus 3, see Example 10.6.

Example 10.5. Let us compute $u_{g,even}$ when $(\mathbf{n};\mathbf{r}) = ((1,1);(2,2,2))$. We see that $u_{-1} = 1$ and Theorem 9.11 gives $u_0 = q^2 - 3q + 2$. This result is different from the 1 in the case of odd characteristic, see Example 4.16. Continued use of Theorem 9.11 gives $u_1 = q^4 - 3q^3 + 5q^2 - 6q + 3$ and then Theorem 9.12 gives

$$u_g = 2u_{g-1} - u_{g-2} + q^{2g-1}(q-1)^3 \quad \text{for } g \geq 2.$$

Solving this leaves us with

$$u_{g,even}^{((1,1,1);(2,2,2))} = \frac{(q-1)(q^{2g+3} + g(q^2-1) - 3q - 2)}{(q+1)^2}.$$

Example 10.6. The result for $a_{[1^6]}|_{g,even}$ is

$$a_{[1^6]}|_{g,even} = a_{[1^6]}|_{g,odd} - \frac{5}{8}\, g(g-1)(g-2)\big((g-3)(q-1) - 4\big).$$

EXAMPLE 10.7. The result for $a_{[1^2,4]}|_{g,even}$ is

$$a_{[1^2,4]}|_{g,even} = a_{[1^2,4]}|_{g,odd} - \frac{1}{4}\begin{cases} g(q-1) & \text{if } g \equiv 0 \bmod 4; \\ (g-1)(q-1) & \text{if } g \equiv 1 \bmod 4; \\ (g-2)(q-1) & \text{if } g \equiv 2 \bmod 4; \\ (g-3)(q-1)-4 & \text{if } g \equiv 3 \bmod 4. \end{cases}$$

11. COHOMOLOGICAL RESULTS

11.1. COHOMOLOGICAL RESULTS FOR $\mathcal{H}_{g,n}$. Define the local system $\mathbb{V} := R^1\pi_*(\mathbb{Q}_\ell)$ where $\pi : \mathcal{M}_{g,1} \to \mathcal{M}_g$ is the universal curve. For every partition (note that in this section we use a different notation for partitions) $\lambda = (\lambda_1 \geq \ldots \geq \lambda_g \geq 0)$ there is an irreducible representation of $\mathrm{GSp}(2g)$ with highest weight $(\lambda_1 - \lambda_2)\gamma_1 + \ldots + \lambda_g\gamma_g - |\lambda|\eta$, where the γ_i are suitable fundamental roots and η is the multiplier representation, and we define $\mathbb{V}_\lambda$ to be the corresponding local system. Let us also denote by $\mathbb{V}_\lambda$ its restriction to $\mathcal{H}_g$. In Lemma 13.5 below we will see that making an $\mathbb{S}_{\tilde{n}}$-equivariant count of points of $\mathcal{H}_{g,\tilde{n}}$ over a finite field k, for all $\tilde{n} \leq n$, is equivalent to computing the trace of Frobenius on the compactly supported ℓ-adic Euler characteristic $\mathbf{e}_c(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)$, for every λ with $|\lambda| \leq n$ (where $\ell \nmid |k|$). For more details, see [14] and [15].

Thus, we can use the results of Section 7 together with Theorem 3.2 in [1] to compute the ℓ-adic Euler characteristic $\mathbf{e}_c(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda)$ in $K_0(\mathsf{Gal}_\mathbb{Q})$, the Grothendieck group of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-representations, for every λ with $|\lambda| \leq 7$. Specifically, Theorem 3.2 in [1] tells us that if there is a polynomial P such that $\mathrm{Tr}(F, \mathbf{e}_c(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)) = P(q)$ for all finite fields k, possibly with the exception of a finite number of characteristics, then $\mathbf{e}_c(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda) = P(\mathbf{q})$, where $\mathbf{q}$ is the class of $\mathbb{Q}_\ell(-1)$ in $K_0(\mathsf{Gal}_\mathbb{Q})$. By excluding even characteristic, Section 7 (see Remark 7.11) and Lemma 13.5 shows that there is indeed such a polynomial for all g and all $|\lambda| \leq 7$.

EXAMPLE 11.1. For $g = 8$ and $\lambda = (5,1)$ we have

$$\mathbf{e}_c(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda) = 5\mathbf{q}^5 - 28\mathbf{q}^4 + 4\mathbf{q}^3 + 96\mathbf{q}^2 - 34\mathbf{q} - \mathbf{88}.$$

11.2. COHOMOLOGICAL RESULTS FOR $\overline{\mathcal{M}}_{2,n}$ AND $\mathcal{M}_{2,n}$. Using the stratification of $\overline{\mathcal{M}}_{g,n}$ we can make an $\mathbb{S}_n$-equivariant count of its number of points using the $\mathbb{S}_n$-equivariant counts of the points of $\mathcal{M}_{\tilde{g},\tilde{n}}$ for all $\tilde{g} \leq g$ and $\tilde{n} \leq n+2(g-\tilde{g})$ (see [13, Thm 8.13] and also [2]). Since all curves of genus 2 are hyperelliptic, $\mathcal{M}_{2,n}$ is equal to $\mathcal{H}_{2,n}$. Above, we have made $\mathbb{S}_n$-equivariant counts of $\mathcal{H}_{2,n}$ for $n \leq 7$ and they were all found to be polynomial in q. These $\mathbb{S}_n$-equivariant counts can now be complemented with ones of $\mathcal{M}_{1,n}$ for $n \leq 9$ (see [1, Section 15]) and of $\mathcal{M}_{0,n}$ for $n \leq 11$ (see [18, Prop 2.7]), which are also found to be polynomial in q. We can then apply Theorem 3.4 in [2] to conclude, for all $n \leq 7$, the $\mathbb{S}_n$-equivariant $\mathsf{Gal}_\mathbb{Q}$ (resp. Hodge) structure of the ℓ-adic (resp. Betti) cohomology of $\overline{\mathcal{M}}_{2,n}$.

In the theorems below we give the $\mathbb{S}_n$-equivariant Hodge Euler characteristic (which by purity is sufficient to conclude the Hodge structure) in terms of the Schur polynomials and $\mathbf{L}$, the class of the Tate Hodge structure of weight 2 in $K_0(\mathsf{HS}_\mathbb{Q})$, the Grothendieck group of rational Hodge structures. That is, the action of $\mathbb{S}_n$ on $\overline{\mathcal{M}}_{2,n}$ induces an action on its cohomology, and hence $H^i(\overline{\mathcal{M}}_{2,n} \otimes \mathbb{C}, \mathbb{Q})$ may be written as a direct sum of $H^i_\lambda(\overline{\mathcal{M}}_{2,n} \otimes \mathbb{C}, \mathbb{Q})$, which correspond to the irreducible representations of $\mathbb{S}_n$ indexed by $\lambda \vdash n$ and with characters χ_λ. In terms of this, the coefficient of the Schur polynomial s_λ is equal to $1/\chi_\lambda(id) \cdot \sum_i (-1)^i [H^i_\lambda(\overline{\mathcal{M}}_{2,n} \otimes \mathbb{C}, \mathbb{Q})]$. The results for $n \leq 3$ were previously known by the work of Getzler in [14, Section 8].

THEOREM 11.2. *The $\mathbb{S}_n$-equivariant Hodge Euler characteristic of $\overline{\mathcal{M}}_{2,4}$ is equal to*

$$\begin{gathered}(\mathbf{L}^7 + 8\mathbf{L}^6 + 33\mathbf{L}^5 + 67\mathbf{L}^4 + 67\mathbf{L}^3 + 33\mathbf{L}^2 + 8\mathbf{L} + \mathbf{1})s_4 \\ +(4\mathbf{L}^6 + 26\mathbf{L}^5 + 60\mathbf{L}^4 + 60\mathbf{L}^3 + 26\mathbf{L}^2 + 4\mathbf{L})s_{31} \\ +(2\mathbf{L}^6 + 12\mathbf{L}^5 + 28\mathbf{L}^4 + 28\mathbf{L}^3 + 12\mathbf{L}^2 + 2\mathbf{L})s_{2^2} \\ +(3\mathbf{L}^5 + 10\mathbf{L}^4 + 10\mathbf{L}^3 + 3\mathbf{L}^2)s_{21^2}\end{gathered}$$

THEOREM 11.3. *The $\mathbb{S}_n$-equivariant Hodge Euler characteristic of $\overline{\mathcal{M}}_{2,5}$ is equal to*

$$\begin{gathered}(\mathbf{L}^8 + 9\mathbf{L}^7 + 49\mathbf{L}^6 + 128\mathbf{L}^5 + 181\mathbf{L}^4 + 128\mathbf{L}^3 + 49\mathbf{L}^2 + 9\mathbf{L} + \mathbf{1})s_5 \\ +(6\mathbf{L}^7 + 48\mathbf{L}^6 + 156\mathbf{L}^5 + 227\mathbf{L}^4 + 156\mathbf{L}^3 + 48\mathbf{L}^2 + 6\mathbf{L})s_{41} \\ +(3\mathbf{L}^7 + 31\mathbf{L}^6 + 106\mathbf{L}^5 + 159\mathbf{L}^4 + 106\mathbf{L}^3 + 31\mathbf{L}^2 + 3\mathbf{L})s_{32} \\ +(8\mathbf{L}^6 + 42\mathbf{L}^5 + 65\mathbf{L}^4 + 42\mathbf{L}^3 + 8\mathbf{L}^2)s_{31^2} \\ +(6\mathbf{L}^6 + 26\mathbf{L}^5 + 43\mathbf{L}^4 + 26\mathbf{L}^3 + 6\mathbf{L}^2)s_{2^21} \\ +(\mathbf{L}^5 + 3\mathbf{L}^4 + \mathbf{L}^3)s_{21^3}\end{gathered}$$

THEOREM 11.4. *The $\mathbb{S}_n$-equivariant Hodge Euler characteristic of $\overline{\mathcal{M}}_{2,6}$ is equal to*

$$\begin{gathered}(\mathbf{L}^9 + 11\mathbf{L}^8 + 68\mathbf{L}^7 + 229\mathbf{L}^6 + 420\mathbf{L}^5 + 420\mathbf{L}^4 + 229\mathbf{L}^3 + 68\mathbf{L}^2 + 11\mathbf{L} + \mathbf{1})s_6 \\ +(7\mathbf{L}^8 + 75\mathbf{L}^7 + 317\mathbf{L}^6 + 641\mathbf{L}^5 + 641\mathbf{L}^4 + 317\mathbf{L}^3 + 75\mathbf{L}^2 + 7\mathbf{L})s_{51} \\ +(5\mathbf{L}^8 + 62\mathbf{L}^7 + 292\mathbf{L}^6 + 615\mathbf{L}^5 + 615\mathbf{L}^4 + 292\mathbf{L}^3 + 62\mathbf{L}^2 + 5\mathbf{L})s_{42} \\ +(\mathbf{L}^8 + 21\mathbf{L}^7 + 108\mathbf{L}^6 + 236\mathbf{L}^5 + 236\mathbf{L}^4 + 108\mathbf{L}^3 + 21\mathbf{L}^2 + \mathbf{L})s_{3^2} \\ +(17\mathbf{L}^7 + 118\mathbf{L}^6 + 278\mathbf{L}^5 + 278\mathbf{L}^4 + 118\mathbf{L}^3 + 17\mathbf{L}^2)s_{41^2} \\ +(16\mathbf{L}^7 + 115\mathbf{L}^6 + 277\mathbf{L}^5 + 277\mathbf{L}^4 + 115\mathbf{L}^3 + 16\mathbf{L}^2)s_{321} \\ +(3\mathbf{L}^7 + 22\mathbf{L}^6 + 53\mathbf{L}^5 + 53\mathbf{L}^4 + 22\mathbf{L}^3 + 3\mathbf{L}^2)s_{2^3} \\ +(9\mathbf{L}^6 + 29\mathbf{L}^5 + 29\mathbf{L}^4 + 9\mathbf{L}^3)s_{31^3} \\ +(6\mathbf{L}^6 + 21\mathbf{L}^5 + 21\mathbf{L}^4 + 6\mathbf{L}^3)s_{2^21^2}\end{gathered}$$

Theorem 11.5. *The $\mathbb{S}_n$-equivariant Hodge Euler characteristic of $\overline{\mathcal{M}}_{2,7}$ is equal to*

$$
\begin{aligned}
&(\mathbf{L}^{10}+12\mathbf{L}^9+90\mathbf{L}^8+363\mathbf{L}^7+854\mathbf{L}^6+1125\mathbf{L}^5+854\mathbf{L}^4+363\mathbf{L}^3+90\mathbf{L}^2+\ldots)s_7\\
&+(9\mathbf{L}^9+109\mathbf{L}^8+580\mathbf{L}^7+1529\mathbf{L}^6+2109\mathbf{L}^5+1529\mathbf{L}^4+580\mathbf{L}^3+109\mathbf{L}^2+9\mathbf{L})s_{61}\\
&+(6\mathbf{L}^9+100\mathbf{L}^8+606\mathbf{L}^7+1728\mathbf{L}^6+2430\mathbf{L}^5+1728\mathbf{L}^4+606\mathbf{L}^3+100\mathbf{L}^2+6\mathbf{L})s_{52}\\
&+(3\mathbf{L}^9+58\mathbf{L}^8+389\mathbf{L}^7+1153\mathbf{L}^6+1647\mathbf{L}^5+1153\mathbf{L}^4+389\mathbf{L}^3+58\mathbf{L}^2+3\mathbf{L})s_{43}\\
&+(28\mathbf{L}^8+258\mathbf{L}^7+831\mathbf{L}^6+1221\mathbf{L}^5+831\mathbf{L}^4+258\mathbf{L}^3+28\mathbf{L}^2)s_{51^2}\\
&+(34\mathbf{L}^8+331\mathbf{L}^7+1133\mathbf{L}^6+1675\mathbf{L}^5+1133\mathbf{L}^4+331\mathbf{L}^3+34\mathbf{L}^2)s_{421}\\
&+(12\mathbf{L}^8+140\mathbf{L}^7+489\mathbf{L}^6+738\mathbf{L}^5+489\mathbf{L}^4+140\mathbf{L}^3+12\mathbf{L}^2)s_{3^21}\\
&+(8\mathbf{L}^8+91\mathbf{L}^7+335\mathbf{L}^6+502\mathbf{L}^5+335\mathbf{L}^4+91\mathbf{L}^3+8\mathbf{L}^2)s_{32^2}\\
&+(28\mathbf{L}^7+143\mathbf{L}^6+228\mathbf{L}^5+143\mathbf{L}^4+28\mathbf{L}^3)s_{41^3}\\
&+(34\mathbf{L}^7+170\mathbf{L}^6+275\mathbf{L}^5+170\mathbf{L}^4+34\mathbf{L}^3)s_{321^2}\\
&+(10\mathbf{L}^7+47\mathbf{L}^6+77\mathbf{L}^5+47\mathbf{L}^4+10\mathbf{L}^3)s_{2^31}\\
&+(4\mathbf{L}^6+7\mathbf{L}^5+4\mathbf{L}^4)s_{31^4}\\
&+(2\mathbf{L}^6+6\mathbf{L}^5+2\mathbf{L}^4)s_{2^21^3}
\end{aligned}
$$

In Table 1 we present the nonequivariant information (remember that all cohomology is Tate) in the form of Betti numbers of $\overline{\mathcal{M}}_{2,n}$ for all $n \leq 7$. Notice that the table only contains as many numbers as we need to be able to fill in the missing ones using Poincaré duality. These results agree with Table 2 of ordinary Euler characteristics for $\overline{\mathcal{M}}_{2,n}$ for $n \leq 6$ found in [4].

Table 1. Dimensions of $H^i(\overline{\mathcal{M}}_{2,n} \otimes \mathbb{C}, \mathbb{Q})$ for $n \leq 7$.

	H^0	H^2	H^4	H^6	H^8	H^{10}
$\overline{\mathcal{M}}_2$	1	2				
$\overline{\mathcal{M}}_{2,1}$	1	3	5			
$\overline{\mathcal{M}}_{2,2}$	1	6	14			
$\overline{\mathcal{M}}_{2,3}$	1	12	44	67		
$\overline{\mathcal{M}}_{2,4}$	1	24	144	333		
$\overline{\mathcal{M}}_{2,5}$	1	48	474	1668	2501	
$\overline{\mathcal{M}}_{2,6}$	1	96	1547	8256	18296	
$\overline{\mathcal{M}}_{2,7}$	1	192	4986	39969	129342	189289

The theorem used above also gives the corresponding results for $\mathcal{M}_{2,n}$ for $n \leq 7$, which we will present in terms of local systems $\mathbb{V}_\lambda$ defined as above, but starting from $\mathbb{V} := R^1\pi_*\mathbb{Q}$. See [14, Section 8] for the results on $\mathbf{e}_c(\mathcal{M}_2 \otimes \mathbb{C}, \mathbb{V}_\lambda)$, for all λ of weight at most 3.

Theorem 11.6. *The Hodge Euler characteristics of the local systems* $\mathbb{V}_\lambda$ *on* $\mathcal{M}_2' := \mathcal{M}_2 \otimes \mathbb{C}$ *of weight* 4 *or* 6 *are equal to*

$$\mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(4,0)}) = \mathbf{0},\ \ \mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(3,1)}) = \mathbf{L}^2 - \mathbf{1},\ \ \mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(2,2)}) = -\mathbf{L}^4,$$
$$\mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(6,0)}) = -\mathbf{1},\ \ \mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(5,1)}) = \mathbf{L}^2 - \mathbf{L} - \mathbf{1},$$
$$\mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(4,2)}) = \mathbf{L}^3,\ \ \mathbf{e}_c(\mathcal{M}_2', \mathbb{V}_{(3,3)}) = -\mathbf{L} - \mathbf{1}.$$

12. Appendix: Introducing b_i, c_i and r_i

This section will give an interpretation of the information carried by the u_g's. It will be in terms of counts of hyperelliptic curves together with prescribed inverse images of points on $\mathbf{P}^1$ under their unique degree 2 morphism.

Definition 12.1. Let C_φ be a curve defined over k together with a separable degree 2 morphism φ over k from C to $\mathbf{P}^1$. We then define

$$b_i(C_\varphi) := |\{\alpha \in A(i) : |\varphi^{-1}(\alpha)| = 2, \varphi^{-1}(\alpha) \subseteq C(k_i)\}|,$$
$$c_i(C_\varphi) := |\{\alpha \in A(i) : |\varphi^{-1}(\alpha)| = 2, \varphi^{-1}(\alpha) \nsubseteq C(k_i)\}|$$

and put $r_i(C_\varphi) := b_i(C_\varphi) + c_i(C_\varphi)$.

The number of ramification points of f that lie in $A(i)$ is then equal to $|A(i)| - r_i(C_\varphi)$. Let λ_i denote the partition of i consisting of one element. We then find that

$$|C_\varphi(\lambda_i)| = |A(i)| + b_i(C_\varphi) - c_i(C_\varphi) + \begin{cases} 2c_{i/2}(C_\varphi) & \text{if } i \text{ is even;} \\ 0 & \text{if } i \text{ is odd.} \end{cases}$$

and thus

$$a_n(C_\varphi) = \sum_{i|n\,:\,2i\nmid n} \big(c_i(C_\varphi) - b_i(C_\varphi)\big) + \sum_{i:2i|n} \big(-b_i(C_\varphi) - c_i(C_\varphi)\big).$$

Definition 12.2. For partitions μ and ν, $g \geq 2$ and odd characteristic, define

$$b_\mu c_\nu|_g := \sum_{[C_f] \in \mathcal{H}_g(k)/\cong_k} \frac{1}{|\mathrm{Aut}_k(C_f)|} \prod_{i=1}^{l(\mu)} b_i(C_f)^{\mu_i} \prod_{j=1}^{l(\nu)} c_j(C_f)^{\nu_j}.$$

The number $|\mu| + |\nu|$ will be called the weight of this expression.

Remark 12.3. We can, in the obvious way, also define $a_\lambda b_\mu c_\nu|_g$, but from the relation between $a_i(C_f)$, $b_i(C_f)$ and $c_i(C_f)$ we see that this gives no new phenomena.

Directly from the definitions we get the following lemma.

Lemma 12.4. *Let the characteristic be odd and let* f *be an element of* P_g. *We then have*

$$b_i(C_f) = \frac{1}{2} \sum_{\alpha \in A(i)} \Big(\chi_{2,i}\big(f(\alpha)\big)^2 + \chi_{2,i}\big(f(\alpha)\big)\Big)$$

and

$$c_i(C_f) = \frac{1}{2} \sum_{\alpha \in A(i)} \Big(\chi_{2,i}\big(f(\alpha)\big)^2 - \chi_{2,i}\big(f(\alpha)\big) \Big).$$

If the characteristic is odd we then use the same arguments as in Section 3 to conclude that

$$b_\mu c_\nu|_g = \frac{I}{2^{|\mu|+|\nu|}} \sum_{f \in P_g} \prod_{i=1}^{l(\mu)} \Big(\sum_{\alpha \in A(i)} \chi_{2,i}\big(f(\alpha)\big) + \chi_{2,i}\big(f(\alpha)\big)^2 \Big)^{\mu_i} \cdot$$
$$\cdot \prod_{j=1}^{l(\nu)} \Big(\sum_{\alpha \in A(j)} \chi_{2,j}\big(f(\alpha)\big) - \chi_{2,j}\big(f(\alpha)\big)^2 \Big)^{\nu_j}.$$

Note that this expression is defined for all $g \geq -1$. It can be decomposed in terms of u_g's (that is, we can find a result corresponding to Lemma 3.8) for tuples $(\mathbf{n};\mathbf{r}) \in \mathcal{N}_m$ such that

$$|\mathbf{n}| \leq |\mu| + |\nu|. \tag{12.1}$$

Remark 12.5. The corresponding results clearly hold for elements (h, f) in P_g in even characteristic and the decomposition of $b_\mu c_\nu|_g$ is independent of characteristic.

Example 12.6. For each N we have the decomposition:

$$b_{[N]}|_g = \frac{1}{2}(u_g^{((N);(2))} + u_g^{((N);(1))}) \text{ and } c_{[N]}|_g = \frac{1}{2}(u_g^{((N);(2))} - u_g^{((N);(1))}).$$

Example 12.7. Let us decompose $b_{[1^2]}c_{[2]}|_g$ into u_g's:

$$b_{[1^2]}c_{[2]}|_g = \frac{1}{8}(u_g^{((2,1,1);(2,2,2))} + u_g^{((2,1,1);(2,1,1))} + 2u_g^{((2,1);(2,2))}$$
$$- u_g^{((2,1,1);(1,2,2))} - u_g^{((2,1,1);(1,1,1))} - 2u_g^{((2,1);(1,2))}).$$

In this expression we have removed the u_g's for which $\sum_{i=1}^m r_i n_i$ is odd, since they are always equal to 0.

Lemma 12.8. *For each N, the following information is equivalent:*

1. *all u_g's of degree at most N;*
2. *all $b_\mu c_\nu|_g$ of weight at most N.*

Proof: From property (12.1) of the decomposition of $b_\mu c_\nu|_g$ into u_g's we directly find that if we know (1) we can compute (2). For the other direction we note on the one hand that

$$I \sum_{f \in P_g} \prod_{i=1}^{j} \big(b_i(C_f) - c_i(C_f)\big)^{s_i} \big(b_i(C_f) + c_i(C_f)\big)^{t_i} \tag{12.2}$$

can be formulated in terms of $b_\mu c_\nu|_g$'s of weight at most

$$S := \sum_{i=1}^{j} i\,(s_i + t_i).$$

If we on the other hand decompose (12.2) into u_g's we find that there is a unique u_g of degree S. The corresponding pair $(\mathbf{n};\mathbf{r})$ contains, for each i, precisely s_i entries of the form i^1 and t_i entries of the form i^2. Every u_g of degree S can be created in this way and hence if we know (2) we can compute (1). $\square$

Remark 12.9. From the definitions of $a_i(C_f)$ and $r_i(C_f)$ we see that knowing (1) and (2) in Lemma 12.8 is also equivalent to knowing

(3) all $a_\lambda r_\xi|_g$ of weight at most N,

where $a_\lambda r_\xi|_g$ is defined in the obvious way. Moreover, $a_\lambda r_\xi|_g = 0$ if $|\lambda|$ is odd.

13. Appendix: The stable part of the counts

Remark 13.1. All results in this section are independent of characteristic.

Definition 13.2 ([8, Def. 1.2.1, 1.2.2]). Let $\mathcal{F}$ be a constructible (ℓ-adic) sheaf on a scheme X of finite type over $\mathbb{Z}$. The sheaf $\mathcal{F}$ is said to be *pure* of weight m if, for every closed point x in X and eigenvalue α of Frobenius F (relative to $k = k(x)$) acting on $\mathcal{F}_{\bar{x}}$, α is an algebraic integer of weight equal to m, i.e., such that all its conjugates have absolute value equal to $q^{m/2}$. The sheaf $\mathcal{F}$ is said to be *mixed* of weight $\leq m$ if there exists a filtration $0 = \mathcal{F}_{-1} \subset \mathcal{F}_0 \subset \ldots \subset \mathcal{F}_m = \mathcal{F}$ of constructible subsheaves such that, for all j, $\mathcal{F}_j/\mathcal{F}_{j-1}$ is pure of weight j.

Theorem 13.3 ([8, Cor. 3.3.3, 3.3.4]). *Let $X \xrightarrow{f} \mathbb{Z}$ be a scheme of finite type, and $\mathcal{F}$ a constructible sheaf mixed of weight $\leq m$. Then $R^i f_! \mathcal{F}$ is mixed of weight $\leq m+i$. Thus, for every finite field k, there is a filtration $0 = W_{-1} \subset W_0 \subset \ldots \subset W_{i+m} = H^i_c(X_{\bar{k}}, \mathcal{F})$ of $\mathrm{Gal}(\bar{k}/k)$-representations such that, for all j, W_j/W_{j-1} is pure of weight j.*

Definition 13.4. Let $K_0(\mathsf{Gal}_k)$ be the Grothendieck group of $\mathrm{Gal}(\bar{k}/k)$-representations. In this category, and with the notation of Theorem 13.3, we have $[H^i_c(X_{\bar{k}}, \mathcal{F})] = \sum_{j=0}^{i+m} [W_j/W_{j-1}]$. For any $w \geq 0$, let us define $[H^i_c(X_{\bar{k}}, \mathcal{F})]^w := \sum_{j=w}^{i+m} [W_j/W_{j-1}]$ and $\mathbf{e}^w_c(X_{\bar{k}}, \mathcal{F}) := \sum_{i \geq 0} (-1)^i [H^i_c(X_{\bar{k}}, \mathcal{F})]^w$ in $K_0(\mathsf{Gal}_k)$. We make the corresponding definition of $\mathbf{e}^w_c(X_{\overline{\mathbb{Q}}}, \mathcal{F})$ in $K_0(\mathsf{Gal}_{\mathbb{Q}})$.

Recall the definition in Section 11.1, for a prime $\ell \nmid q$, of the ℓ-adic local system $\mathbb{V}_\lambda$ on $\mathcal{H}_g$. If τ is the canonical morphism from $\mathcal{H}_g \otimes \bar{k}$ to H_g, we put $\mathbb{V}'_\lambda = \tau_* \mathbb{V}_\lambda$. This is a constructible sheaf pure of weight $|\lambda|$.

In this section we will see that if g and w are large enough we can compute the trace of Frobenius on $\mathbf{e}^w_c(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)$, which by definition (cf. Section 2 in [3]) is equal to $\mathbf{e}^w_c(H_g, \mathbb{V}'_\lambda)$. We first make the connection to $\mathbb{S}_n$-equivariant counts of points of $H_{g,n}$ explicit.

Lemma 13.5. *Let the symmetric polynomial $\mathbf{s}_{<\lambda>}$ be the Schur polynomial in the symplectic case (see [11, A.45]), and $\mathbf{p}_\lambda$ the power sum. If $\mathbf{s}_{<\lambda>} = \sum_{|\mu| \leq |\lambda|} m_\mu \, \mathbf{p}_\mu$ then*

$$\mathrm{Tr}\big(F, \mathbf{e}_c(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}'_\lambda)\big) = \sum_{|\mu| \leq |\lambda|} m_\mu \, q^{\frac{1}{2}(|\lambda| - |\mu|)} \, a_\mu|_g. \tag{13.1}$$

From Theorems 4.14 and 9.12 we see that only the u_g's with all $r_i = 2$ have inhomogeneous recurrence relations. Theorem 5.2 dealt with the homogeneous part of the linear recurrence relations for $a_\lambda|_g$. The following lemma, which is a direct consequence of Theorems 4.14, 9.12 and 5.2, deals with the "inhomogeneities".

LEMMA 13.6. *Denote by* $t_{\mathbf{n}}$ *the coefficient of* $u_g^{(\mathbf{n};(2,\ldots,2))}$ *in the decomposition of* $a_\lambda|_g$ *(given in Construction-Lemma 3.8). Each value of* $|\mathbf{n}|$ *for a pair* $(\mathbf{n};(2,\ldots,2))$ *appearing in this decomposition of* $a_\lambda|_g$ *is at most equal to* $|\lambda|/2$. *Define the polynomial*

$$f_{\mathbf{n}}(x) := \Big(\prod_{i=1}^{m}(x^{n_i}-1)\Big)/(x-1).$$

For $g \geq 0$, *let* $R_\lambda(q)|_g$ *be the sum, over the pairs* $(\mathbf{n};(2,\ldots,2))$ *that occur in the decomposition of* $a_\lambda|_g$, *of the polynomial quotients of,*

$$t_{\mathbf{n}}\, q^{2g+|\mathbf{n}|}\, J\,(q-1)\, f_{\mathbf{n}}(q) \quad by \quad f_{\mathbf{n}}(q^2), \tag{13.2}$$

which is of degree at most $(|\lambda|+4g-2)/2$. *The polynomial* $R_\lambda(q)|_g$ *is a particular solution to the recurrence relation, described in Section 5, for* $a_\lambda|_g$.

Since the power sums form a rational basis of the ring of symmetric polynomials, equation (13.1) and Theorem 13.3 show that $a_\lambda|_g$ is of the form $\sum_j z_j\alpha_j$ for a finite set of rational numbers z_j and distinct algebraic integers α_i of weight at most $|\lambda|+4g-2$ (note that $2g-1$ is the dimension of H_g). If our base field k is replaced by an extension k_m of degree m then $a_\lambda|_g$ is equal to $\sum_j z_j\alpha_j^m$. For $g \geq |\lambda|-1$, the linear recurrence relation for $a_\lambda|_g$ (see Section 5) shows that it can be written as the particular solution $R_\lambda(q)|_g$ plus the homogeneous part, an integer sum of $a_\lambda|_{\tilde{g}} - R_\lambda(q)|_{\tilde{g}}$ for $\tilde{g} \leq |\lambda|-2$. We then see that if $g \geq |\lambda|-1$ and $w = 5\,|\lambda|-9$, the homogeneous part of the solution to the linear recurrence relation for $a_\lambda|_g$ does not contribute to $\mathrm{Tr}\big(F, \mathbf{e}_c^w(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}'_\lambda)\big)$. To conclude this we used the fact that $\sum_i z_i\alpha_i^m = 0$ for all m implies that $z_i = 0$ for all i, where the z_i and α_i are complex numbers and the α_i are distinct and nonzero. We can now summarize using Theorem 3.2 in [1].

DEFINITION 13.7. For a polynomial $f(x) = \sum_i f_i x^i$ put $f^w(x) := \sum_{i \geq w} f_i x^i$.

THEOREM 13.8. *Let* $\mathbf{q}$ *denote the class of* $\mathbb{Q}_\ell(-1)$ *in* $K_0(\mathsf{Gal}_\mathbb{Q})$. *For* $g \geq |\lambda|-1$ *and* $w = 5\,|\lambda|-9$ *we have an equality in* $K_0(\mathsf{Gal}_\mathbb{Q})$,

$$\mathbf{e}_c^w(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda) = \sum_{|\mu| \leq |\lambda|} m_\mu\, \mathbf{q}^{\frac{1}{2}(|\lambda|-|\mu|)}\, R_\mu^{w-|\lambda|+|\mu|}(\mathbf{q})|_g.$$

EXAMPLE 13.9. In the case $\lambda = (4,2,2)$, for $w = 31$ and $g \geq 7$, we find that $\mathrm{Tr}\big(F, \mathbf{e}_c^w(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)\big)$ is equal to $f_g^w(q)$, where f_g is the polynomial quotient of $q^{2g+4}(3q^2+3q+2)$ by $(q^2+1)^2(q+1)^3$.

REMARK 13.10. By Poincaré duality (cf. Section 2 in [3]) we find that there is a filtration $0 = W'_{i+|\lambda|-1} \subset W'_{i+|\lambda|} \subset \ldots \subset W'_{2(2g-1+|\lambda|)} = H^i(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)$

of $\mathrm{Gal}(\bar{k}/k)$-representations such that W'_j/W'_{j-1} is pure of weight j. Let us define $[H^i(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)]_w := \sum_{j=i+|\lambda|}^{w} [W'_j/W'_{j-1}]$ and $\mathbf{e}_w(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda) := \sum_{i\geq 0}(-1)^i [H^i(\mathcal{H}_g \otimes \bar{k}, \mathbb{V}_\lambda)]_w$ in $K_0(\mathsf{Gal}_k)$ and similarily $\mathbf{e}_w(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda)$. Theorem 13.8 shows that, for $g \geq \tilde{g} \geq |\lambda| - 1$ and $w = 4\tilde{g} - 3\,|\lambda| + 7$, one has that $\mathbf{e}_w(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda)$ is *stable*, in the sense that it is independent of g.

Computations for λ's of low weight lead us to make a conjecture, which is true for $|\lambda| \leq 30$.

Conjecture 13.11. For $g \geq |\lambda| - 1$ and $w = 5\,|\lambda| - 9$, we have $\mathbf{e}_c^w(\mathcal{H}_g \otimes \overline{\mathbb{Q}}, \mathbb{V}_\lambda) = 0$ for all λ such that $\lambda_1 > |\lambda|/2$.

References

[1] Jonas Bergström. Cohomology of moduli spaces of curves of genus three via point counts. *J. Reine Angew. Math.*, 622:155–187, 2008.

[2] Jonas Bergström and Orsola Tommasi. The rational cohomology of $\overline{\mathcal{M}}_4$. *Math. Ann.*, 338(1):207–239, 2007.

[3] Jonas Bergström and Gerard van der Geer. The Euler characteristic of local systems on the moduli of curves and abelian varieties of genus three. *J. Topol.*, 1(3):651–662, 2008.

[4] Gilberto Bini and John Harer. Euler characteristics of moduli spaces of curves. Available at `arXiv:math.AG/0506083`.

[5] Gilberto Bini and Gerard van der Geer. The Euler characteristic of local systems on the moduli of genus 3 hyperelliptic curves. *Math. Ann.*, 332(2):367–379, 2005.

[6] Theo van den Bogaart and Bas Edixhoven. Algebraic stacks whose number of points over finite fields is a polynomial. In *Number fields and function fields—two parallel worlds*, volume 239 of *Progr. Math.*, pages 39–49. Birkhäuser Boston, Boston, MA, 2005.

[7] Bradley W. Brock and Andrew Granville. More points than expected on curves over finite field extensions. *Finite Fields Appl.*, 7(1):70–91, 2001.

[8] Pierre Deligne. La conjecture de Weil. II. *Inst. Hautes Études Sci. Publ. Math.*, (52):137–252, 1980.

[9] Carel Faber and Gerard van der Geer. Sur la cohomologie des systèmes locaux sur les espaces de modules des courbes de genre 2 et des surfaces abéliennes. I. *C. R. Math. Acad. Sci. Paris*, 338(5):381–384, 2004.

[10] Carel Faber and Gerard van der Geer. Sur la cohomologie des systèmes locaux sur les espaces de modules des courbes de genre 2 et des surfaces abéliennes. II. *C. R. Math. Acad. Sci. Paris*, 338(6):467–470, 2004.

[11] William Fulton and Joe Harris. *Representation theory*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991.

[12] Gerard van der Geer and Marcel van der Vlugt. Supersingular curves of genus 2 over finite fields of characteristic 2. *Math. Nachr.*, 159:73–81, 1992.

[13] E. Getzler and M. M. Kapranov. Modular operads. *Compositio Math.*, 110(1):65–126, 1998.

[14] Ezra Getzler. Topological recursion relations in genus 2. In *Integrable systems and algebraic geometry (Kobe/Kyoto, 1997)*, pages 73–106. World Sci. Publishing, River Edge, NJ, 1998.
[15] Ezra Getzler. Resolving mixed Hodge modules on configuration spaces. *Duke Math. J.*, 96(1):175–203, 1999.
[16] Ezra Getzler. Euler characteristics of local systems on $\mathcal{M}_2$. *Compositio Math.*, 132(2):121–135, 2002.
[17] Nicholas M. Katz and Peter Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45 of *AMS Colloquium Publications*. AMS, Providence, RI, 1999.
[18] Mark Kisin and Gus Lehrer. Equivariant Poincaré polynomials and counting points over finite fields. *J. Algebra*, 247(2):435–451, 2002.
[19] Qing Liu. *Algebraic geometry and arithmetic curves*, volume 6 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, 2002. Translated from the French by Reinie Erné, Oxford Science Publications.
[20] Orsola Tommasi. *Geometry of discriminants and cohomology of moduli spaces*. PhD thesis, Radboud Universiteit Nijmegen, November 2005.

Jonas Bergström
Korteweg-de Vries Instituut
Universiteit van Amsterdam
Postbus 94248
1090 GE, Amsterdam
The Netherlands
o.l.j.bergstrom@uva.nl

Spectral Analysis of Relativistic Atoms – Dirac Operators with Singular Potentials

Matthias Huber

Received: February 26, 2009

Communicated by Heinz Siedentop

Abstract. This is the first part of a series of two papers, which investigate spectral properties of Dirac operators with singular potentials. We examine various properties of complex dilated Dirac operators. These operators arise in the investigation of resonances using the method of complex dilations. We generalize the spectral analysis of Weder [50] and Šeba [46] to operators with Coulomb type potentials, which are not relatively compact perturbations. Moreover, we define positive and negative spectral projections as well as transformation functions between different spectral subspaces and investigate the non-relativistic limit of these operators. We will apply these results in [30] in the investigation of resonances in a relativistic Pauli-Fierz model, but they might also be of independent interest.

2000 Mathematics Subject Classification: 81C05 (47F05; 47N50; 81M05)
Keywords and Phrases: Dirac operator, Coulomb Potential, Spectral theory of non-self-adjoint operators, Non-relativistic limit

1 Introduction and Definitions

A fascinating question in the mathematical analysis of operators describing atomic systems is the fate of eigenvalues embedded in the continuous spectrum if a perturbation is "turned on". Typically, these eigenvalues "vanish" and one has absolutely continuous spectrum. But the eigenvalues leave a trace: For example, the scattering cross section shows bumps near the eigenvalues, or certain states with energies close to the eigenvalues have an extended lifetime (described by the famous "Fermi Golden Rule" [13, Equation (VIII.2), p. 142] on a certain time scale). These energies are called resonances or resonance energies. Mathematically, resonances are described by poles of a holomorphic

continuation of the resolvent (or matrix elements of it) or the scattering amplitude to a second sheet.
The generic systems in which resonances occur are many-particle systems. This can be many-electron systems, in which the electron-electron interaction is the perturbation. The corresponding physical effect is called "Auger effect": Excited states ("autoionizing states") relax by emission of electrons. Another typical system in a one- or many-electron atom interacting with the quantized electromagnetic field, in which case excited states can relax by emitting photons. Resonances can also occur in one-particle systems, although this is not typically the case. It is well known (see [8] for example) that for a Schrödinger operator with Coulomb potential the set of resonances is empty.
During the last decades numerous results were obtained in the mathematical investigation of resonances so that it seems hopeless to give a complete account of the available literature. Nevertheless we would like to give an overview and mention at least some of the relevant works.
The investigation of resonances as poles of holomorphic continuations of scattering amplitude and resolvent goes back to Weisskopf and Wigner [53] and Schwinger [45]. The mathematical theory of resonances was pushed further by Friedrichs [14], Livsic [36], and Howland [27, 28]. One of the mathematical methods in the spectral analysis is the method of complex dilation, which associates the "vanished" embedded eigenvalue with a non-real eigenvalue of a certain non-selfadjoint operator and was investigated by Aguilar and Combes [2] and Balslev and Combes [6] (see [43] for an overview). Resonances in the case of the Stark effect were investigated by Herbst [24] and by Herbst and Simon [25]. Simon [48] initiated the mathematical investigation of the time-dependent perturbation theory. This was carried on by Hunziker [32]. Herbst [23] proved exponential temporal decay for the Stark effect.
The spectral analysis of non-relativistic atoms in interaction with the radiation field was initiated by Bach, Fröhlich, and Sigal [4, 5]. It was carried on by Griesemer, Lieb und Loss [18], by Fröhlich, Griesemer und Schlein (see for example [15]) and many others (see for example Hiroshima [26], Arai and Hirokawa [3], Dereziński and Gérard [9], Hiroshima and Spohn [12]), Loss, Miyao and Spohn [37] or Hasler and Herbst [21, 20]). In particular, Bach, Fröhlich, and Sigal [5] proved a lower bound on the lifetime of excited states in non-relativistic QED. Later, an upper bound was proven by Hasler, Herbst, and Huber [22] (see also [29]) and by Abou Salem et al. [1]. Recently, Miyao and Spohn [38] showed the existence of a groundstate for a semi-relativistic electron coupled to the quantized radiation field.
Our overall aim is to show that the lifetime of excited states of a relativistic one-electron atom obeys Fermi's Golden Rule [30] and coincides with the non-relativistic result in leading order in the fine structure constant. We will investigate the necessary spectral properties of a Dirac operator with potential, projected to its positive spectral subspace, coupled to the quantized radiation field. Following Bach et al. [5] and Hasler et al. [22], our main technical tool is complex dilation in connection with the Feshbach projection method.

In this first part of the work, we investigate the necessary properties of one-particle Dirac operators with singular potentials. In particular, we will derive the necessary properties of complex dilated spectral projections and discuss the non-relativistic limit of complex dilated Dirac operators. This serves mainly as a technical input for the second part of our work [30]. However, we believe that some of the results presented in the first part are also of independent interest. Note that the method of complex dilation has already successfully been applied to Dirac operators (see Weder [50] and Šeba [46]). However, these authors assume the relative compactness of the electric potential so that their method does not apply to Coulomb type potentials. Note moreover that Weder [51] considers very general operators including relativistic spin-0-Hamiltonians with potentials with Coulomb singularity. The basic assumption of this work is, however, that the unperturbed operator is sectorial, which is not fulfilled for the Dirac operator. Our results cover a class of Dirac operators which includes Coulomb and Yukawa potentials (with exception of Lemma 11 and Lemma 12 which we prove for the Coulomb case only).
Our results about the spectral projections of the dilated Dirac operator can be used to generalize the Douglas-Kroll transformation (see Siedentop and Stockmeyer [47] and Huber and Stockmeyer [31]) to dilated operators.

2 Definitions and Overview

The free Dirac operator (with velocity of light $c > 0$)

$$D_{c,0} := -ic\boldsymbol{\alpha} \cdot \nabla + c^2\beta \tag{1}$$

is an operator on the Hilbert space $\mathcal{H} := L^2(\mathbb{R}^3; \mathbb{C}^4)$. It is self-adjoint on the domain $\mathrm{Dom}(D_{c,0}) := H^1(\mathbb{R}^3; \mathbb{C}^4)$ [49, Chapter 1.4]. Here $\boldsymbol{\alpha}$ is the vector of the usual Dirac α- matrices, and β is the Dirac β-matrix.
We define for $\epsilon > 0$ the strip $S_\epsilon := \{z \in \mathbb{C} || \mathrm{Im}\, z| < \epsilon\}$. Let $\chi : \mathbb{R}^3 \to \mathbb{R}$ a bounded, measurable function. We will suppose that there is a $\Theta > 0$ such that $\theta \mapsto \chi(e^\theta x)$ admits a holomorphic continuation to $\theta \in S_\Theta$ for all $x \in \mathbb{R}^3$. We abbreviate $\chi_\theta := \chi(e^\theta \cdot)$. We will need the following two properties at different places:

$$\sup_{\theta \in S_\Theta,\, x \in \mathbb{R}^3} |\chi(e^\theta x)| \leq 1 \tag{H1}$$

$$\sup_{x \in \mathbb{R}^3} |\chi(e^\theta x) - \chi(x)| \leq \tilde{C}|\theta| \quad \text{for some } \tilde{C} > 0 \tag{H2}$$

It is easy to see that these properties are fulfilled for the Coulomb potential ($\chi(x) = 1$) or the Yukawa potential ($\chi(x) = e^{-ax}$ for some $a > 0$). The Dirac operator with potential $V := \chi/|\cdot|$

$$D_{c,\gamma} := -ic\boldsymbol{\alpha} \cdot \nabla + c^2\beta - \gamma V \tag{2}$$

is an operator on the Hilbert space $L^2(\mathbb{R}^3; \mathbb{C}^4)$ as well. It is self-adjoint on the domain $\mathrm{Dom}(D_{c,\gamma}) := \mathrm{Dom}(D_{c,0}) = H^1(\mathbb{R}^3; \mathbb{C}^4)$ for $\gamma \in \mathbb{R}$ with $|\gamma| <$

$c\sqrt{3}/2$ [49, Chapter 4.3.3]. γ is called coupling constant. The interacting Dirac operator describes a relativistic electron in the field of a nucleus, where the free operator yields the kinetic energy of the electron, whereas the electric potential gives its potential energy in the electric field of the nucleus.
The operator $D_{c,\gamma}$ has the set $(-\infty,-c^2]\cup[c^2,\infty)$ as essential spectrum. We assume that the operator has a nonempty set of positive eigenvalues, all of which have finite multiplicity. We number the eigenvalues by $\tilde{E}_{n,l}(c,\gamma)$ (not counting multiplicities). Here $n\in\mathbb{N}$ (or $n\in\{1,\dots,N_{\max}\}$ for some $N_{\max}\in\mathbb{N}$ if there are only finitely many eigenvalues) denotes the principal quantum number and $l\in\{1,\dots,N_n\}$ for some $N_n\in\mathbb{N}$ labels the fine structure components. We choose the numbering in such a way that for all $n'>n$, all $l\in\{1,\dots,N_n\}$ and all $l'\in\{1,\dots,N_{n'}\}$ the inequality $\tilde{E}_{n,l}(c,\gamma)<\tilde{E}_{n',l'}(c,\gamma)$ holds and such that $\tilde{E}_{n,l}(c,\gamma)<\tilde{E}_{n,l'}(c,\gamma)$ for $l<l'$. This numbering is natural for all values of c for the Coulomb potential, where the eigenvalues are explicitly known (see [35]). The spectrum of a Dirac operators can be shown to have this structure if c is large enough for general potentials (see [49]). We set $E_{n,l}(c,\gamma):=\tilde{E}_{n,l}(c,\gamma)-c^2$.
We define for $\theta\in\mathbb{C}$ and $\gamma\in\mathbb{R}$ the dilated operators

$$D_{c,0}(\theta):=-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla+c^2\beta \tag{3}$$

and

$$D_{c,\gamma}(\theta):=-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla+c^2\beta-\gamma V(\theta) \tag{4}$$

with $V(\theta):=e^{-\theta}\chi_\theta V_C$ on $\mathrm{Dom}(D_{c,0}(\theta))=\mathrm{Dom}(D_{c,\gamma}(\theta))=H^1(\mathbb{R}^3;\mathbb{C}^4)$, where $V_C=1/|\cdot|$ is the Coulomb potential. It is clear that $D_{c,0}(\theta)$ is closed on this domain and that (because of Hardy's inequality) $D_{c,\gamma}(\theta)$ is at least well defined under assumption (H1). We shall prove further properties in Section 4. For technical reasons, we will assume $c\geq 1$ in the following. We will assume moreover that $\gamma\geq 0$. Further, we define for $\theta\in\mathbb{R}$ the unitary dilation $\mathcal{U}(\theta):L^2(\mathbb{R}^3;\mathbb{C}^4)\to L^2(\mathbb{R}^3;\mathbb{C}^4)$, $(\mathcal{U}(\theta)f)(x):=e^{\frac{3}{2}\theta}f(e^\theta x)$. It fulfills the identity $\mathcal{U}(\theta)D_{c,\gamma}\mathcal{U}(\theta)^*=D_{c,\gamma}(\theta)$. The operators $D_{c,\gamma}(\theta)$ are extensions of the operators $\mathcal{U}(\theta)D_{c,\gamma}\mathcal{U}(\theta)^*$ for complex θ. Note that the mapping $\mathcal{U}(\theta)$ cannot be continued as a bounded operator to a complex domain, but the mapping $\theta\mapsto\mathcal{U}(\theta)\psi$ for an analytic vector ψ admits such an continuation, whose radius of convergence depends on the vector ψ (cf. [42, Chapter X.6]). However, we will prove in Section 8, that under certain conditions the restrictions of $\mathcal{U}(\theta)$ to certain spectral subspaces have bounded, bounded invertible extensions.
We add a short guide through the paper: We define a version of the Foldy-Wouthuysen transformation for non-self-adjoint Dirac operators in Section 3. Just as its analog for self-adjoint operators, it diagonalizes the free Dirac operator. It is however not a unitary operator any more so that one has to prove explicit estimates on its norm (see Theorem 1). The Foldy-Wouthuysen transformation serves as a technical input for the following sections.
We prove in Section 4 that the method of complex dilation can be successfully applied to Dirac operators with potentials with Coulomb singularities. In particular, we shall see that the dilated operators define a holomorphic family of

type (A) in the sense of Kato (see Theorem 2). Moreover, we provide a spectral analysis of such operators in Theorem 3. Just as in the case of Schrödinger operator, the real eigenvalues remain fixed under the complex dilation, whereas the essential spectrum swings into the complex plane and thus reveals possible non-real eigenvalues, which correspond to resonances of the original self-adjoint operator (see Figure ??). Note that there are no resonances for the Coulomb potential (see Remark 3).

In Section 5 we extend the notion of positive and negative spectral projections to the complex dilated Dirac operators. The definition of the spectral projections in Formula (32) is a straightforward extension of a well known formula from Kato's book (see [33, Lemma VI.5.6]). The rest of this section is devoted to the proof that the operators defined in (32) are actually well defined projections (see Theorem 4), that they commute with the dilated Dirac operator (see Theorem 5), and that their range is what one expects it to be (see Theorem 5 as well), which is not completely obvious in the non-self-adjoint case. Note that the projections themselves are not orthogonal projections.

These results enable us to define transformation functions between the positive spectral projections of the dilated and not dilated Dirac operators in Section 6, which is essential in order to show that also the projected Dirac operators are holomorphic families – even if they are coupled to the quantized radiation field. This will be accomplished in [30]. Moreover, these results can be used to generalize [47] to complex dilated operators. Transformation functions as defined in Formula (60) are similarity transformations between two (not necessarily orthogonal) projections (see Formula (57) in Theorem 6). Note that our definition requires that the norm difference between the projections be smaller than one, but there are more general approaches. For details on transformation functions we refer the reader to [33, Chapter II.4].

In Theorem 7 in Section 7 we prove a resolvent estimate for the dilated Dirac operator projected and restricted onto its positive spectral subspace. In particular, we prove that the norm of the resolvent converges (essentially) to zero as the inverse distance to the right complex half plane. Note that this really requires the restriction of the operator to its positive spectral subspace and that the norm of the resolvent of a non-self-adjoint operator is not bounded from above by the inverse distance of the spectral parameter to the spectrum.

In Section 8 we will investigate the non-relativistic limit of dilated Dirac operators and thereby generalize and extend the results in Thaller's book [49] in various directions. We prove in Theorem 8 and Corollary 2 that complex dilated Dirac operators converge to the corresponding (complex dilated) Schrödinger operators in the sense of norm resolvent convergence as the velocity of light goes to infinity. As in the undilated case, this convergence is needed to gain information about the spectral projections onto the eigenspaces belonging to the real eigenvalues and their behaviour in the nonrelativistic limit (see for example Lemma 7 or Lemma 8). In particular, the complex dilation, restricted to an eigenspace is a bounded operator (uniformly in the dilation parameter and the velocity of light – see Lemma 9) and the projections onto the fine structure components are uniformly bounded as well (see Corollary 5). These statements will be needed in [30]. Note that for Schrödinger operators and non-relativistic QED the above mentioned problems are absent, since there is neither a fine structure splitting nor the additional parameter of the velocity of light which has to be controlled.

Moreover, we show in Theorem 9 and Theorem 10 that the lower Pauli spinor of a normed eigenfunction of the Dirac operator converges to zero in the sense of the Sobolev space $H^1(\mathbb{R}^3;\mathbb{C}^2)$ and that the upper Pauli spinor is bounded in the sense of $H^1(\mathbb{R}^3;\mathbb{C}^2)$ as the velocity of light tends to infinity. This shows that the notion of "large" and "small" components of a Dirac spinor, which is frequently used by physicists, is also justified for dilated operators. Moreover, it follows that certain expectation values of the Dirac $\boldsymbol{\alpha}$-matrix vanish as the velocity of light tends to infinity. We will apply this fact in [30].

Note that in the discussion of the non-relativistic limit in Section 8 we need some estimates from Bach, Fröhlich, and Sigal [5] which we cite in Appendix A for the convenience of the reader.

3 Foldy-Wouthuysen-Transformation

In this section we investigate the complex continuation of the Foldy-Wouthuysen transformation and show some important properties in Theorem 1. We need this as a technical input for the spectral analysis in the following sections. Let us mention that a complex continuation of the Foldy-Wouthuysen transformation was implicitly used by Evans, Perry, and Siedentop [11] for the investigation of the spectrum of the Brown-Ravenhall operator. Also Balslev and Helffer [7] use holomorphic continuations of the Foldy-Wouthuysen transformation.

For $p \in \mathbb{R}^3$ we define the matrix $D_{c,0}(p;\theta) := ce^{-\theta}\boldsymbol{\alpha}\cdot p + c^2\beta$. We use the convention $\sqrt{\cdot} : \mathbb{C}\setminus\mathbb{R}_0^- \to \mathbb{C} : \sqrt{z} = re^{\mathrm{i}\,\phi/2}$ for the complex square root, where $z = re^{\mathrm{i}\,\phi}$ with $r \geq 0$ and $-\pi < \phi < \pi$. Note that for $w \in \mathbb{C}$ with $|\arg w| \leq \frac{\pi}{4}$ the estimate

$$\operatorname{Re}\sqrt{w} \geq \sqrt{\operatorname{Re} w} \geq 0 \tag{5}$$

holds, which follows immediately from the formula $\cos(2\phi) = (\cos\phi)^2 - (\sin\phi)^2 \leq (\cos\phi)^2$. Next, we define for $p \in \mathbb{R}^3$ and $\theta \in S_{\pi/2}$ the matrix

$$\hat{U}_{\mathrm{FW}}(c,p;\theta) := \frac{1}{N_c(p;\theta)}\begin{pmatrix} (c^2+E_c(p;\theta))\mathbf{1}_{2\times 2} & ce^{-\theta}\boldsymbol{\sigma}\cdot p \\ -ce^{-\theta}\boldsymbol{\sigma}\cdot p & (c^2+E_c(p;\theta))\mathbf{1}_{2\times 2}\end{pmatrix}, \tag{6}$$

where $E_c(p;\theta) := \sqrt{e^{-2\theta}c^2p^2+c^4}$ and $N_c(p;\theta) := \sqrt{2E_c(p;\theta)(c^2+E_c(p;\theta))}$. $\hat{U}_{\mathrm{FW}}(c;\theta)$ is the maximal multiplication operator on $L^2(\mathbb{R}^3;\mathbb{C}^4)$ which is generated by $U_{\mathrm{FW}}(p,c;\theta)$. Analogously, we define

$$\hat{V}_{\mathrm{FW}}(p,c;\theta) := \frac{c^2+E_c(p;\theta) - ce^{-\theta}\beta\boldsymbol{\alpha}\cdot p}{N_c(p;\theta)} \tag{7}$$

and $V_{\mathrm{FW}}(c;\theta)$. The corresponding Fourier transforms are $U_{\mathrm{FW}}(c;\theta) := \mathcal{F}^{-1}\hat{U}_{\mathrm{FW}}(c;\theta)\mathcal{F}$ and $V_{\mathrm{FW}}(c;\theta) := \mathcal{F}^{-1}\hat{V}_{\mathrm{FW}}(c;\theta)\mathcal{F}$. Note that these operators coincide with the usual Foldy-Wouthuysen transformation for $\theta = 0$ (see [49]), but are not unitary for $\theta \notin \mathbb{R}$. Nevertheless they define a similarity transformation, which diagonalizes the free Dirac operator. This will be important in the following sections, since the diagonalized operator $\sqrt{-c^2 e^{-2\theta}\Delta + c^4}\beta$ is normal, contrary to the operator $D_{c,0}(\theta)$.

Theorem 1. *Let $\theta \in S_{\pi/4}$. Then the following statements hold:*

a) The operator $U_{\mathrm{FW}}(c;\theta)$ is a bounded operator on $L^2(\mathbb{R}^3;\mathbb{C}^4)$ with bounded inverse $V_{\mathrm{FW}}(c;\theta)$. There is a constant C_{FW} (independent of c and θ) such that

$$\|U_{\mathrm{FW}}(c;\theta)\| \le \sqrt{1 + C_{\mathrm{FW}}|\sin \operatorname{Im}\theta|} \tag{8}$$

and

$$\|V_{\mathrm{FW}}(c;\theta)\| \le \sqrt{1 + C_{\mathrm{FW}}|\sin \operatorname{Im}\theta|}. \tag{9}$$

b) The Foldy-Wouthuysen transformation diagonalizes the Dirac operator:

$$U_{\mathrm{FW}}(c;\theta)D_{c,0}(\theta)V_{\mathrm{FW}}(c;\theta) = \sqrt{-c^2 e^{-2\theta}\Delta + c^4}\beta. \tag{10}$$

Proof.
a) A simple calculation shows

$$\hat{U}_{\mathrm{FW}}(p,c;\theta)\hat{V}_{\mathrm{FW}}(p,c;\theta) = \hat{V}_{\mathrm{FW}}(p,c;\theta)\hat{U}_{\mathrm{FW}}(p,c;\theta) = \mathbf{1}. \tag{11}$$

We have $\|U_{\mathrm{FW}}(c;\theta)\| \le \sup_{p\in\mathbb{R}^3} \|\hat{U}_{FW,c}(p;\theta)\|$. Thus, it suffices to consider the case $c = 1$ and $\operatorname{Re}\theta = 0$. In view of the identity $\|\hat{U}_{\mathrm{FW,c}}(p;\theta)\|^2 = \|\hat{U}_{\mathrm{FW,c}}(p;\theta)^*\hat{U}_{\mathrm{FW,c}}(p;\theta)\|$ we find with $\vartheta \in (-\pi/4, \pi/4)$

$$\begin{aligned}\hat{U}_{\mathrm{FW,c}}(p;\mathrm{i}\,\vartheta)^*\hat{U}_{\mathrm{FW,c}}(p;\mathrm{i}\,\vartheta) &= \frac{(1+E_1(p;\mathrm{i}\,\vartheta))(1+E_1(p;-\mathrm{i}\,\vartheta)) + p^2}{\tilde{N}} \\ &+ \frac{\beta\boldsymbol{\alpha}\cdot p(e^{-\mathrm{i}\,\vartheta}(1+E_1(p;-\mathrm{i}\,\vartheta)) - e^{\mathrm{i}\,\theta}(1+E_1(p;\mathrm{i}\,\vartheta)))}{\tilde{N}},\end{aligned} \tag{12}$$

where $\tilde{N} := \sqrt{4E_1(p;\mathrm{i}\,\vartheta)E_1(p;-\mathrm{i}\,\vartheta)(1+E_1(p;\mathrm{i}\,\vartheta))(1+E_1(p;-\mathrm{i}\,\vartheta))}$. Note that the expression under the square root is real, and that $|1 + E_1(p;\pm\mathrm{i}\,\vartheta)| \ge |E_1(p;\pm\mathrm{i}\,\vartheta)| = \sqrt[4]{1 + 2\cos(2\vartheta)p^2 + p^4} \ge \sqrt[4]{1+p^4}$, where we used $|\vartheta| < \pi/4$. Thus the denominator in (12) can be estimated as

$$|\tilde{N}| \ge 2\sqrt{1+|p|^4}. \tag{13}$$

Next, observe that

$$|e^{\mathrm{i}\,\vartheta}E_1(p;\mathrm{i}\,\vartheta) - e^{-\mathrm{i}\,\vartheta}E_1(p;-\mathrm{i}\,\vartheta)| \le \frac{|\sin(2\vartheta)|}{\sqrt{p^2+\cos(2\vartheta)}}, \tag{14}$$

where we used the estimate $|w| \ge |\operatorname{Re} w|$ and (5). From (14) it follows that

$$\|\beta\boldsymbol{\alpha}\cdot p(e^{\mathrm{i}\,\vartheta}(1+E_1(p;\mathrm{i}\,\vartheta)) - e^{-\mathrm{i}\,\vartheta}(1+E_1(p;-\mathrm{i}\,\vartheta)))\| \le 2|p||\sin(\vartheta)| + |\sin(2\vartheta)|. \tag{15}$$

Moreover, we have

$$1-\frac{(1+E_1(p;-\mathrm{i}\,\vartheta))\,(1+E_1(p;\mathrm{i}\,\vartheta))+p^2}{\tilde N} = \frac{\tilde N^2+\left((1+E_1(p;-\mathrm{i}\,\vartheta))\,(1+E_1(p;\mathrm{i}\,\vartheta))+p^2\right)^2}{\tilde N\left(\tilde N+((1+E_1(p;-\mathrm{i}\,\vartheta))\,(1+E_1(p;\mathrm{i}\,\vartheta))+p^2)\right)}. \tag{16}$$

Using $((1+E_1(p;-\mathrm{i}\,\vartheta))(1+E_1(p;\mathrm{i}\,\vartheta))+p^2)>0$ and (13) we estimate the denominator by

$$|\tilde N(\tilde N+((1+E_1(p;-\mathrm{i}\,\vartheta))(1+E_1(p;\mathrm{i}\,\vartheta))+p^2))|\geq 4(1+|p|^4). \tag{17}$$

In order to estimate the enumerator we find after some calculations

$$\begin{aligned}&4E_1(p;-\mathrm{i}\,\vartheta)E_1(p;\mathrm{i}\,\vartheta)(1+E_1(p;-\mathrm{i}\,\vartheta))(1+E_1(p;\mathrm{i}\,\vartheta))\\ &\quad-((1+E_1(p;-\mathrm{i}\,\vartheta))(1+E_1(p;\mathrm{i}\,\vartheta))+p^2)^2\\ &\quad=2p^4+2(e^{2\mathrm{i}\,\vartheta}+e^{-2\mathrm{i}\,\vartheta})p^2+2p^2(e^{-2\mathrm{i}\,\vartheta}E_1(p;-\mathrm{i}\,\vartheta)+e^{2\mathrm{i}\,\vartheta}E_1(p;\mathrm{i}\,\vartheta))\\ &\quad-2p^2-2p^2(E_1(p;-\mathrm{i}\,\vartheta)+E_1(p;\mathrm{i}\,\vartheta))-2p^2E_1(p;-\mathrm{i}\,\vartheta)E_1(p;\mathrm{i}\,\vartheta).\end{aligned} \tag{18}$$

We combine suitable terms in (18): We have

$$(e^{2\mathrm{i}\,\vartheta}+e^{-2\mathrm{i}\,\vartheta})p^2-2p^2=2(\cos(2\vartheta)-1)p^2, \tag{19}$$

$$\begin{aligned}&|2p^2(e^{-2\mathrm{i}\,\vartheta}E_1(p;-\mathrm{i}\,\vartheta)+e^{2\mathrm{i}\,\vartheta}E_1(p;\mathrm{i}\,\vartheta))-2p^2(E_1(p;-\mathrm{i}\,\vartheta)+E_1(p;\mathrm{i}\,\vartheta))|\leq 4p^2\\ &\times|\sqrt{p^2+e^{2\mathrm{i}\,\vartheta}}-\sqrt{p^2+e^{-2\mathrm{i}\,\vartheta}}|\leq 4p^2\frac{2\sin(2\vartheta)}{|\sqrt{p^2+e^{2\mathrm{i}\,\vartheta}}+\sqrt{p^2+e^{-2\mathrm{i}\,\vartheta}}|}\leq 4|p|\sin(2\vartheta),\end{aligned} \tag{20}$$

and

$$|2p^4+2\cos(2\vartheta)p^2-2p^2E_1(p;-\mathrm{i}\,\vartheta)E_1(p;\mathrm{i}\,\vartheta)|\leq 2|\sin(2\vartheta)|^2. \tag{21}$$

Summarizing the estimates (13) and (15) through (21), we finally obtain

$$\|\hat U_{\mathrm{FW}}(\mathrm{i}\,\vartheta,p)^*\hat U_{\mathrm{FW}}(\mathrm{i}\,\vartheta,p)-1\|\leq\left[\frac{|p|+1}{\sqrt{1+|p|^4}}+\frac{p^2+2|p|+1}{1+|p|^4}\right]|\sin(\vartheta)|, \tag{22}$$

where we used that $|\sin(2\vartheta)|\leq 2|\sin\vartheta|$ for $|\vartheta|\leq\pi/4$. If we set $C_{\mathrm{FW}}:=\sup_{t\in\mathbb{R}_0^+}\left[\frac{t+1}{\sqrt{1+t^4}}+\frac{t^2+2t+1}{1+t^4}\right]<\infty$, equation (22) shows the claim on $U_{\mathrm{FW}}(c;\theta)$. The claim on the inverse operator $V_{\mathrm{FW}}(c;\theta)$ can be proven analogously.
b) We have $\hat U_{\mathrm{FW}}(c,p;\theta)D_{c,0}(p;\theta)\hat V_{\mathrm{FW}}(c,p;\theta)=D_{c,0}(p;\theta)\hat V_{\mathrm{FW}}(c,p;\theta)^2$ as well as $\hat V_{\mathrm{FW}}(c,p;\theta)=\hat U_{\mathrm{FW}}(c,p;\theta)-2ce^{-\theta}\beta\boldsymbol{\alpha}\cdot p/N_c(p;\theta)$. From this it follows that

$\hat{U}_{\rm FW}(c,p;\theta)D_{c,0}(p;\theta)\hat{V}_{\rm FW}(c,p;\theta) = D_{c,0}(p;\theta) - A$, where $A := \frac{1}{N_c(p;\theta)^2}$ $D_{c,0}(p;\theta)[2ce^{-\theta}\beta\boldsymbol{\alpha}\cdot p][c^2 + E_c(p;\theta) - ce^{-\theta}\beta\boldsymbol{\alpha}\cdot p]$. A little calculation shows $A = -\frac{2c^2e^{-2\theta}p^2E_c(p;\theta)\beta}{N_c(p;\theta)^2} + ce^{-\theta}\boldsymbol{\alpha}\cdot p$, which implies

$$\hat{U}_{\rm FW}(c,p;\theta)D_{c,0}(p;\theta)\hat{V}_{\rm FW}(c,p;\theta) = E_c(p;\theta)\beta \tag{23}$$

and thus proves (10). □

4 Dilation Analyticity and Spectrum

We show that the operators in equations (3) and (4) define holomorphic families of closed operators. Since we will be interested in the non-relativistic limit later on, we consider only such values of c and γ which can be dealt with using Hardy's inequality. For $\theta \in S_{\pi/2}$ we define the set $M_{\gamma/c} := \{\theta \in \mathbb{C} | \frac{2\gamma}{c} < \cos(\operatorname{Im}\theta)\}$. We define $V_1(\theta) := e^{-\theta/2}\chi_\theta\sqrt{V_C}$ and $V_2(\theta) := e^{-\theta/2}\sqrt{V_C}$. Note that $V(\theta) = V_1(\theta)V_2(\theta)$.

Theorem 2. *Let $\theta \in S_{\min\{\Theta,\pi/2\}}$ and suppose that (H1) holds. Then the operator $D_{c,\gamma}(\theta)$ is closed for $\frac{2\gamma}{c} < \cos(\operatorname{Im}\theta)$ on $\operatorname{Dom}(D_{c,\gamma}(\theta)) = H^1(\mathbb{R}^3;\mathbb{C}^4)$, and we have $D_{c,\gamma}(\theta)^* = D_{c,\gamma}(\bar{\theta})$. $D_{c,\gamma}(\theta)$ is a holomorphic family of type (A) in the sense of Kato for $\theta \in M_{\gamma/c}$. $D_{c,0}(\theta)$ is an entire family of type (A).*

Proof. For $f \in H^1(\mathbb{R}^3;\mathbb{C}^4)$ the estimate $\|D_{c,0}(\theta)f\|^2 \geq |\operatorname{Re} e^{-\theta}|^2c^2\|\nabla f\|^2$ holds. Hardy's inequality implies $\|\gamma V(\theta)f\|^2 \leq 4\gamma^2|e^{-\theta}|^2\|\nabla f\|^2$ and thus $\|\gamma V(\theta)f\| \leq \frac{2\gamma}{c\cos(\operatorname{Im}\theta)}\|D_{c,0}(\theta)f\|$, which proves that the operator $D_{c,\gamma}(\theta)$ is closed and has a bounded inverse. Thus, the domain $\operatorname{Dom}(D_{c,\gamma}(\theta)) = H^1(\mathbb{R}^3;\mathbb{C}^4)$ is independent of $\theta \in M_{\gamma/c}$. It is clear that for $f \in \operatorname{Dom}(D_{c,\gamma}(\theta))$ the mapping $M_{\gamma/c} \to L^2(\mathbb{R}^3;\mathbb{C}^4)$, $\theta \mapsto D_{c,\gamma}(\bar{\theta})f$ is holomorphic, which implies that $D_{c,\gamma}(\theta)$ is a holomorphic family of type (A) [33, Chapter VII-2.1].
Moreover, obviously $D_{c,\gamma}(\bar{\theta})^* \supset D_{c,\gamma}(\theta)$ holds. Thus, it suffices to prove the inclusion $\operatorname{Dom}(D_{c,\gamma}(\bar{\theta})^*) \subset \operatorname{Dom}(D_{c,\gamma}(\theta)) = \operatorname{Ran}(D_{c,\gamma}(\theta)^{-1})$. We adapt a well known strategy from the case of self-adjoint operators (cf. [52, Satz 5.14]). We have $\operatorname{Dom}(D_{c,\gamma}(\theta)^{-1}) = \operatorname{Ran}(D_{c,\gamma}(\bar{\theta})) = L^2(\mathbb{R}^3;\mathbb{C}^4)$. For $f \in \operatorname{Dom}(D_{c,\gamma}(\bar{\theta})^*)$ we find $f_0 := D_{c,\gamma}(\theta)^{-1}D_{c,\gamma}(\bar{\theta})^*f \in \operatorname{Dom}(D_{c,\gamma}(\theta)) \subset \operatorname{Dom}(D_{c,\gamma}(\bar{\theta})^*)$. Thus $D_{c,\gamma}(\theta)f_0 = D_{c,\gamma}(\bar{\theta})^*f_0$, and the definition of f_0 implies $D_{c,\gamma}(\bar{\theta})^*f = D_{c,\gamma}(\theta)f_0$. From this it follows that $D_{c,\gamma}(\bar{\theta})^*(f - f_0) = 0$ and thus $f - f_0 \in \mathrm{N}(D_{c,\gamma}(\bar{\theta})^*) = \operatorname{Ran}(D_{c,\gamma}(\bar{\theta}))^\perp = \{0\}$, implying $f = f_0 \in \operatorname{Dom}(D_{c,\gamma}(\theta))$. □

Remark 1. *Note that if V is the Coulomb potential or the Yukawa potential, then $D_{c,\gamma}(\theta)$ is equal to a multiple of the self-adjoint operator $-ic\boldsymbol{\alpha}\cdot\nabla + V_C$ up to a bounded operator so that the proof of the above theorem is trivial. Note moreover, that for $V = V_C$, the operator $D_{c,\gamma}(\theta)$ is entire.*

Remark 2. *Theorem 2 and its proof imply that $H^1(\mathbb{R}^3;\mathbb{C}^4)$ is the maximal domain of the operator on $L^2(\mathbb{R}^3;\mathbb{C}^4)$ generated by the differential expression*

$\tilde{D}_{c,\gamma}(\theta) := -e^{-\theta} ic\boldsymbol{\alpha}\cdot\nabla + c^2\beta - \gamma V(\theta)$. *To see this set*

$$M_{\max} := \{f \in L^2(\mathbb{R}^3;\mathbb{C}^4) | \tilde{D}_{c,\gamma}(\theta) f \in L^2(\mathbb{R}^3;\mathbb{C}^4)\},$$

where the gradient is to be understood in distributional sense. Note that $f \in M_{\max}$ *implies* $\nabla f \in L^1_{loc}(\mathbb{R}^3;\mathbb{C}^4)$*, since* $V(\theta) \in L^2(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$*. If* $M_{\max} \supsetneq H^1(\mathbb{R}^3;\mathbb{C}^4)$*, then the operator* $D'_{c,\gamma}(\theta)$ *defined by the differential expression* $\tilde{D}_{c,\gamma}(\theta)$ *on the domain* $\mathcal{D}(D'_{c,\gamma}(\theta)) := M_{\max}$ *is a strict extension of the operator* $D_{c,\gamma}(\theta)$*. As in the proof of Theorem 2 it would follow that there was a* $0 \neq g \in M_{\max}$ *such that* $D'_{c,\gamma}(\theta) g = 0$*. It follows by partial integration from* $\nabla g \in L^1_{loc}(\mathbb{R}^3;\mathbb{C}^4)$ *that* $(\tilde{D}_{c,\gamma}(\bar{\theta}) f, g) = 0$ *for all* $f \in C_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$*. By density of* $C_0^\infty(\mathbb{R}^3;\mathbb{C}^4)$ *in* $H^1(\mathbb{R}^3;\mathbb{C}^4)$ *this equality extends to* $(D_{c,\gamma}(\bar{\theta}) f, g) = 0$ *for all* $f \in H^1(\mathbb{R}^3;\mathbb{C}^4) = \mathcal{D}(D_{c,\gamma}(\bar{\theta}))$*. Since* $D_{c,\gamma}(\bar{\theta})$ *is onto, it follows* $g = 0$*, a contradiction, which implies* $H^1(\mathbb{R}^3;\mathbb{C}^4) = M_{\max}$*.*

The following lemma, whose simple proof we omit, contains a useful fact:

Lemma 1. *Let* $a, b > 0$*. Then* $\sup_{p\in\mathbb{R}^3} \frac{\sqrt{a^2c^2p^2+c^4}}{\sqrt{b^2c^2p^2+c^4}} \leq \max\{1, \frac{a}{b}\}$*.*

Now we need the spectrum of the operator $D_{c,\gamma}(\theta)$. Theorem 1 shows (see Figure 1) $\sigma(D_{c,0}(\theta)) = \Sigma_c^-(\theta) \cup \Sigma_c^+(\theta)$, where $\Sigma_c^\pm(\theta) = \pm E_c(\mathbb{R};\theta)$.
In the case of self-adjoint operators the compactness of the difference of free and interacting resolvent would imply that $D_{c,0}(\theta)$ and $D_{c,\gamma}(\theta)$ with $\gamma \neq 0$ have the same essential spectrum. This is however not true for non-self-adjoint operators in general. In particular there exist several different definitions of the essential spectrum, which do not coincide in general and have different invariance properties.
In the case of relatively compact perturbations this difficulty can be mastered using the analytic Fredholm theorem [50]. Since Coulomb type potentials are not relatively compact, we adapt a strategy invented by Nenciu [40] for the self-adjoint case. We need the following lemma:

Lemma 2. *Let* $\theta \in S_{\pi/4}$ *and* $z \notin \sigma(D_{c,0}(\theta))$*. Then the operator* $V_C^{1/2}(D_{c,0}(\theta) - z)^{-1}$ *is compact.*

Proof. It suffices to consider the case $z = 0$. We write $V_C^{1/2} D_{c,0}(\theta)^{-1} = V_C^{1/2} \left(\sqrt{-c^2e^{-2\theta}\Delta + c^4}\beta\right)^{-1} \left(\sqrt{-c^2e^{-2\theta}\Delta + c^4}\beta\right) D_{c,0}(\theta)^{-1}$. Because of $V_C^{1/2} \in L^6_w(\mathbb{R}^3)$ and $1/(\pm\sqrt{c^2e^{-2\theta}(\cdot)^2 + c^4} - z) \in L^6(\mathbb{R}^3)$, the operator $V_C^{1/2}(\sqrt{-c^2e^{-2\theta}\Delta + c^4}\beta - z)^{-1}$ is compact [44]. Moreover, Theorem 1 implies $(\sqrt{-c^2e^{-2\theta}\Delta + c^4}\beta) D_{c,0}(\theta)^{-1}\| \leq 1 + C_{\mathrm{FW}}|\mathrm{Im}\,\theta|$. This shows the claim. ☐

For $z \notin \sigma(D_{c,0}(\theta))$ we define the operator $M_{c;\theta}(z) := V_2(\theta)(D_{c,0}(\theta) - z)^{-1} V_1(\theta)$. Moreover, let $B_{c;\theta;+}$ and $B_{c;\theta;-}$ (see Figure 1) the closed subsets of $\{z \in \mathbb{C} | \mathrm{Re}\, z > 0\}$ and $\{z \in \mathbb{C} | \mathrm{Re}\, z < 0\}$ respectively, which are enclosed be the

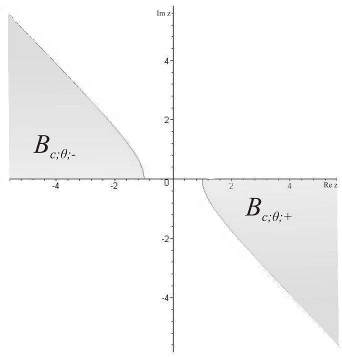


Figure 1: The spectrum of the operator $D_{c,0}(\theta)$ and sets $B_{c;\theta;\pm}$ for $c = 1$ and $\theta = \mathrm{i}\,\pi/4$.

curves $[c^2, \infty)$ and $E_c(\mathbb{R};\theta)$ ($(-\infty, -c^2]$ and $-E_c(\mathbb{R};\theta)$ respectively). We set $B_{c;\theta} = B_{c;\theta;+} \cup B_{c;\theta;-}$.
Furthermore, for $\theta \in S_{\pi/4}$ we define the constants

$$C(\operatorname{Im}\theta) := \frac{1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|}{\sqrt{\cos(2\operatorname{Im}\theta)}}, \quad C_1(\operatorname{Im}\theta) := C(\operatorname{Im}\theta) + \frac{1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|}{\cos(\operatorname{Im}\theta)}. \tag{24}$$

Note the inequality $1/\cos(\operatorname{Im}\theta) \leq C(\operatorname{Im}\theta)$.
The following theorem yields a precise description of the spectrum of the operator $D_{c,\gamma}(\theta)$. In particular, outside the set $B_{c,\theta}$ the spectra of $D_{c,\gamma}(\theta)$ and $D_{c,\gamma}(0)$ coincide so that one particle resonances – if any exist – can be located only within the set $B_{c,\theta}$.
Let $\mathcal{B}(L^2(\mathbb{R}^3;\mathbb{C}^4))$ be the set of bounded and everywhere defined operators on $L^2(\mathbb{R}^3;\mathbb{C}^4)$. Moreover, we set $B_a(x_0) := \{x \in \mathbb{R}^3 | |x - x_0| < a\}$ for $a > 0$ and $x_0 \in \mathbb{R}^3$

Theorem 3. *Let $\theta \in S_{\min\{\pi/4,\Theta\}}$ and $\frac{2\gamma}{c}C(\operatorname{Im}\theta) < 1$. Suppose that (H1) holds. Then $\sigma(D_{c,\gamma}(\theta)) = \sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta}$, where $A_{c,\gamma;\theta}$ is a discrete subset of $\mathbb{C} \setminus \sigma(D_{c,0}(\theta)$, and we have $A_{c,\gamma;\theta} \cap (\mathbb{C} \setminus B_{c;\theta}) = \sigma_{\mathrm{disc}}(D_{c,\gamma}(0))$. The set $A_{c,\gamma;\theta}$ has at most the accumulation points $\pm c^2$. For $z \notin \sigma(D_{c,\gamma}(\theta))$ the resolvent identity*

$$\begin{aligned}(D_{c,\gamma}(\theta) - z)^{-1} &= (D_{c,0}(\theta) - z)^{-1} + \\ &+ \gamma(D_{c,0}(\theta) - z)^{-1}V_1(\theta)(1 - e^{-\theta}\gamma M_{c;\theta}(z))^{-1}V_2(\theta)(D_{c,0}(\theta) - z)^{-1}\end{aligned} \tag{25}$$

holds.

Proof. We denote the r.h.s. of (25) by $R_{c,\gamma;\theta}(z)$.

Step 1: Proof of (25) *for* $z = \mathrm{i}\,\eta$, $\eta \in \mathbb{R}$. Using Kato's inequality and Theorem 1 we obtain

$$\begin{aligned}\|\gamma M_{c;\theta}(\mathrm{i}\,\eta)\| = \|\gamma V_2(\theta)(D_{c,0}(\theta) - \mathrm{i}\,\eta)^{-1} V_1(\theta)\| \le \frac{\gamma\pi e^{-\mathrm{Re}\,\theta}(1 + C_{\mathrm{FW}}|\mathrm{Im}\,\theta|)}{2} \\ \times \|\frac{|\nabla|}{\sqrt{-\cos(2\mathrm{Im}\,\theta)c^2 e^{-2\mathrm{Re}\,\theta}\Delta + c^4}}\| \le \frac{\gamma}{c}\frac{\pi}{2}C(\mathrm{Im}\,\theta), \quad (26)\end{aligned}$$

where we used additionally (5) and Lemma 1. Equation (26) shows that (25) holds for $z = \mathrm{i}\,\eta$, $\eta \in \mathbb{R}$.
Step 2: Proof of (25), *general case.* We have

$$1 - \gamma M_{c;\theta}(z) = 1 - \gamma M_{c;\theta}(0) - \gamma(M_{c;\theta}(z) - M_{c;\theta}(0)) = (1 - \gamma M_{c;\theta}(0))(1 - N(z)),$$

where $N(z) := z\,(1 - \gamma M_{c;\theta}(0))^{-1}\left[V_2(\theta)D_{c,0}(\theta)^{-1}(D_{c,0}(\theta) - z)^{-1}V_1(\theta)\right]$. Using Step 1 and Lemma 2 we see that $N(z)$ is compact and a holomorphic function of z for $z \in \mathbb{C} \setminus \sigma(D_{c,0}(\theta))$. Applying the analytic Fredholm theorem [41, Theorem VI.14] yields that $(1 - N(z))^{-1}$ is a meromorphic function on $\mathbb{C} \setminus \sigma(D_{c,0}(\theta))$ with values in $\mathcal{B}(L^2(\mathbb{R}^3; \mathbb{C}^4))$, whose residues are operators of finite rank. Using Step 1 once more, we see that this also holds for $(1 - e^{-\theta}\gamma M_{c;\theta}(z))^{-1}$. In particular, there is a set $A_{c,\gamma;\theta} \subset \mathbb{C} \setminus \sigma(D_{c,0}(\theta))$ which has no accumulation point in $\mathbb{C} \setminus \sigma(D_{c,0}(\theta))$ such that $z \mapsto R_{c,\gamma;\theta}(z)$ is holomorphic in $\mathbb{C} \setminus (\sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta})$.
Step 3: The mapping $z \mapsto R_{c,\gamma;\theta}(z)\,(D_{c,\gamma}(\theta) - z)\,f$ with $f \in \mathrm{Dom}(D_{c,\gamma}(\theta))$ is holomorphic on $\mathbb{C} \setminus (\sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta})$. Because of Step 1 the operator $R_{c,\gamma;\theta}(z)$ equals the resolvent of $D_{c,\gamma}(\theta)$ for $z = \mathrm{i}\,\eta$, $\eta \in \mathbb{R}$. It follows that $R_{c,\gamma;\theta}(z)\,(D_{c,\gamma}(\theta) - z)\,f = f$ for all $z \in \mathbb{C} \setminus (\sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta})$ and $f \in \mathrm{Dom}(D_{c,\gamma}(\theta))$.
Moreover, it is easy to see that $\mathrm{Ran}\,R_{c,\gamma;\theta}(z) \subset H^{1/2}(\mathbb{R}^3; \mathbb{C}^4)$. Thus, we obtain as before $(g, (D_{c,\gamma}(\theta) - z)\,R_{c,\gamma;\theta}(z)f) = (g, f)$ for all $f \in L^2(\mathbb{R}^3; \mathbb{C}^4)$, $g \in H^{1/2}(\mathbb{R}^3; \mathbb{C}^4)$ and $z \in \mathbb{C} \setminus (\sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta})$. It follows that $\mathrm{Ran}\,R_{c,\gamma;\theta}(z) \subset H^1(\mathbb{R}^3; \mathbb{C}^4)$ and $(D_{c,\gamma}(\theta) - z)R_{c,\gamma;\theta}(z)f = f$ for $f \in L^2(\mathbb{R}^3; \mathbb{C}^4)$ and $z \in \mathbb{C} \setminus (\sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta})$. Summarizing, we find $R_{c,\gamma;\theta}(z) = (D_{c,\gamma}(\theta) - z)^{-1}$ for all $z \in \mathbb{C} \setminus (\sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta})$. In particular, it follows that $\sigma(D_{c,\gamma}(\theta)) \subset \sigma(D_{c,0}(\theta)) \cup A_{c,\gamma;\theta}$.
Let now $z_0 \in A_{c,\gamma;\theta}$. Then the analytic Fredholm theorem implies the existence of $f \in L^2(\mathbb{R}^3; \mathbb{C}^4)$ with $(1 - N(z_0))f = 0$, and thus also $(1 - \gamma M_{c;\theta}(z_0))f = 0$. We proceed as follows: Since $(D_{c,0}(\theta) - z)^{-1}V_1(\theta)$ is bounded, we find $f \in \mathrm{Ran}(V_2(\theta))$, i.e. $f = V_2(\theta)g$ for $g = (D_{c,0}(\theta) - z)^{-1}V_1(\theta)f \in L^2(\mathbb{R}^3; \mathbb{C}^4)$. It follows that $(D_{c,0}(\theta) - z_0)g = \gamma V_1(\theta)f = \gamma V(\theta)g$ in $H^{-1/2}(\mathbb{R}^3; \mathbb{C}^4)$. Rewriting this equality (in the sense of $H^{-1/2}(\mathbb{R}^3; \mathbb{C}^4)$) we find $-\mathrm{i}\,ce^{-\theta}\boldsymbol{\alpha} \cdot \nabla g - \beta c^2 g - \gamma V(\theta)g = z_0 g$. Since the r.h.s. of this equality is a (regular distribution generated by a) function in $L^2(\mathbb{R}^3; \mathbb{C}^4)$, the l.h.s. is. This implies that $g \in H^1(\mathbb{R}^3; \mathbb{C}^4) = \mathcal{D}(D_{c,\gamma}(\theta))$ by Remark 2, i.e. $z_0 \in \sigma(D_{c,\gamma}(\theta))$ which in turn proves $\sigma(D_{c,\gamma}(\theta)) \cap (\mathbb{C} \setminus \sigma(D_{c,0}(\theta))) = A_{c,\gamma;\theta}$.

Step 4: It remains to show that $\sigma(D_{c,\gamma}(\theta)) \cap \sigma(D_{c,0}(\theta)) = \sigma(D_{c,0}(\theta))$ holds. To show this, we pick $E \in \sigma(D_{c,0}(\theta))$ and $p \in \mathbb{R}^3$ with $E = E_c(p;\theta)$ in order to construct a suitable Weyl sequence. Let us define $\psi_{p,c;\theta} \in C^\infty(\mathbb{R}^3;\mathbb{C}^4)$ by

$$\psi_{p,c;\theta}(x) := N_c(p;\theta)^{-1}(c^2 + E_c(p;\theta)\xi, ce^{-\theta}\boldsymbol{\sigma}\cdot p\xi)^T e^{-ipx} \tag{27}$$

with $\xi = (1,0)^T$. Equations (7) and (23) imply

$$(-ic\boldsymbol{\alpha}\cdot\nabla + \beta c^2)\psi_{p,c;\theta}(x) = E_c(p;\theta)\psi_{p,c;\theta}(x). \tag{28}$$

We pick a function $0 \neq \phi \in C_0^\infty(\mathbb{R}^3)$ with $\operatorname{supp}\phi \subset B_1(0)$ and set for $n \in \mathbb{N}$ $\phi_n(x) := \phi(\frac{1}{n}x - ne_1)$ with $e_1 = (1,0,0)^T$ as well as $f_n := \phi_n\psi_{p,c;\theta}$. Obviously, we have $f_n \in \mathrm{Dom}(D_{c,\gamma}(\theta))$. First, we calculate

$$\|f_n\| \geq (1 + C_{\mathrm{FW}})^{-1/2}\|\phi_n\| = n^{3/2}(1 + C_{\mathrm{FW}})^{-1/2}\|\phi\|, \tag{29}$$

where we used the definition (27) of $\psi_{p,c;\theta}$, Equation (7), Equation (11), Equation (8) and the identity $\int \mathrm{d}x\, \phi_n(x)^2 = \int \mathrm{d}x\, \phi(\frac{1}{n}x - ne_1) = n^3 \int \mathrm{d}x\, \phi(x)^2$. Furthermore, we find for $n \geq 2$

$$\begin{aligned}\|V_C f_n\|^2 =& \int \mathrm{d}x\, \frac{1}{|x|^2}\phi_n(x)^2 \|\psi_{p,c;\theta}(0)\|^2 \\ \leq& (1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|)\frac{4}{n^4}\int \mathrm{d}x\, \phi_n(x)^2 \frac{4(1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|)}{n^4} n^3\|\phi\|^2,\end{aligned} \tag{30}$$

since $\operatorname{supp}\phi_n \subset B_n(n^2 e_1)$ and $\|\psi_{p,c;\theta}(0)\| \leq \sqrt{1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|}$ because of Formula (9). Moreover, we obtain

$$\|(c\boldsymbol{\alpha}\cdot\nabla\phi_n)\psi_{p,c;\theta}(\cdot)\| \leq \frac{c\sqrt{1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|}}{n} n^{3/2}\|\nabla\phi\|. \tag{31}$$

Formulas (28) through (31) imply

$$\frac{\|(D_{c,\gamma}(\theta) - E_c(p;\theta))f_n\|}{\|f_n\|} \leq \sqrt{1 + C_{\mathrm{FW}}|\operatorname{Im}\theta|}\,\frac{\frac{2n^{3/2}}{n^2}\|\phi\| + \frac{cn^{3/2}}{n}\|\nabla\phi\|}{\frac{n^{3/2}}{\sqrt{1+C_{\mathrm{FW}}}}\|\phi\|} \underset{n\to\infty}{\longrightarrow} 0.$$

Thus $D_{c,\gamma}(\theta) - E_c(p;\theta)$ does not have a bounded inverse and $E_c(p;\theta) \in \sigma(D_{c,\gamma}(\theta))$.
Step 5: The proof of $A_{c,\gamma;\theta} \cap (\mathbb{C}\setminus B_{c;\theta}) = \sigma_{\mathrm{disc}}(D_{c,\gamma}(0))$ is a standard argument, which uses the dilation analyticity of the operators $D_{c,\gamma}(\theta)$ (see [43, Chapter XII.6] or [46]). The same holds for the claim on the accumulation points. □

Remark 3. *Note that for $V = V_C$ the set of resonances is empty. This follows similarly as for the Schrödinger case (see [8]): If there was a resonance, then $D_{c,\gamma}(\pi)$ would have a non-real eigenvalue.*

5 Spectral Projections

In this section we extend the notion of positive and negative spectral projections to dilated Dirac operators. We define for $p \in \mathbb{R}^3$ the matrices $\Lambda_{c,0}^{(\pm)}(p;\theta) := \frac{1}{2}(\mathbf{1} \pm \frac{cp\cdot\boldsymbol{\alpha}+c^2\beta}{E_c(p;\theta)})$. A calculation shows that $\Lambda_{c,0}^{(\pm)}(p;\theta)^2 = \Lambda_{c,0}^{(\pm)}(p;\theta)$ and $\Lambda_{c,0}^{(\pm)}(p;\theta)D_{c,0}(p;\theta) = \pm E_c(p;\theta)\Lambda_{c,0}^{(\pm)}(p;\theta)$. Moreover, one verifies the identity $\Lambda_{c,0}^{(\pm)}(p;\theta) = \frac{1}{2} \pm \frac{1}{2\pi} \lim_{R\to\infty} \int_{-R}^{R} \mathrm{d}\eta \, \frac{1}{D_{c,0}(p;\theta)-\mathrm{i}\,\eta}$. These observations motivate the following definition for the dilated interacting operators:

$$\Lambda_{c,\gamma}^{(\pm)}(\theta) := \frac{1}{2} \pm \frac{1}{2\pi} \operatorname*{s-lim}_{R\to\infty} \int_{-R}^{R} \mathrm{d}\eta \, \frac{1}{D_{c,\gamma}(\theta)-\mathrm{i}\,\eta} \tag{32}$$

It is well known [33, Chapter VI-5.2, Lemma 5.6] that Equation (32) yields the positive and negative spectral projections for real θ. Note that similar formulas for not necessarily self-adjoint operators are known (see [16, Chapter VX]). These authors use a different definition for the spectral projections, however. First, we show in Theorem 4 that these operators are well defined and bounded projections even if $\theta \notin \mathbb{R}$. We need the following technical lemma:

Lemma 3. *Let $\theta \in S_{\pi/4}$. Then for all $\eta \in \mathbb{R}$*

$$\|\frac{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\| \le C_1(\operatorname{Im}\theta), \tag{33}$$

where $C_1(\operatorname{Im}\theta)$ is defined in (24).

Proof. We prove the estimate

$$\begin{aligned} \|\frac{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}{\sqrt{-e^{-2\theta}c^2\Delta+c^4}\beta-\mathrm{i}\,\eta}\| &\le \|\frac{|D_{c,0}(\operatorname{Re}\theta)|}{\sqrt{-e^{-2\theta}c^2\Delta+c^4}\beta-\mathrm{i}\,\eta}\| \\ +\|\frac{\eta}{\sqrt{-e^{-2\theta}c^2\Delta+c^4}\beta-\mathrm{i}\,\eta}\| &\le \frac{1}{\sqrt{\cos(2\operatorname{Im}\theta)}}+\frac{1}{\cos\operatorname{Im}\theta}. \end{aligned} \tag{34}$$

We estimate the first summand using inequality (5) and Lemma 1. For the second summand we restrict ourselves to the case $\operatorname{Im}\theta < 0$. The proof for $\operatorname{Im}\theta > 0$ works analogously, and (33) holds obviously if $\operatorname{Im}\theta = 0$. Moreover, it suffices to consider $\operatorname{Re}\theta = 0$. We investigate the term $|\sqrt{e^{-2\theta}c^2p^2+c^4}-\mathrm{i}\,\eta|$. For $\eta > 0$ the inequality $\operatorname{Im}\sqrt{e^{-2\theta}c^2p^2+c^4} < 0$ yields $|-\sqrt{e^{-2\theta}c^2p^2+c^4}+\mathrm{i}\,\eta| \ge |\eta|$. For $\eta < 0$ the inequality $\operatorname{Im}\sqrt{c^2p^2+e^{+2\theta}c^4} > 0$ implies $|\sqrt{c^2p^2+e^{+2\theta}c^4}-ie^{+\theta}\eta| \ge -\cos(\operatorname{Im}\theta)\eta = \cos(\operatorname{Im}\theta)|\eta|$, which proves (34). The claim follows using Theorem 1. ☐

Theorem 4. *Let $\theta \in S_{\min\{\pi/4,\Theta\}}$ and $\frac{2\gamma}{c}C(\operatorname{Im}\theta) < 1$. Suppose that (H1) holds. Then the following statements hold: $\Lambda_{c,\gamma}^{(\pm)}(\theta) \in \mathcal{B}(L^2(\mathbb{R}^3;\mathbb{C}^4))$, $\Lambda_{c,\gamma}^{(\pm)}(\theta) = \Lambda_{c,\gamma}^{(\pm)}(\theta)^2$ and $\Lambda_{c,\gamma}^{(+)}(\theta)+\Lambda_{c,\gamma}^{(-)}(\theta) = \mathbf{1}$. The operators $\Lambda_{c,\gamma}^{(\pm)}(\theta)$ are bounded holomorphic families in θ for $\theta \in M_{\gamma/c}$.*

Proof. The proof is inspired by similar estimates in [47].
Step 1: The resolvent equation (25) and the estimate (26) yield the convergence of the series

$$\begin{aligned}(D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}-(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}=\gamma\sum_{n=1}^{\infty}(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}V_1(\theta)\\ \times[\gamma V_2(\theta)(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}V_1(\theta)]^{n-1}V_2(\theta)(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}\end{aligned}\tag{35}$$

in norm.
Step 2: We show that the expression

$$\lim_{R\to\infty}\int_{-R}^{R}\mathrm{d}\eta\,\big(f,[\frac{1}{D_{c,\gamma}(\theta)-\mathrm{i}\,\eta}-\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}]g\big),\quad f,g\in L^2(\mathbb{R}^3;\mathbb{C}^4)\tag{36}$$

defines a bounded operator on $L^2(\mathbb{R}^3;\mathbb{C}^4)$. In order to achieve this, we estimate

$$\begin{aligned}&\big|\big(f,\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}V_1(\theta)\big[\gamma V_2(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}V_1(\theta)\big]^{n-1}V_2(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}g\big)\big|\\ &\quad\le\frac{\pi}{2}\Big\|\frac{|\nabla|^{1/2}}{D_{c,0}(\bar\theta)+\mathrm{i}\,\eta}f\Big\|\Big\|\frac{|\nabla|^{1/2}}{D_{c,0}(\theta)-\mathrm{i}\,\eta}g\Big\|\big(\frac{\gamma}{c}\frac{\pi}{2}C(\operatorname{Im}\theta)\big)^{n-1}\le\frac{\pi}{2ce^{-\operatorname{Re}\theta}}\\ &\quad\times\Big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|+\mathrm{i}\,\eta}f\Big\|\Big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}g\Big\|C_1(\operatorname{Im}\theta)^2\big(\frac{\gamma}{c}\frac{\pi}{2}C(\operatorname{Im}\theta)\big)^{n-1},\end{aligned}$$

where we used (26) in the first estimate and Lemma 3 in the second estimate. $C(\operatorname{Im}\theta)$ and $C_1(\operatorname{Im}\theta)$ were defined in (24). As in [47, Proof of Lemma 1] we obtain $\int_{-\infty}^{\infty}\mathrm{d}\eta\,\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|+\mathrm{i}\,\eta}f\big\|\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}g\big\|\le\pi\|f\|\|g\|$ and thus

$$\begin{aligned}\int_{-\infty}^{\infty}\mathrm{d}\eta\,\big|\big(f,[\frac{1}{D_{c,\gamma}(\theta)-\mathrm{i}\,\eta}-\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}]g\big)\big|\le\\ \le\pi\frac{\gamma}{c}\frac{\pi}{2}\|f\|\|g\|C_1(\operatorname{Im}\theta)^2\frac{1}{1-\big(\frac{\gamma}{c}\frac{\pi}{2}C(\operatorname{Im}\theta)\big)}\end{aligned}\tag{37}$$

Step 3: The expressions

$$\Big(f,\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}V_1(\theta)\big[\gamma V_2(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}V_1(\theta)\big]^{n-1}V_2(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}g\Big)$$

are holomorphic functions of $\theta\in S_{\min\{\pi/4,\Theta\}}$. These estimates show the existence of an integrable and summable majorant, independent of θ for $\theta\in M_{\gamma/c}$. Thus, the operator in Equation (36) is a holomorphic function of θ [33, Chapter VII-1.1], and the identity $\Lambda^{(+)}_{c,\gamma}(\theta)=\Lambda^{(+)}_{c,\gamma}(\theta)^2$, which is obviously true for $\theta\in\mathbb{R}$, extends to $\theta\in M_{\gamma/c}$, i.e. $\Lambda^{(+)}_{c,\gamma}(\theta)$ is a projection.

Step 4: We show that the limit exists as a strong limit and estimate for $g \in H^{1/2}(\mathbb{R}^3;\mathbb{C}^4)$ as follows:

$$\begin{aligned}
&\big|\big(f, \frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big[\gamma V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big]^{n-1}V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}g\big)\big| \\
&\quad\le \frac{2}{ce^{-\operatorname{Re}\theta}}\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|+\mathrm{i}\,\eta}\big\|\|f\|\big\|\frac{1}{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}\big\|\big\||D_{c,0}(\operatorname{Re}\theta)|^{1/2}g\big\| \\
&\qquad\qquad\times C_1(\operatorname{Im}\theta)^2\big(\frac{2\gamma}{c}C(\operatorname{Im}\theta)\big)^{n-1}
\end{aligned}$$

Here we estimated the expression in the square brackets similarly to (26), but used Hardy's inequality instead of Kato's inequality. Moreover, we used the estimate (33) twice. Since $\sigma(D_{c,0}(\operatorname{Re}\theta)) = (-\infty, c^2] \cup [c^2,\infty)$, we have

$$\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|+\mathrm{i}\,\eta}\big\| = \sup_{|\lambda|\ge c^2}\frac{\sqrt{|\lambda|}}{\sqrt{\lambda^2+\eta^2}} \le \min\{\frac{1}{c},\frac{1}{\sqrt{|\eta|}}\}.$$

This estimate shows that the convergence in formula (36) is uniform in $f \in L^2(\mathbb{R}^3;\mathbb{C}^4)$, which implies the strong convergence [33, Theorem III.1.32 and Lemma III.3.5], since $H^{1/2}(\mathbb{R}^3;\mathbb{C}^4)$ is dense in $L^2(\mathbb{R}^3;\mathbb{C}^4)$. $\square$

Obviously, the identity $\Lambda^{(+)}_{c,\gamma}(\theta) + \Lambda^{(-)}_{c,\gamma}(\theta) = \mathbf{1}$ holds. We set $\mathcal{H}^{(\pm)}_{c,\gamma}(\theta) := \Lambda^{(\pm)}_{c,\gamma}(\theta)L^2(\mathbb{R}^3;\mathbb{C}^4)$ and find $L^2(\mathbb{R}^3;\mathbb{C}^4) = \mathcal{H}^{(+)}_{c,\gamma}(\theta) \dot{+} \mathcal{H}^{(-)}_{c,\gamma}(\theta)$, wehre $\dot{+}$ denotes the direct sum. We call the $\Lambda^{(\pm)}_{c,\gamma}(\theta)$ positive and negative spectral projections and $\mathcal{H}^{(\pm)}_{c,\gamma}(\theta)$ positive and negative spectral subspaces, respectively. This is justified because of Theorem 5.
The following corollary generalizes [47, Lemma 1] to dilated spectral projections.

Corollary 1. *Let $\theta \in S_{\min\{\pi/4,\Theta\}}$ and suppose that (H1) holds. Then there exists a constant $C_{\mathrm{NR}} > 0$ such that for $\frac{2\gamma}{c}C(\operatorname{Im}\theta) < 1$ the estimate*

$$\|\Lambda^{(\pm)}_{c,\gamma}(\theta) - \Lambda^{(\pm)}_{c,0}(\theta)\| \le C_{\mathrm{NR}}\frac{\gamma}{c}$$

holds.

Proof. This follows directly from Equation (37) in the proof of Theorem 4. $\square$

The next theorem shows that the spaces $\mathcal{H}^{(\pm)}_{c,\gamma}(\theta)$ are invariant under $D_{c,\gamma}(\theta)$ and describes the spectrum of the restriction of the operator to these spaces. If a part of the spectrum is contained in a Jordan curve, analogous statements can be found in [33, Theorem III-6.17]. The following theorem describes a more general situation, but the essential elements of the proof of [33, Theorem III-6.17] can be adapted.
For a closed operator A we denote its resolvent set by $\rho(A)$.

Theorem 5. *Let* $\theta \in S_{\min\{\pi/4,\Theta\}}$ *and* $\frac{2\gamma}{c}C(\operatorname{Im}\theta) < 1$. *Suppose that (H1) holds. Then the identity*

$$\Lambda_{c,\gamma}^{(\pm)}(\theta)(D_{c,\gamma}(\theta)-z)^{-1} = (D_{c,\gamma}(\theta)-z)^{-1}\Lambda_{c,\gamma}^{(\pm)}(\theta) \tag{38}$$

holds for all $z \in \rho(D_{c,\gamma}(\theta))$. *The subspaces* $\operatorname{Ran}\Lambda_{c,\gamma}^{(+)}(\theta)$ *and* $\operatorname{Ran}\Lambda_{c,\gamma}^{(-)}(\theta)$ *are invariant subspaces for* $D_{c,\gamma}(\theta)$. *In particular,*

$$\sigma(D_{c,\gamma}(\theta)|_{\operatorname{Ran}\Lambda_{c,\gamma}^{(+)}(\theta)}) = \sigma(D_{c,\gamma}(\theta)) \cap \{z \in \mathbb{C}|\operatorname{Re} z > 0\} \tag{39}$$

and

$$\sigma(D_{c,\gamma}(\theta)|_{\operatorname{Ran}\Lambda_{c,\gamma}^{(-)}(\theta)}) = \sigma(D_{c,\gamma}(\theta)) \cap \{z \in \mathbb{C}|\operatorname{Re} z < 0\} \tag{40}$$

hold.

Proof. Obviously, for all $z \notin \sigma(D_{c,\gamma}(\theta))$, all $\eta \in \mathbb{R}$ and all $f \in L^2(\mathbb{R}^3;\mathbb{C}^4)$ the equation $(D_{c,\gamma}(\theta)-z)^{-1}(D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}f = (D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}(D_{c,\gamma}(\theta)-z)^{-1}f$ is true. This immediately implies

$$\begin{aligned}(D_{c,\gamma}(\theta)-z)^{-1}\lim_{R\to\infty}\int_{-R}^{R}\mathrm{d}\eta\;(D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}f &= \\ = \lim_{R\to\infty}\int_{-R}^{R}\mathrm{d}\eta\;(D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}&(D_{c,\gamma}(\theta)-z)^{-1}f\end{aligned}$$

and thus (38). It follows that [33, Chapter III-5.6 and Theorem III.6.5] $(D_{c,\gamma}(\theta)-z)^{-1}\operatorname{Ran}\Lambda_{c,\gamma}^{(\pm)}(\theta) \subset \operatorname{Ran}\Lambda_{c,\gamma}^{(\pm)}(\theta)$ and $\Lambda_{c,\gamma}^{(\pm)}(\theta)\operatorname{Dom}(D_{c,\gamma}(\theta)) \subset \operatorname{Dom}(D_{c,\gamma}(\theta))$ as well as $D_{c,\gamma}(\theta)\mathcal{H}_{c,\gamma}^{(\pm)}(\theta) \subset \mathcal{H}_{c,\gamma}^{(\pm)}(\theta)$. We define the operators $D_{c,\gamma}^{(\pm)}(\theta) := D_{c,\gamma}(\theta)|_{\mathcal{H}_{c,\gamma}^{(\pm)}(\theta)}$ and (for $z \notin \sigma(D_{c,\gamma}(\theta))$ at the moment) the resolvents $R_{c,\gamma;\theta}^{(\pm)}(z) := (D_{c,\gamma}^{(\pm)}(\theta)-z)^{-1} = (D_{c,\gamma}(\theta)-z)^{-1}|_{\mathcal{H}_{c,\gamma}^{(\pm)}(\theta)}$. In particular, $\sigma(D_{c,\gamma}^{(\pm)}(\theta)) \subset \sigma(D_{c,\gamma}(\theta))$.

On the other side, we have $f \in \mathcal{H}_{c,\gamma}^{(\pm)}(\theta)$ and $z \notin \sigma(D_{c,\gamma}(\theta))$ $R_{c,\gamma;\theta}^{(\pm)}(z)f = (D_{c,\gamma}(\theta)-z)^{-1}f = (D_{c,\gamma}(\theta)-z)^{-1}\Lambda_{c,\gamma}^{(\pm)}(\theta)f$. Using the first resolvent identity, we find for $z \in \mathbb{C}$ with $\operatorname{Re} z < 0$ respectively $\operatorname{Re} z > 0$

$$(D_{c,\gamma}(\theta)-z)^{-1}\Lambda_{c,\gamma}^{(\pm)}(\theta)f = -\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\eta\frac{1}{z-\mathrm{i}\,\eta}(D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}f, \tag{41}$$

since for $z \in \mathbb{C}$ with $\operatorname{Re} z < 0$ respectively $\operatorname{Re} z > 0$ the residue theorem implies $\lim_{R\to\infty}\int_{-R}^{R}\mathrm{d}\eta\,\frac{1}{z-\mathrm{i}\,\eta} = \lim_{R\to\infty}\int_{-R}^{R}\mathrm{d}\eta\,\frac{z}{z^2+\eta^2} = \mp\pi$.

The r.h.s. of equation (41) is holomorphic in $z \notin \mathrm{i}\,\mathbb{R}$. Thus, $R_{c,\gamma;\theta}^{(+)}(z)$ has a holomorphic continuation to $\{z \in \mathbb{C}|\operatorname{Re} z < 0\}$, and $R_{c,\gamma;\theta}^{(-)}(z)$ has a holomorphic continuation to $\{z \in \mathbb{C}|\operatorname{Re} z > 0\}$. The holomorphicity of the resolvent implies $\{z \in \mathbb{C}|\operatorname{Re} z < 0\} \subset \rho(D_{c,\gamma}^{(+)}(\theta))$ and $\{z \in \mathbb{C}|\operatorname{Re} z >$

$0\} \subset \rho(D_{c,\gamma}^{(-)}(\theta))$. This proves $\sigma(D_{c,\gamma}^{(-)}(\theta)) \subset \{z \in \mathbb{C} | \mathrm{Re}\, z < 0\}$ and $\sigma(D_{c,\gamma}^{(+)}(\theta)) \subset \{z \in \mathbb{C} | \mathrm{Re}\, z > 0\}$. On the other side, $z \in \sigma(D_{c,\gamma}(\theta))$ cannot fulfill both $z \in \rho(D_{c,\gamma}^{(-)}(\theta))$ and $z \in \rho(D_{c,\gamma}^{(+)}(\theta))$, because otherwise the identity $(D_{c,\gamma}(\theta)-z)^{-1} = (D_{c,\gamma}^{(+)}(\theta)-z)^{-1}\Lambda_{c,\gamma}^{(+)}(\theta)+(D_{c,\gamma}^{(-)}(\theta)-z)^{-1}\Lambda_{c,\gamma}^{(-)}(\theta)$ would imply the contradiction $z \in \rho(D_{c,\gamma}(\theta))$. This shows (39) and (40). □

Next, we need spectral projections for the eigenvalues: We define for all $n \geq 1$ (and $n \leq N_{\max}$ if there only finitely many eigenvalues) the spectral projections

$$P_n(c,\gamma;\theta) := -\frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n(c,\gamma)} \frac{1}{D_{c,\gamma}(\theta) - z} \mathrm{d}z\,, \tag{42}$$

where z runs through $\Gamma_n(c,\gamma)$ in the positive sense. $\Gamma_n(c,\gamma)$ is chosen such that for all $1 \leq l \leq N_n$ the eigenvalues $\tilde{E}_{n,l}(c,\gamma)$ are located within the contour, but no other elements of the spectrum $D_{c,\gamma}(\theta)$.
For later, we need spectral projections for the fine structure components. We set for $n \geq 1$ and $1 \leq l \leq N_n$

$$P_{n,l}(c,\gamma;\theta) := -\frac{1}{2\pi \mathrm{i}} \int_{\Gamma_{n,l}(c,\gamma)} \frac{1}{D_{c,\gamma}(\theta) - z} \mathrm{d}z\,, \tag{43}$$

where z runs through $\Gamma_{n,l}(c,\gamma)$ in the positive sense, and $\Gamma_{n,l}(c,\gamma)$ is chosen such that only the eigenvalue $\tilde{E}_{n,l}(c,\gamma)$ lies within the contour. We denote the corresponding normed eigenfunctions by $\phi_{n,l}(c,\gamma;\theta)$.

6 Transformation Functions

We need transformation functions between the spectral subspaces of dilated and not dilated operators for the resolvent estimate in Section 7 and in order to establish the dilation analyticity of a relativistic Pauli-Fierz model in [30].
Another example for a transformation function is the Douglas-Kroll transformation, which was investigated by Siedentop and Stockmeyer [47] (see also Huber and Stockmeyer [31]). Contrary to the situation there, our spectral projections are not self-adjoint and thus the transformation function is a non-unitary similarity transformation. The estimates in this section can be used to generalize the Douglas-Kroll transformation to complex dilated operators.
In order to prove the existence of the transformation function, we need norm estimates on the difference between the spectral projections.

Lemma 4. *Let $\theta \in S_{\min\{\pi/4,\Theta\}}$. Suppose that (H1) and (H2) hold. Then the following statements hold:*

a) *There is a constant $C_{\mathrm{DL}} > 0$ (independent of c, γ and θ) such that for $\frac{2\gamma}{c}C(\mathrm{Im}\,\theta) < 1$ the estimate*

$$\|\Lambda_{c,\gamma}^{(\pm)}(0) - \Lambda_{c,\gamma}^{(\pm)}(\theta)\| \leq C_{\mathrm{DL}}|\theta| \tag{44}$$

holds. The operator $|D_{c,0}(0)|^{1/2}[\Lambda^{(\pm)}_{c,\gamma}(0)-\Lambda^{(\pm)}_{c,\gamma}(\theta)]|D_{c,0}(0)|^{-1/2}$ *is a holomorphic function of* $\theta \in M_{\gamma/c}$.

b) Let moreover $0 < q < 1$*. Then there is a constant* $C_{\mathrm{DLS}} > 0$ *(independent of* c*,* γ *and* θ*) such that for* $\frac{2\gamma}{c}C(\operatorname{Im}\theta) < q$ *the estimate*

$$\||D_{c,0}(0)|^{1/2}[\Lambda^{(\pm)}_{c,\gamma}(0) - \Lambda^{(\pm)}_{c,\gamma}(\theta)]|D_{c,0}(0)|^{-1/2}\| \le C_{\mathrm{DLS}}|\theta| \tag{45}$$

holds.

Proof. We adapt method which was used by Siedentop and Stockmeyer [47] and by Griesemer, Lewis and Siedentop [19] for other choices of projections. We start with the difference of resolvents

$$\begin{aligned}(D_{c,0}(\theta) - \mathrm{i}\,\eta)^{-1} &- (D_{c,0}(0) - \mathrm{i}\,\eta)^{-1}\\ &= ic[e^{-\theta} - 1](D_{c,0}(\theta) - \mathrm{i}\,\eta)^{-1}\boldsymbol{\alpha}\cdot\nabla(D_{c,0}(0) - \mathrm{i}\,\eta)^{-1}\end{aligned} \tag{46}$$

and note that $|e^{-\theta} - 1| \le B|\theta|$ holds with $B = e^{\pi/4}$ for all $|\theta| \le \pi/4$.
Step 1: Proof for the free projections. Equation (46) it and Lemma 3 imply that

$$\begin{aligned}&|(f, [(D_{c,0}(\theta) - \mathrm{i}\,\eta)^{-1} - (D_{c,0}(0) - \mathrm{i}\,\eta)^{-1}]g)|\\ \le& B|\theta|\||D_{c,0}(\operatorname{Re}\theta)|^{1/2}(|D_{c,0}(\operatorname{Re}\theta)| + \mathrm{i}\,\eta)^{-1}f\|\||D_{c,0}(0)|^{1/2}(D_{c,0}(0) - \mathrm{i}\,\eta)^{-1}g\|\\ &\times \||D_{c,0}(\operatorname{Re}\theta)|^{-1/2}c\boldsymbol{\alpha}\cdot\nabla|D_{c,0}(0)|^{-1/2}\|\|\frac{|D_{c,0}(\operatorname{Re}\theta)| - \mathrm{i}\,\eta}{D_{c,0}(\theta) - \mathrm{i}\,\eta}\|\\ \le& \frac{B|\theta|}{e^{-\operatorname{Re}\theta/2}}C_1(\operatorname{Im}\theta)\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)| + \mathrm{i}\,\eta}f\|\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0) - \mathrm{i}\,\eta}g\|,\end{aligned}$$

where we used the estimate $\|c|\nabla||D_{c,0}(\operatorname{Re}\theta)|^{-1}\| \le 1/e^{-\operatorname{Re}\theta}$.
This proves (cf. [47, Proof of Lemma 1] and proof of Corollary 1) $\|\Lambda^{(\pm)}_{c,0}(0) - \Lambda^{(\pm)}_{c,0}(\theta)\| \le \tilde{C}_{\mathrm{DL}}|\theta|$ with a $\tilde{C}_{\mathrm{DL}} > 0$ and analogously $\||D_{c,0}(0)|^{1/2}[\Lambda^{(\pm)}_{c,0}(0) - \Lambda^{(\pm)}_{c,0}(\theta)]|D_{c,0}(0)|^{-1/2}\| \le \tilde{C}_{\mathrm{DL}}|\theta|$, since $|D_{c,0}(0)|^{1/2}$ commutes with all operators in (46).
Step 2: Proof of (44). We write

$$\begin{aligned}&\|\big[V_2(\theta)\frac{1}{D_{c,0}(\theta) - \mathrm{i}\,\eta}V_1(\theta)\big] - \big[V_2(0)\frac{1}{D_{c,0}(0) - \mathrm{i}\,\eta}V_1(0)\big]\| \qquad (47)\\ \le&\|\big[V_1(\theta)\frac{e^{-\theta}}{D_{c,0}(\theta) - \mathrm{i}\,\eta}\chi_\theta V_2(\theta)\big] - \big[V_1(\theta)\frac{e^{-\theta}}{D_{c,0}(0) - \mathrm{i}\,\eta}\chi_\theta V_2(\theta)\big]\|\\ &+ \|\big[V_C^{1/2}\frac{1}{D_{c,0}(0) - \mathrm{i}\,\eta}(\chi_\theta e^{-\theta} - 1)V_C^{1/2}\big]\| \le \frac{B|\theta|\pi}{2c}(C(\operatorname{Im}\theta) + 1 + \tilde{C}),\end{aligned}$$

where we estimated the second summand by $B(1+\tilde{C})|\theta|\pi/(2c)$ from above, and the second summand – similarly as in (26) – according to

$$\begin{aligned}&\big\|(e^{-\theta}-1)\big[V_2(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}c\boldsymbol{\alpha}\cdot\nabla\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}V_1(\theta)\big]\big\|\leq\\&\qquad\leq\frac{B|\theta|\pi}{2c}\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big\|\big\|\frac{|D_{c,0}(0)|}{D_{c,0}(0)-\mathrm{i}\,\eta}\big\|\leq\frac{B|\theta|\pi}{2c}C(\operatorname{Im}\theta).\end{aligned}$$

In the same way we obtain

$$\begin{aligned}&\big\|\big[e^{-\theta/2}V_C^{1/2}(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}-V_C^{1/2}(D_{c,0}(0)-\mathrm{i}\,\eta)^{-1}\big]g\big\| &(48)\\&\leq|e^{-\theta/2}-e^{\theta/2}|\big\|e^{-\theta}V_C^{1/2}(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}c\boldsymbol{\alpha}\cdot\nabla(D_{c,0}(0)-\mathrm{i}\,\eta)^{-1}g\big\|\\&+|e^{-\theta/2}-1|\big\|V_C^{1/2}\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}g\big\|\leq B|\theta|\sqrt{\frac{\pi}{2c}}\big(C(\operatorname{Im}\theta)+1/2\big)\big\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)-\mathrm{i}\,\eta}g\big\|.\end{aligned}$$

Lemma 3 implies

$$\big\|V_C^{1/2}\frac{e^{-\theta/2}}{D_{c,0}(\theta)-\mathrm{i}\,\eta}g\big\|\leq C_1(\operatorname{Im}\theta)\sqrt{\frac{\pi}{2c}}\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}g\big\| \tag{49}$$

and (see Formula (26))

$$\big\|V_C^{1/2}e^{-\theta}(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}V_C^{1/2}\big\|\leq\frac{\pi C(\operatorname{Im}\theta)}{2c}. \tag{50}$$

Formulas (47) through (50) show

$$\begin{aligned}&\big|\gamma^n\big(f,\frac{e^{-\theta/2}}{D_{c,0}(\theta)-\mathrm{i}\,\eta}V_1(\theta)\big[V_2(\theta)\frac{e^{-\theta}}{D_{c,0}(\theta)-\mathrm{i}\,\eta}V_1(\theta)\big]^{n-1}V_2(\theta)\frac{e^{-\theta/2}}{D_{c,0}(\theta)-\mathrm{i}\,\eta}g\big)-\\&-\gamma^n\big(f,\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}V_1(0)\big[V_2(0)\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}V_1(0)\big]^{n-1}V_2(0)\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}g\big)\big|\\&\qquad\leq B|\theta|\big(\frac{\pi\gamma C(\operatorname{Im}\theta)}{2c}\big)^{n-1}\big(\frac{\pi\gamma C_1(\operatorname{Im}\theta)}{2c}\big)\big(C(\operatorname{Im}\theta)+1+\tilde{C}\big)\\&\qquad\times\big\|\frac{|D_{c,0}(0)|^{1/2}}{|D_{c,0}(0)|-\mathrm{i}\,\eta}f\big\|\big[n\big\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|+\mathrm{i}\,\eta}g\big\|+\big\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)+\mathrm{i}\,\eta}g\big\|\big],\end{aligned}$$

which implies (44).
Step 3: Proof of (45). We use the expansion

$$\begin{aligned}&(D_{c,\gamma}(\theta)-\mathrm{i}\,\eta)^{-1}-(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}\\&\qquad=\sum_{n=1}^{\infty}\gamma^n\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big[V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big]^{n-1}V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\qquad(51)\end{aligned}$$

and start with the necessary estimates on the differences of the resolvents: Using Hardy's inequality, we obtain as in (26)

$$\begin{aligned}&\|[V(\theta)(D_{c,0}(\theta)-\mathrm{i}\,\eta)^{-1}-V(0)(D_{c,0}(0)-\mathrm{i}\,\eta)^{-1}]|D_{c,0}(0)|^{-1/2}g\| \\ &\leq \frac{2B|\theta|}{c}(C(\operatorname{Im}\theta)+1+\tilde{C})\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)-\mathrm{i}\,\eta}g\|,\end{aligned} \tag{52}$$

and we find analogously

$$\|V_C\big[\chi_\theta\frac{e^{-\theta}}{D_{c,0}(\theta)-\mathrm{i}\,\eta}-\chi\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}\big]\|\leq\frac{2B|\theta|}{c}(C(\operatorname{Im}\theta)+1+\tilde{C}) \tag{53}$$

as well as

$$\begin{aligned}&\|[(D_{c,0}(\bar\theta)+\mathrm{i}\,\eta)^{-1}-(D_{c,0}(0)+\mathrm{i}\,\eta)^{-1}]|D_{c,0}(0)|^{1/2}f\| \\ &\leq\|[e^{\bar\theta}-1]\frac{e^{-\bar\theta}}{D_{c,0}(\bar\theta)+\mathrm{i}\,\eta}c\boldsymbol{\alpha}\cdot\nabla\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)+\mathrm{i}\,\eta}f\|\leq 2B|\theta|C(\operatorname{Im}\theta)\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)+\mathrm{i}\,\eta}f\|.\end{aligned} \tag{54}$$

For the terms with the resolvents we use Lemma 3 and Lemma 1 to estimate

$$\begin{aligned}&\|V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}|D_{c,0}(0)|^{-1/2}g\|\leq\frac{2}{c}\|\frac{|D_{c,0}(\operatorname{Re}\theta)|}{D_{c,0}(\theta)-\mathrm{i}\,\eta}|D_{c,0}(0)|^{-1/2}g\| \\ &\leq\frac{2C_1(\operatorname{Im}\theta)e^{\pi/8}}{c}\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{|D_{c,0}(\operatorname{Re}\theta)|-\mathrm{i}\,\eta}g\|,\end{aligned} \tag{55}$$

and (cf. Formula (26))

$$\|V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\|\leq\frac{2}{c}\||D_{c,0}(\operatorname{Re}\theta)|\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\|\leq\frac{2C(\operatorname{Im}\theta)}{c}. \tag{56}$$

Formulas (52) through (56) show

$$\begin{aligned}&\gamma^n\big|\big(f,|D_{c,0}(0)|^{1/2}\Big[\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big[V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big]^{n-1}V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta} \\ &-\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}\big[V\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}\big]^{n-1}V\frac{1}{D_{c,0}(0)-\mathrm{i}\,\eta}\Big]|D_{c,0}(0)|^{-1/2}g\big)\big| \\ &\leq e^{\pi/4}B|\theta|\big(\frac{2\gamma C(\operatorname{Im}\theta)}{c}\big)^{n-1}\big(\frac{2\gamma C_1(\operatorname{Im}\theta)}{c}\big)(C(\operatorname{Im}\theta)+1+\tilde{C})\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)+\mathrm{i}\,\eta}f\| \\ &\times\big[n\|\frac{|D_{c,0}(\operatorname{Re}\theta)|^{1/2}}{D_{c,0}(\operatorname{Re}\theta)-\mathrm{i}\,\eta}g\|+\|\frac{|D_{c,0}(0)|^{1/2}}{D_{c,0}(0)-\mathrm{i}\,\eta}g\|\big],\end{aligned}$$

which in turn proves (45).

Step 4: Holomorphicity. This follows as in the proof of Theorem 4, since $\big(f,|D_{c,0}(0)|^{1/2}\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big[V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}\big]^{n-1}V(\theta)\frac{1}{D_{c,0}(\theta)-\mathrm{i}\,\eta}|D_{c,0}(0)|^{-1/2}g\big)$ are holomorphic functions of θ and the above estimates imply the existence of summable and integrable majorant which does not depend on θ. ☐

Before we turn to the existence of a transformation function in Theorem 6, we need two operator inequalities, one of which was proven in [19]. Since the other inequality can be proven completely analogously, we omit the proof. Let us mention that there exits an improved version of one of these inequalities (see [39]). But since we will be interested in the non-relativistic limit only, it is sufficient to use the original version.

Lemma 5 ([19], Lemma 2). *Suppose that $\vartheta \in \mathbb{R}$ and $\frac{\gamma}{c} < \frac{1}{2}$. Then the operator inequalities*

$$(1 - \frac{2\gamma}{c})|D_{c,0}(\vartheta)\| \leq |D_{c,\gamma}(\vartheta)| \leq (1 + \frac{2\gamma}{c})|D_{c,0}(\vartheta)|$$

hold.

Now we can turn to the transformation function $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)$ defined below. It enables us to consider the operator $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)D^{(\pm)}_{c,\gamma}(\theta)\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1}$ instead of the operator $D^{(\pm)}_{c,\gamma}(\theta)$. This is necessary for technical reasons, since the latter operates on a fixed space (i.e. $\operatorname{Ran}\Lambda^{(+)}_{c,\gamma}(0)$). We will prove in [30] that this operator defines a holomorphic family of operators. Moreover, we will need the transformation function in the proof of the resolvent estimate in Theorem 7.

Theorem 6. *Suppose that $\theta \in S_{\min\{\pi/4,\Theta\}}$, $\frac{2\gamma}{c}C(\operatorname{Im}\theta) < 1$ and $C_{\mathrm{DL}}|\theta| < q$ for some $0 < q < 1$. Suppose moreover that (H1) and (H2) hold. Then the following statements hold:*

a) There is a bounded mapping $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) : L^2(\mathbb{R}^3;\mathbb{C}^4) \to L^2(\mathbb{R}^3;\mathbb{C}^4)$ with the property

$$\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)\Lambda^{(+)}_{c,\gamma}(\theta)\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1} = \Lambda^{(+)}_{c,\gamma}(0) \tag{57}$$

and bounded inverse $\mathcal{V}_{\mathrm{DL}}(c,\gamma;\theta) := \mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1}$. There is a constant $C_{UDL} > 0$, independent of c, γ and θ, such that

$$\|\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) - \mathbf{1}\| \leq C_{UDL}|\theta| \tag{58}$$

holds.

b) Suppose that additionally $C_{\mathrm{DLS}}|\theta| < q$ holds. Then there is a constant C_{UDLS}, independent of c, γ and θ, such that

$$\||D_{c,0}(0)|^{1/2}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)|D_{c,0}(0)|^{-1/2} - \mathbf{1}\| \leq C_{UDLS}|\theta| \tag{59}$$

is true. The same estimates hold for $\mathcal{V}_{\mathrm{DL}}(c,\gamma;\theta)$.

c) The operator $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)$, and for $C_{\mathrm{DLS}}|\theta| < q$ the operator

$$|D_{c,0}(0)|^{1/2}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)|D_{c,0}(0)|^{-1/2}$$

and the operator $|D_{c,0}(0)|^{-1/2}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)|D_{c,0}(0)|^{1/2}$, are holomorphic functions of θ. The same statements hold for $\mathcal{V}_{\mathrm{DL}}(c,\gamma;\theta)$.

Proof. We follow [47, Theorem 1] and [33, Chapter I-4.6.] and define

$$\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) := [\Lambda^{(+)}_{c,\gamma}(0)\Lambda^{(+)}_{c,\gamma}(\theta)+\Lambda^{(-)}_{c,\gamma}(0)\Lambda^{(-)}_{c,\gamma}(\theta)][\mathbf{1}-(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))^2]^{-1/2}. \tag{60}$$

It is easy to see that $(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))^2$ commutes with $\Lambda^{(+)}_{c,\gamma}(\theta)$ and $\Lambda^{(+)}_{c,\gamma}(0)$ and that $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)$ is invertible with inverse $\mathcal{V}_{\mathrm{DL}}(c,\gamma;\theta) := [\Lambda^{(+)}_{c,\gamma}(\theta)\Lambda^{(+)}_{c,\gamma}(0)+ +\Lambda^{(-)}_{c,\gamma}(\theta)\Lambda^{(-)}_{c,\gamma}(0)t][\mathbf{1}-(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))^2]^{-1/2}$, and that Equation (57) holds. Lemma 4 implies that $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)$ is a holomorphic function θ, since $(1-A)^{-1/2}$ has a norm convergent series expansion for bounded operators A with $\|A\|<1$.
Proof of (58)*:* We follow [47, Proof of Lemma 5]. We have $\Lambda^{(+)}_{c,\gamma}(0)\Lambda^{(+)}_{c,\gamma}(\theta)+\Lambda^{(-)}_{c,\gamma}(0)\Lambda^{(-)}_{c,\gamma}(\theta)=\mathbf{1}-\left[\Lambda^{(-)}_{c,\gamma}(0)-\Lambda^{(+)}_{c,\gamma}(0)\right]\left[\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0)\right]$ and thus

$$\begin{aligned}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) := \Big\{\mathbf{1}-\left[\Lambda^{(-)}_{c,\gamma}(0)-\Lambda^{(+)}_{c,\gamma}(0)\right]\left[\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0)\right]\Big\}&\\ \times\left[\mathbf{1}-(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))^2\right]^{-1/2}&.\end{aligned}$$

Using the representation $(1-a^2)^{-1/2}=\frac{1}{\pi}\int_{-1}^{1}\frac{1}{\sqrt{1-y^2}}\frac{1}{1-ya}\mathrm{d}y$ (see [17, Formula 3.197.4]) we obtain

$$\begin{aligned}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) = \Big\{\mathbf{1}-\left[\Lambda^{(-)}_{c,\gamma}(0)-\Lambda^{(+)}_{c,\gamma}(0)\right]\left[\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0)\right]\Big\}&\\ \times\frac{1}{\pi}\int_{-1}^{1}\frac{1}{\sqrt{1-y^2}}\frac{1}{1-y(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))}\mathrm{d}y&.\end{aligned}$$

Lemma 4 implies that the estimates $\|\left[\Lambda^{(-)}_{c,\gamma}(0)-\Lambda^{(+)}_{c,\gamma}(0)\right]\left[\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0)\right]\|\leq 2C_{\mathrm{DL}}|\theta|$ and

$$\begin{aligned}&\Big\|\frac{1}{\pi}\int_{-1}^{1}\frac{1}{\sqrt{1-y^2}}\frac{1}{1-y(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))^2}\mathrm{d}y\ -\mathbf{1}\Big\| =\\ &\qquad = \Big\|\frac{1}{\pi}\int_{-1}^{1}\frac{1}{\sqrt{1-y^2}}\frac{y(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))}{1-y(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0))}\mathrm{d}y\,\Big\|\leq C'C_{\mathrm{DL}}|\theta|\end{aligned}$$

hold for some $C'>0$.
Proof of (59)*:* The strategy is similar to the proof of (58). We write

$$\begin{aligned}&|D_{c,0}(0)|^{1/2}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)|D_{c,0}(0)|^{-1/2} =\\ &=\Big\{\mathbf{1}-|D_{c,0}(0)|^{1/2}|D_{c,\gamma}(0)|^{-1/2}\left[\Lambda^{(-)}_{c,\gamma}(0)-\Lambda^{(+)}_{c,\gamma}(0)\right]|D_{c,\gamma}(0)|^{1/2}|D_{c,0}(0)|^{-1/2}\\ &\quad\times|D_{c,0}(0)|^{1/2}\left[\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0)\right]|D_{c,0}(0)|^{-1/2}\Big\}\\ &\quad\times\frac{1}{\pi}\int_{-1}^{1}\frac{1}{\sqrt{1-y^2}}\frac{1}{1-y|D_{c,0}(0)|^{1/2}\left(\Lambda^{(+)}_{c,\gamma}(\theta)-\Lambda^{(+)}_{c,\gamma}(0)\right)|D_{c,0}(0)|^{-1/2}}\mathrm{d}y\,,\end{aligned}$$

where we used that $|D_{c,\gamma}(0)|^{-1/2}$ commutes with $\Lambda^{(\pm)}_{c,\gamma}(0)$. Using Lemma 5 and Lemma 4 we obtain the claim as before. $\square$

A first application of the transformation function $\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)$ is the following lemma, which estimates the difference between the dilated Dirac operator and its original version.

Lemma 6. *Under the assumptions of Theorem 6 b) there is a constant $C_{UD} > 0$, independent of γ, c and θ, such that*

$$\Big\| |D_{c,0}(0)|^{-1/2} \big[\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) D_{c,\gamma}(\theta) \mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1} - D_{c,\gamma}(0)\big] |D_{c,0}(0)|^{-1/2} \Big\| \le C_{UD}|\theta| \quad (61)$$

holds.

Proof. We have

$$\begin{aligned}
&|D_{c,0}(0)|^{-1/2}[\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)D_{c,\gamma}(\theta)\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1} - D_{c,\gamma}(0)]|D_{c,0}(0)|^{-1/2}\\
=&|D_{c,0}(0)|^{-1/2}[\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) - \mathbf{1}]|D_{c,0}(0)|^{1/2}|D_{c,0}(0)|^{-1/2}D_{c,\gamma}(\theta)|D_{c,0}(0)|^{-1/2}\\
&\times|D_{c,0}(0)|^{1/2}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1}|D_{c,0}(0)|^{-1/2}+\\
&+|D_{c,0}(0)|^{-1/2}[D_{c,\gamma}(\theta) - D_{c,\gamma}(0)]|D_{c,0}(0)|^{-1/2}\\
&\times|D_{c,0}(0)|^{1/2}\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1}|D_{c,0}(0)|^{-1/2} + |D_{c,0}(0)|^{-1/2}D_{c,\gamma}(0)|D_{c,0}(0)|^{-1/2}\\
&\times|D_{c,0}(0)|^{1/2}[\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1} - \mathbf{1}]|D_{c,0}(0)|^{-1/2},
\end{aligned}$$

which implies the claim, if we use additionally

$$\begin{aligned}
&\||D_{c,0}(0)|^{-1/2}[D_{c,\gamma}(\theta) - D_{c,\gamma}(0)]|D_{c,0}(0)|^{-1/2}\| = \||D_{c,0}(0)|^{-1/2} \qquad (62)\\
&\times[-(e^{-\theta}-1)ic\boldsymbol{\alpha}\cdot\nabla - \gamma(V(\theta) - V)]|D_{c,0}(0)|^{-1/2}\| \le (B+\tilde{C})|\theta|(1+\frac{\pi\gamma}{2c})
\end{aligned}$$

and Theorem 6. Moreover, we used the inequality $|e^{-\theta} - 1| \le B|\theta|$ with $B = e^{\pi/4}$ and Kato's inequality in the proof of (62). $\square$

7 A resolvent estimate for the Dirac operator

In the following, we choose an $\eta > 0$ such that for some $\tilde{n} > 1$ and all $c \ge 1$ the inequalities $\tilde{E}_{\tilde{n},\tilde{n}}(c,\gamma) < c^2 - \eta$ and $\tilde{E}_{\tilde{n}+1,1}(c,\gamma) > c^2 - \eta$ hold. If $\tilde{n} = N_{\max}$, then the second condition has to be omitted.
Using the notation of Section 5 we define $P_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta) := \sum_{1\le n\le\tilde{n}} P_n(c,\gamma;\theta)$ and $\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta) := \mathbf{1} - (\Lambda^{(-)}_{c,\gamma}(\theta) + P_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta))$. Note that $\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)$ projects onto a subspace of the positive spectral subspace.

The following theorem partly generalizes [5, Lemma 3.8] for Dirac operators (see also Theorem A.1). We will slightly extend this theorem in the non-relativistic limit (see Lemma 7 and Corollary 4). This theorem and Corollary 4 enable us to control the norm of the resolvent of the non-self-adjoint operator $D_{c,\gamma}(\theta)|_{\operatorname{Ran}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)}$. Note that the usual theorems about the norm of the resolvent of a self-adjoint operator fail in general, and that for the following to hold it is essential that to restrict the operator to (a subspace of) the positive spectral subspace.

THEOREM 7. *Suppose that the assumptions of Theorem 6 b) hold. Assume additionally that the inequalities* $C_{\mathrm{UD}}|\theta|(1+2\gamma/c) < q$ *and* $2\gamma(1+C_{\mathrm{FW}}|\operatorname{Im}\theta|) < q$ *are fulfilled for some* $0 < q < 1$. *Then the following statements are true: The operator* $D_{c,\gamma}(\theta)|_{\operatorname{Ran}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)} - z$ *has a bounded inverse for all* $z \in \mathbb{C}$ *with* $\operatorname{Re} z \leq c^2 - 1$. *There is a constant* $C_R > 0$, *independent of* c, γ *and* θ, *such that for all* $z \in \mathbb{C}$ *with* $\operatorname{Re} z \leq c^2 - 1$ *the estimate*

$$\left\| \left[D_{c,\gamma}(\theta)|_{\operatorname{Ran}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)} - z \right]^{-1} \bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta) \right\| \leq \frac{C_R \|\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)\|}{c^2 - \eta - \operatorname{Re} z}$$

holds.

Proof. We make a case distinction:
Case 1: $\operatorname{Re} z \leq 0$. Theorem 6 implies the inclusion $\operatorname{Ran}(\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) \times \bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1}) \subset \operatorname{Ran}(\Lambda^{(+)}_{c,\gamma}(0))$. Thus, using Theorem 6 again, it suffices to show

$$\left\| [(\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) D_{c,\gamma}(\theta) \mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1})|_{\operatorname{Ran}\Lambda^{(+)}_{c,\gamma}(0)} - z]^{-1} \Lambda^{(+)}_{c,\gamma}(0) \right\| \leq \frac{C}{c^2 - \eta - \operatorname{Re} z}.$$

As in [5, Proof of Lemma 3.8], we use a resolvent expansion:

$$\begin{aligned}
&[(\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) D_{c,\gamma}(\theta) \mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1})|_{\operatorname{Ran}\Lambda^{(+)}_{c,\gamma}(0)} - z]^{-1} \Lambda^{(+)}_{c,\gamma}(0) \qquad (63)\\
&= \sum_{n=0}^{\infty} [D_{c,\gamma}(0)|_{\operatorname{Ran}\Lambda^{(+)}_{c,\gamma}(0)} - z]^{-1} \Lambda^{(+)}_{c,\gamma}(0) |D_{c,\gamma}(0)|^{1/2} \\
&\quad \times \Big[\Lambda^{(+)}_{c,\gamma}(0) |D_{c,\gamma}(0)|^{-1/2} |D_{c,0}(0)|^{1/2} |D_{c,0}(0)|^{-1/2} \\
&\quad \times [\mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta) D_{c,\gamma}(\theta) \mathcal{U}_{\mathrm{DL}}(c,\gamma;\theta)^{-1} - D_{c,\gamma}(0)] \\
&\quad \times |D_{c,0}(0)|^{-1/2} |D_{c,0}(0)|^{1/2} |D_{c,\gamma}(0)|^{-1/2} \Lambda^{(+)}_{c,\gamma}(0) \\
&\quad \times |D_{c,\gamma}(0)|^{1/2} [D_{c,\gamma}(0)|_{\operatorname{Ran}\Lambda^{(+)}_{c,\gamma}(0)} - z]^{-1} \Lambda^{(+)}_{c,\gamma}(0) |D_{c,\gamma}(0)|^{1/2} \Big]^n |D_{c,\gamma}(0)|^{-1/2}
\end{aligned}$$

In order to prove the convergence of the series, we have to estimate the terms in (63). First, we note that

$$|D_{c,\gamma}(0)| (D_{c,\gamma}(0)|_{\operatorname{Ran}\Lambda^{(+)}_{c,\gamma}(0)} - z)^{-1} \Lambda^{(+)}_{c,\gamma}(0) = \sup_{\lambda \geq c^2} \frac{\lambda}{|\lambda - z|} \leq 1 \qquad (64)$$

holds, since $\operatorname{Re} z \leq 0$. Moreover, the spectral theorem implies

$$\begin{aligned}
&\| |D_{c,\gamma}(0)|^{1/2} [D_{c,\gamma}(0)|_{\operatorname{Ran} \Lambda_{c,\gamma}^{(+)}(0)} - z]^{-1} \Lambda_{c,\gamma}^{(+)}(0) \|^2 \\
\leq & \| [D_{c,\gamma}(0)|_{\operatorname{Ran} \Lambda_{c,\gamma}^{(+)}(0)} - z]^{-1} \Lambda_{c,\gamma}^{(+)}(0) \| \\
& \times \| |D_{c,\gamma}(0)| [D_{c,\gamma}(0)|_{\operatorname{Ran} \Lambda_{c,\gamma}^{(+)}(0)} - z]^{-1} \Lambda_{c,\gamma}^{(+)}(0) \| \leq \frac{C}{c^2 - \eta - \operatorname{Re} z}.
\end{aligned} \tag{65}$$

Lemma 6, Lemma 5 and (64) prove the convergence of the series in (63). Using Formula (65), the claim follows for $\operatorname{Re} z \leq 0$ from (63).
Case 2: $0 \leq \operatorname{Re} z \leq c^2 - 1$. We use the resolvent expansion

$$[D_{c,\gamma}(\theta) - z]^{-1} = \sum_{n=0}^{\infty} [D_{c,0}(\theta) - z]^{-1} [\gamma V(\theta) [D_{c,0}(\theta) - z]^{-1}]^n. \tag{66}$$

Hardy's inequality and Theorem 1 yield $\|\gamma V(\theta)[D_{c,0}(\theta) - z]^{-1}\| \leq 2\gamma e^{-\operatorname{Re}\theta}(1 + C_{\mathrm{FW}} |\operatorname{Im}\theta|) \| \frac{|\nabla|}{\sqrt{-e^{-2\theta} c^2 \Delta + c^4} \beta - z} \|$. In order to control this norm, we estimate as follows:

$$\sup_{p \in \mathbb{R}^3} \frac{e^{-\operatorname{Re}\theta} |p|}{|\sqrt{e^{-2\theta} c^2 p^2 + c^4} \pm z|} \leq \frac{1}{\sqrt{\cos(2 \operatorname{Im}\theta)}} \sup_{p \in \mathbb{R}^3} \frac{|p|}{|\sqrt{c^2 p^2 + c^4} \pm \operatorname{Re} z}$$

Since $\frac{|p|}{\sqrt{c^2 p^2 + c^4} + \operatorname{Re} z} \leq 1/c$, it suffices to consider the case with the minus sign. We need to find the supremum of the function $f_{c,l} : [0, \infty) \to \mathbb{R}$, $f_{c,l}(r) := \frac{r}{\sqrt{c^2 r^2 + c^4} - l}$ for $0 \leq l \leq (c^2 - 1)$. If we differentiate this function, we find that it attains its maximum at the point $r_0 := \frac{\sqrt{c^4 - l^2} c}{l}$. Now, we define the function $g_c(l) := f_{c,l}(r_0) = \frac{c}{\sqrt{c^4 - l^2}}$ for $0 \leq l \leq (c^2 - 1)$. This function is obviously monotonously increasing in l and therefore attains its maximum at the point $l_0 := c^2 - 1$. We have $g_c(l_0) = \frac{c}{\sqrt{c^4 - (c^2 - 1)^2}} = \frac{c}{\sqrt{2c^2 - 1}} \leq 1$.
Thus, Equation (66) and Theorem 1 yield the estimate $\|[D_{c,\gamma}(\theta) - z]^{-1}\| \leq \tilde{C} \| 1 / (\sqrt{e^{-2\theta} c^2 p^2 + c^4} \beta - z) \|$ with some $\tilde{C} > 0$. Since $\sqrt{e^{-2\theta} c^2 p^2 + c^4} \beta$ is normal, we find $\|[D_{c,\gamma}(\theta) - z]^{-1}\| \leq C_R / (c^2 - \eta - \operatorname{Re} z)$, which remains true, if we restrict the resolvent to $\operatorname{Ran} \bar{P}_{\mathrm{disc}, \tilde{n}}(c, \gamma; \theta)$.

□

8 Non-relativistic limit

In this section we investigate the non-relativistic limit of complex dilated Dirac operators. We will use these results in [30], where we will discuss the interaction with the second quantized radiation field. Moreover, we can extend the resolvent estimate of Theorem 7 to the region close to the spectrum of the operator and control the norm of the projection occurring there.

8.1 General Theory

We extend some statements from [49] to the non-self-adjoint case. We define $\beta_\pm := \frac{1}{2}(\mathbf{1} \pm \beta)$ as well as $M := \{z \in \mathbb{C} | -1 \leq \operatorname{Re} z < 0, |\operatorname{Im} z| \leq 1\}$ and fix a $\gamma > 0$ such that $D_{c,\gamma}(\theta) - c^2$ has no eigenvalues E with $\operatorname{Re} E \leq -1$. This is at least true for $0 \leq \gamma < 1$ in the case of $V = V_C$, which can be seen, for example, using the explicit formula for the eigenvalues, see [35]. We define as operators on $L^2(\mathbb{R}^3; \mathbb{C}^4)$:

$$D_{\infty,0}(\theta) := -\frac{e^{-2\theta}}{2}\Delta, \qquad D_{\infty,\gamma}(\theta) := -\frac{e^{-2\theta}}{2}\Delta - \gamma V(\theta)\beta_+$$

$$K_{c,0}(\theta) := (D_{\infty,0}(\theta) - z - \frac{z^2}{2c^2})^{-1}, \quad K_{c,\gamma}(\theta) := (D_{\infty,\gamma}(\theta) - z - \frac{z^2}{2c^2})^{-1}$$

as well as

$$R_{\infty,0;\theta}(z) := (D_{\infty,0}(\theta) - z)^{-1}, \qquad R_{c,0;\theta}(z) := (D_{c,0}(\theta) - z)^{-1}$$

$$R_{\infty,\gamma;\theta}(z) := (D_{\infty,\gamma}(\theta) - z)^{-1}, \qquad R_{c,\gamma;\theta}(z) := (D_{c,\gamma}(\theta) - z)^{-1}.$$

First, we generalize [49, Theorem 6.1 and Theorem 6.4] to dilated operators. As in [49], Theorem 8 is the starting point for the investigation of the non-relativistic limit.

Theorem 8. *a) Suppose that $\theta \in S_{\pi/4}$ and $c \geq 1$. Then for $z \notin \sigma(D_{c,0}(\theta)) \cup \sigma(D_{\infty,0}(\theta))$ the resolvent relation*

$$\begin{aligned}(D_{c,0}(\theta) \mp c^2 - z)^{-1} &= \left(\beta_\pm \pm \frac{1}{2c^2}\left(-ic\boldsymbol{\alpha}\cdot\nabla \pm z\right)\right) \\ &\times \left(\mathbf{1} \mp \frac{1}{2c^2} z^2 \left(\pm D_{\infty,0}(\theta) - z\right)^{-1}\right)^{-1} \left(\pm D_{\infty,0}(\theta) - z\right)^{-1} \qquad (67)\end{aligned}$$

holds.

b) Suppose that $\theta \in S_{\min\{\pi/4,\vartheta\}}$, $\frac{2\gamma}{c} C(\operatorname{Im}\theta) < 1$ and that (H1) holds. Then for $z \in M \setminus \mathbb{R}$ the relations

$$\begin{aligned}(D_{c,\gamma}(\theta) - c^2 - z)^{-1} &= \left(\beta_+ + \frac{1}{2c^2}(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla + z)\right) \\ &\times K_{c,\gamma}(\theta)\left(\mathbf{1} - \frac{\gamma}{2c^2}V(\theta)(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla + z)K_{c,\gamma}(\theta)\right)^{-1} \qquad (68)\end{aligned}$$

and

$$K_{c,\gamma}(\theta) = \left(\mathbf{1} - \frac{z^2}{2c^2}(D_{\infty,\gamma}(\theta) - z)^{-1}\right)^{-1} (D_{\infty,\gamma}(\theta) - z)^{-1} \qquad (69)$$

hold.

Proof.
a) We follow the proof of [49, Theorem 6.1], noting that $z \in \mathbb{C}$ with $z(1+\frac{z}{2c^2}) \notin e^{-2\mathrm{i}\,\mathrm{Im}\,\theta}[0,\infty)$ is equivalent to $z+c^2 \notin \sigma(D_{c,0}(\theta))$. In order to show Equation (67), we define the operators $A_\pm(\theta) := D_{c,0}(\theta) \pm c^2 \pm z = -ic\boldsymbol{\alpha}\cdot\nabla \pm 2c^2\beta_\pm \pm z$ and note that $A_+(\theta)A_-(\theta) = A_-(\theta)A_+(\theta) = -c^2 e^{-2\mathrm{i}\,\theta}\Delta - 2c^2 z - z^2$. This yields

$$A_\pm(\theta)^{-1} = \frac{A_\mp(\theta)}{2c^2}\left(D_{\infty,0}(\theta) - z - \frac{z^2}{2c^2}\right)^{-1}, \tag{70}$$

which in turn implies the claim. Note that all operators are equivalent to multiplication operators.
b) We follow the proof of [49, Theorem 6.2]. Theorem 3 implies that $z + c^2 \notin \sigma(D_{c,\gamma}(\theta))$. It follows that $D_{c,\gamma}(\theta) - (c^2+z) = A_-(\theta) - \gamma V(\theta) = (\mathbf{1} + \gamma V(\theta)A_-(\theta)^{-1})A_-(\theta)$. Since $D_{c,\gamma}(\theta) - (c^2+z)$ and $A_-(\theta)$ have bounded inverses, the bounded operator $\mathbf{1} + \gamma V(\theta)A_-(\theta)^{-1}$ is bijective, and is thus in particular bounded invertible. From Equation (70) it follows that

$$(D_{c,\gamma}(\theta) - c^2 - z)^{-1} = (A_-(\theta) - \gamma V(\theta))^{-1} = A_-(\theta)^{-1}(1 - \gamma V(\theta)A_-(\theta)^{-1})^{-1} \tag{71}$$

$$= \left(\beta_+ + \frac{1}{2c^2}(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla + z)\right)(D_{\infty,0}(\theta) - z - \frac{z^2}{2c^2})^{-1}$$
$$\times\left(1 - \gamma V(\theta)\beta_+ K_{c,0}(\theta) - \frac{\gamma}{2c^2}V(\theta)(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla + z)K_{c,0}(\theta)\right)^{-1}.$$

$z \in M \setminus \mathbb{R}$ implies $z + z^2/(2c) \in M\setminus\mathbb{R}$ and in particular $z(1+\frac{z}{2c^2}) \notin \sigma(D_{\infty,\gamma}(\theta))$, which shows (69). Moreover $K_{c,\gamma}(\theta) = K_{c,0}(\theta) - \gamma V(\theta)\beta_+ = (\mathbf{1} - \gamma V(\theta)\beta_+ K_{c,0}(\theta)^{-1})K_{c,0}(\theta)$ holds. To see this, note that $z + c^2 \notin \sigma(D_{c,\gamma}(\theta))$ implies $z + c^2 \notin \sigma(D_{c,0}(\theta))$, which in turn implies $z(1+\frac{z}{2c^2}) \notin \sigma(D_{\infty,0}(\theta))$, i.e. $K_{c,0}(\theta)$ is bounded invertible. Thus, the bounded operator $\mathbf{1} - \gamma V(\theta)\beta_+ K_{c,0}(\theta)^{-1}$ has a bounded inverse, and

$$K_{c,\gamma}(\theta)^{-1} = K_{c,0}(\theta)^{-1}(\mathbf{1} - \gamma V(\theta)\beta_+ K_{c,0}(\theta)^{-1})^{-1} \tag{72}$$

as well as

$$\left(1 - \gamma V(\theta)\beta_+ K_{c,0}(\theta) - \frac{\gamma}{2c^2}V(\theta)(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla + z)K_{c,0}(\theta)\right)^{-1}$$
$$= (\mathbf{1} - \gamma V(\theta)K_{c,0}(\theta)^{-1}\beta_+)^{-1}$$
$$\times\left(\mathbf{1} - \frac{\gamma}{2c^2}V(\theta)(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla + z)K_{c,0}(\theta)(\mathbf{1} - \gamma V(\theta)K_{c,0}(\theta)^{-1}\beta_+)^{-1}\right)^{-1} \tag{73}$$

hold. Using (72) and (73), (68) follows from (71). ∎

We denote the real eigenvalues of $D_{\infty,\gamma}(\theta)$ by $E_n(\infty,\gamma) = E_n(\gamma)$, ordered by size and not counting multiplicities. Note that by dilation analyticity these eigenvalues are the same as the eigenvalues of $D_{\infty,\gamma}(0)$ and that $|E_n(\infty,\gamma) - E_{n,l}(c,\gamma)| = \mathcal{O}(1/c^2)$ for all $l \in \{1,\dots,N_n\}$ by [49, Theorem 6.8]. We pick now η as in Section 7 and define for each $\tilde\epsilon > 0$ the set

$$M_{\eta,\tilde\epsilon} := \{z \in \mathbb{C} | 1 \le \operatorname{Re} z \le -(\eta+\tilde\epsilon),\ |\operatorname{Im} z| \le 1,\ \operatorname{dist}(z, \sigma(D_{c,\gamma}(\theta))) \ge \tilde\epsilon\}.$$

Moreover, we set $D(w,r) := \{z \in \mathbb{C} || z-w| < r\}$ for $w \in \mathbb{C}$ and $r > 0$. Fix $\tilde\epsilon > 0$ so small that for all $n, n' \in \mathbb{N}$ with $n \neq n'$ and $1 \le n, n' \le \tilde n$ the sets $D(E_n(\infty,\gamma), 2\tilde\epsilon)$ and $D(E_{n'}(\infty,\gamma), 2\tilde\epsilon)$ are disjoint and contained in the set $\{z \in \mathbb{C} | 1 \le \operatorname{Re} z \le -(\eta+\tilde\epsilon),\ |\operatorname{Im} z| \le 1\}$. Now we pick for $\tilde\epsilon > 0$ a contour Γ with positive orientation such that Γ is contained $M_{\eta,\tilde\epsilon}$ and has only the eigenvalue $E_n(\gamma)$ in its interior, but no other eigenvalues of $\sigma(D_{\infty,\gamma}(\theta))$. Then we define

$$P_n(\infty,\gamma;\theta) := -\frac{1}{2\pi\mathrm{i}} \int_\Gamma \mathrm{d}z\, R_{\infty,\gamma;\theta}(z)\beta_+.$$

We set $P_{\mathrm{disc}}(\infty,\gamma;\theta) := \sum_{i=1}^{\tilde n} P_i(\infty,\gamma;\theta)$ and $\bar P_{\mathrm{disc}}(\infty,\gamma;\theta) := 1 - P_{\mathrm{disc}}(\infty,\gamma;\theta)$. Note that using the definitions in Appendix A and in slight abuse of notation $P_{\mathrm{disc}}(\infty,\gamma;\theta) = P_{\mathrm{disc}}(\gamma;\theta)\beta_+$ and $\bar P_{\mathrm{disc}}(\infty,\gamma;\theta) = \beta_- + \bar P_{\mathrm{disc}}(\gamma;\theta)\beta_+$.

Now we are in the position to generalize [49, Corollary 6.5] to dilated operators.

Corollary 2. *Suppose that $|\theta| < \theta_0$, where θ_0 is sufficiently small (see Appendix A), and $\theta \in S_{\min\{\pi/4,\Theta\}}$ as well as $\frac{2\gamma}{c} C(\operatorname{Im}\theta) < 1$. Suppose moreover that (H1) holds. Then the resolvent expansion*

$$\left[D_{c,\gamma}(\theta) - (c^2+z)\right]^{-1} = \sum_{n=0}^{\infty} \frac{1}{c^n} R_n(z). \tag{74}$$

holds for all $z \in M_{\eta,\tilde\epsilon}$ and all sufficiently large c. The series converges in norm, uniformly in θ and z. In particular,

$$[D_{c,\gamma}(\theta) - (c^2+z)]^{-1} \underset{c\to\infty}{\longrightarrow} [D_{\infty,\gamma}(\theta) - z]^{-1}\beta_+$$

uniformly in θ and z.

Proof. First, we need an estimate on the resolvent of $D_{\infty,\gamma}(\theta)$. We split the resolvent according to

$$\begin{aligned}[D_{\infty,\gamma}(\theta) - z]^{-1} &= [D_{\infty,\gamma}(\theta)|_{\operatorname{Ran} \bar P_{\mathrm{disc}}(\infty,\gamma;\theta)} - z]^{-1} \bar P_{\mathrm{disc}}(\infty,\gamma;\theta) \\ &\quad + \sum_{n=1}^{\tilde n} [D_{\infty,\gamma}(\theta)|_{\operatorname{Ran} P_n(\infty,\gamma;\theta)} - z]^{-1} P_n(\infty,\gamma;\theta).\end{aligned} \tag{75}$$

Theorem A.1 implies that the norm of the first summand in (75) is bounded by $2/\eta$. The norms of the other summands can be estimated according to $\left\|[D_{\infty,\gamma}(\theta)|_{\operatorname{Ran} P_n(\infty,\gamma;\theta)} - z]^{-1} P_n(\infty,\gamma;\theta)\right\| \le \frac{\|P_n(\infty,\gamma;\theta)\|}{\operatorname{dist}(z,E_n(\gamma))} \le \frac{C|\theta|}{\operatorname{dist}(z,E_n(\gamma))}$ using

Corollary A.1. Thus, we have for sufficiently small $1/c$ (dependent on $\tilde{\epsilon}$) and all $z \in M_{\eta,\tilde{\epsilon}}$ the expansion

$$(\mathbf{1} - \frac{z^2}{2c^2}(D_{\infty,\gamma}(\theta) - z)^{-1})^{-1} = (D_{\infty,\gamma}(\theta) - z)^{-1} \sum_{n=0}^{\infty} (\frac{z^2}{2c^2}(D_{\infty,\gamma}(\theta) - z)^{-1})^n.$$

Hardy's inequality implies for $f \in H^2(\mathbb{R}^3; \mathbb{C}^4)$ the estimates $\|Vf\| \leq 2\|\nabla f\| \leq a\|\Delta f\| + (1/a)\|f\|$ and $e^{-2\operatorname{Re}\theta}\|\Delta f\| \leq 1/(1-2a\gamma)\|D_{\infty,\gamma}(\theta)f\| + 2\gamma/[a(1-2a\gamma)]\|f\|$ with a sufficiently small $a > 0$. It follows that

$$\|\frac{\gamma}{2c^2}V(\theta)(-ice^{-\theta}\boldsymbol{\alpha}\cdot\nabla+z)(D_{\infty,\gamma}(\theta)-z)^{-1}\| \leq \frac{\gamma}{c}[C_1+C_2\|(D_{\infty,\gamma}(\theta)-z)^{-1}\|]$$

holds with $C_1, C_2 > 0$ (independent of γ, c and θ), which implies that the last factor in (68) has a norm convergent series expansion in $1/c$ for $1/c$ small enough. □

REMARK 4. *We find $R_0(z) := \beta_+ R_{\infty,\gamma;\theta}(z)$ as in [49]. As in [49, Remark after Corollary 6.5], the operators occurring with even powers of $1/c$ are even, and the operators occurring with odd powers of $1/c$ are odd .*

LEMMA 7. *Suppose that the assumptions of Corollary 2 hold. Then there is a constant $C_{\mathrm{P},n} > 0$ (independent of c and θ) such that for sufficiently large c the estimate*

$$\|P_n(c,\gamma;\theta) - P_n(\infty,\gamma;\theta)\| \leq \frac{C_{\mathrm{P},n}}{c}$$

holds.

Proof. This follows immediately from Corollary 2. □

The following two corollaries extend Theorem 7.

COROLLARY 3. *Suppose that the assumptions of Corollary 2 hold. Then there is a constant $C > 0$ (possibly dependent on θ) such that for all $z \in \mathbb{C}$ with $-1 \leq \operatorname{Re} z \leq -\eta$ and $|\operatorname{Im} z| \leq 1$ and all sufficiently large c the estimate*

$$\|[D_{c,\gamma}(\theta)|_{\operatorname{Ran} \bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)} - (c^2+z)]^{-1}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)\| \leq C$$

holds.

Proof. Corollary 2 implies that $[D_{c,\gamma}(\theta)|_{\operatorname{Ran} \bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)} - (c^2+z)]^{-1}$ is uniformly bounded in $z \in M_{\eta,\tilde{\epsilon}}$ and c (for sufficiently large c). Lemma 7 and Lemma 4 yield the existence of an upper bound on

$$\|\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)\| = \|\mathbf{1} - (\Lambda_{c,\gamma}^{(-)}(\theta) + P_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta))\|,$$

which does not depend on c. Thus the claim holds for $z \in M_{\eta,\tilde{\epsilon}}$.

Let $z_0 \in D(E_n(\infty,\gamma),\tilde{\epsilon})$. Then $\Gamma := \{z \in \mathbb{C} | |z - E_n(\infty,\gamma)| = 2\tilde{\epsilon}\} \subset M_{\eta,\tilde{\epsilon}}$ holds because of the definition of the set $M_{\eta,\tilde{\epsilon}}$. Since $[D_{c,\gamma}(\theta)|_{\mathrm{Ran}\,\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta))} - (c^2 + z)]^{-1}$ in $z \in \{z \in \mathbb{C}| -1 \le \mathrm{Re}\, z \le -\eta,\ |\mathrm{Im}\, z| \le 1\}$ is holomorphic,

$$[D_{c,\gamma}(\theta)|_{\mathrm{Ran}\,\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)} - (c^2 + z_0)]^{-1}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta) = -\frac{1}{2\pi \mathrm{i}} \\ \times \int_\Gamma [D_{c,\gamma}(\theta)|_{\mathrm{Ran}\,\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta))} - (c^2 + z)]^{-1}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)\frac{1}{z - z_0}\mathrm{d}z$$

holds, where the contour is oriented in the positive sense. This implies the claim for $z_0 \in D(E_n(\infty,\gamma),\tilde{\epsilon})$. □

Corollary 4. *Suppose that the assumptions of Theorem 7 and Corollary 2 hold. Then there is a $C > 0$ (possibly dependent on θ) such that for all $z \in \mathbb{C}$ with $-1 < \mathrm{Re}\, z \le -\eta$ and $|\mathrm{Im}\, z| \le 1$ or with $-\infty < \mathrm{Re}\, z \le -1$ and all sufficiently large c the estimate*

$$\|[D_{c,\gamma}(\theta)|_{\mathrm{Ran}\,\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)} - (c^2 + z)]^{-1}\bar{P}_{\mathrm{disc},\tilde{n}}(c,\gamma;\theta)\| \le \frac{C}{-\eta - \mathrm{Re}\, z}$$

is true.

Proof. This follows immediately from Corollary 3 and Theorem 7 together with Lemma 7. □

Now, we define a transformation function $\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta) : L^2(\mathbb{R}^3;\mathbb{C}^4) \to L^2(\mathbb{R}^3;\mathbb{C}^4)$ by

$$\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta) := [P_n(c,\gamma;\theta)P_n(\infty,\gamma;\theta) + (\mathbf{1} - P_n(c,\gamma;\theta))(\mathbf{1} - P_n(\infty,\gamma;\theta))] \\ \times [\mathbf{1} - (P_n(c,\gamma;\theta) - P_n(\infty,\gamma;\theta))^2]^{-1/2}.$$

Lemma 8. *Suppose that the assumptions of Corollary 2 and the inequality $C_{\mathrm{P},n}/c < q < 1$ hold for some $0 < q < 1$. Then the mapping $\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)$ is bounded with bounded inverse $\mathcal{V}_{\mathrm{NR}}(c,\gamma;\theta)$. The relations*

$$\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)P_l(\infty,\gamma;\theta)\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)^{-1} = P_l(c,\gamma;\theta) \tag{76}$$

and

$$\|\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta) - \mathbf{1}\| \le \frac{C_{\mathrm{NRP}}}{c} \tag{77}$$

hold with a constant $C_{\mathrm{NRP}} > 0$ independent of c and θ. $\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)$ is a holomorphic function of θ.

Proof. Using Lemma 7 this can be proven in the same way as Theorem 6. For the holomorphicity in θ note that the power series (in $1/c$) for $R_{c,\gamma;\theta}(z)$, $P_n(c,\gamma;\theta)$ and $\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)$ converge uniformly in θ. □

Remark 5. *As in [49] we obtain by Remark 4 that in the series expansion of* $\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)$ *the operators occurring with even powers of* $1/c$ *are even and the operators occurring with odd powers of* $1/c$ *are odd. In particular,*

$$\mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta) = \mathcal{U}_{NR,g}(c,\gamma;\theta) + \frac{1}{c}\mathcal{U}_{NR,ug}(c,\gamma;\theta), \tag{78}$$

where $\mathcal{U}_{NR,g}(c,\gamma;\theta)$ *and* $\mathcal{U}_{NR,ug}(c,\gamma;\theta)$ *are even and odd operators holomorphic in* $1/c$.

The following theorem generalizes [49, Theorem 6.7] and shows that the lower component of an eigen-spinor of the Dirac operator converges to zero as $c \to \infty$.

Theorem 9. *Suppose that the assumptions of Lemma 8 hold. Then the normed eigenfunctions* $\phi_n(c,\gamma;\theta)$ *of* $D_{c,\gamma}(\theta)$ *with eigenvalue* $E_{n,l}(c,\gamma)$ *have the form*

$$\phi_{n,l}(c,\gamma;\theta) = \phi_{n,l,+}(c,\gamma;\theta) + \frac{1}{c}\phi_{n,l,-}(c,\gamma;\theta), \qquad \phi_{n,l,\pm}(c,\gamma;\theta) \in \beta_\pm L^2(\mathbb{R}^3;\mathbb{C}^4), \tag{79}$$

where $\phi_{n,l,\pm}(c,\gamma;\theta)$ *are continuous functions of* $1/c$.

Proof. We have $P_n(c,\gamma;\theta)D_{c,\gamma}(\theta)P_n(c,\gamma;\theta) = -\frac{1}{2\pi\mathrm{i}}\int_\Gamma \frac{z}{D_{c,\gamma}(\theta)-z}\mathrm{d}z$. Any eigenvector $\tilde{\phi}_n(c,\gamma;\theta)$ of $P_n(c,\gamma;\theta)D_{c,\gamma}(\theta)P_n(c,\gamma;\theta)$ and thus any eigenvector of $D_{c,\gamma}(\theta)$ with eigenvalue $E_{n,l}(c,\gamma)$ is given by $\tilde{\phi}_{n,l}(c,\gamma;\theta)$ $= \mathcal{U}_{\mathrm{NR}}(c,\gamma;\theta)\phi_{n,l}(\infty,\gamma;\theta)$ for a $\phi_{n,l}(\infty,\gamma;\theta) \in \beta_+L^2(\mathbb{R}^3;\mathbb{C}^4)$. Remark 5 and the analytic perturbation theory imply $\tilde{\phi}_{n,l}(c,\gamma;\theta) = \tilde{\phi}_{n,l,+}(c,\gamma;\theta) + \frac{1}{c}\phi_{n,l,-}(c,\gamma;\theta)$, where $\tilde{\phi}_{n,l}(c,\gamma;\theta)$ and $\tilde{\phi}_{n,l,\pm}(c,\gamma;\theta)$ are holomorphic functions of $1/c$. Since the projections $P_n(c,\gamma;\theta)$ are nor orthogonal, the normed eigenfunctions are in general not holomorphic functions of $1/c$. But nevertheless $\|\tilde{\phi}_{n,l}(c,\gamma;\theta)\| \geq 1 - C\frac{1}{c}$ holds for some $C > 0$ and thus (79) follows. □

We use these statements to prove that eigenfunctions are bounded in the norm of $H^1(\mathbb{R}^3;\mathbb{C}^4)$.

Theorem 10. *Suppose the assumptions of Lemma 8 hold. Then there is a constant* $C_{\mathrm{EF}} > 0$*, independent of* c*, such that the normed eigenfunctions* $\phi_n(c,\gamma;\theta)$ *of* $D_{c,\gamma}(\theta)$ *with eigenvalue* $E_{n,l}(c,\gamma)$ *fulfill the estimates*

$$\|\nabla\phi_{n,l,+}(c,\gamma;\theta)\| \leq C_{\mathrm{EF}} \tag{80}$$

and

$$\|\nabla\phi_{n,l,-}(c,\gamma;\theta)\| \leq \frac{C_{\mathrm{EF}}}{c} \tag{81}$$

for sufficiently large c.

Proof. We follow Esteban and Séré [10, Proof of Lemma 7 and Theorem 3], who considered the non-relativistic limit of self-adjoint Dirac-Fock operators. Since $D_{c,\gamma}(\theta)$ is not self-adjoint, there are some additional difficulties. To simplify the notation, we suppress the dependence of $\phi_{n,l}(c,\gamma;\theta)$ on c, γ and θ. We have

$$\begin{aligned}
&E_{n,l}(c,\gamma)^2\|\phi_{n,l}\|^2 = \|D_{c,\gamma}(\theta)\phi_{n,l}\|^2\\
\geq &e^{-2\operatorname{Re}\theta}[c^2(1-2\sin\operatorname{Im}\theta-\gamma/4)-4c\gamma]\|\nabla\phi_{n,l}\|^2\\
+&[c^4(1-2\sin\operatorname{Im}\theta)-16\gamma c^2]\|\phi_{n,l}\|^2,
\end{aligned}$$

where we used Hardy's inequality. Since $E_{n,l}(c,\gamma)^2 - c^4 \leq 0$, it follows that

$$\begin{aligned}
\|\nabla\phi_{n,l}\|^2 \leq &\frac{E_{n,l}(c,\gamma)^2 - c^4 + 2\sin\operatorname{Im}\theta c^2 + 16c^2}{c^2(1-2\sin\operatorname{Im}\theta-1/4)-4c\gamma}\|\phi_{n,l}\|^2\\
&\leq C(\sin\operatorname{Im}\theta c^2+1)\|\phi_{n,l}\|^2 \qquad (82)
\end{aligned}$$

for sufficiently large c, where $C > 0$ does not depend on c.
Note that the term proportional to c^2 in (82) does not occur for $\operatorname{Im}\theta = 0$, which implies immediately the boundedness of $\|\nabla\phi_{n,l}\|$ in this case. To circumvent this difficulty, we write the Dirac equation in its components, where (in abuse of notation) $\phi_{n,l,\pm}$ denotes the upper and, respectively, lower components of $\phi_{n,l}$:

$$ce^{-\theta}\boldsymbol{\sigma}\cdot\nabla\phi_{n,l,-} - \gamma V(\theta)\phi_{n,l,+} + c^2\phi_{n,l,+} = E_{n,l}(c,\gamma)\phi_{n,l,+} \qquad (83)$$

$$ce^{-\theta}\boldsymbol{\sigma}\cdot\nabla\phi_{n,l,+} - \gamma V(\theta)\phi_{n,l,-} - c^2\phi_{n,l,-} = E_{n,l}(c,\gamma)\phi_{n,l,-} \qquad (84)$$

Dividing (83) by c, using Hardy's inequality and the boundedness of $E_{n,l}(c,\gamma)-c^2$, Formula (82) implies

$$\|\nabla\phi_{n,l,-}\| \leq \frac{2}{c}\|\nabla\phi_{n,l,+}\| + \frac{|E_{n,l}(c,\gamma)-c^2|}{c|e^{-\operatorname{Re}\theta}|}\|\phi_{n,l,+}\| \leq C \qquad (85)$$

for some $C > 0$ independent of c, i.e. $\|\nabla\phi_{n,l,-}\|$ is bounded in c. Dividing (84) by c, we obtain

$$\|\nabla\phi_{n,l,+}\| \leq \frac{2}{c}\|\nabla\phi_{n,l,-}\| + \frac{|E_{n,l}(c,\gamma)+c^2|}{c|e^{-\operatorname{Re}\theta}|}\|\phi_{n,l,-}\| \leq C \qquad (86)$$

for some $C > 0$ independent of c, where we used Theorem 9 and Equation (85). This shows (80). Inserting (86) in (85), Equation (81) follows. □

Remark 6. *Their validity of Theorem 9 and Theorem 10 in the Coulomb case could be derived from the explicit form of the eigenfunctions (see the proof of Lemma 11).*

Moreover, we need a bound on the norm of the dilation operator $\mathcal{U}(\theta)$, restricted to the spaces $\operatorname{Ran} P_n(c,\gamma;\theta)$.

Lemma 9. *Suppose that the assumptions of Lemma 8 hold. Then the family of operators* $\mathcal{U}(\theta)|_{\operatorname{Ran} P_n(c,\gamma;0)} : \operatorname{Ran} P_n(c,\gamma;0) \to \operatorname{Ran} P_n(c,\gamma;\theta)$ *is uniformly bounded in* c *and* θ.

Proof. Surely $\mathcal{U}(\theta)|_{\operatorname{Ran} P_n(\infty,\gamma;0)} : \operatorname{Ran} P_n(\infty,\gamma;0) \to \operatorname{Ran} P_n(\infty,\gamma;\theta)$ is well defined for all $\theta \in \mathbb{C}$ with $|\theta| \le \min\{\pi/4, \Theta\}$ (see [2, 6]) and (as a mapping between finite-dimensional vector spaces) bounded. Since the operator is a holomorphic function of θ for $|\theta| \le \min\{\pi/4, \Theta\}$, there is a bound $C' > 0$ (independent of θ) on its norm.
Let $f \in \operatorname{Ran} P_n(c,\gamma;0)$. Then there is a $\tilde{f} \in \operatorname{Ran} P_n(\infty,\gamma;0)$ with $f = U_{\mathrm{NR}}(c,\gamma;0)\tilde{f}$, and for real θ $f(\theta) := \mathcal{U}_{\mathrm{el}}(\theta) f = \mathcal{U}(\theta) U_{\mathrm{NR}}(c,\gamma;0)\mathcal{U}(\theta)^{-1}\tilde{f}(\theta) = U_{\mathrm{NR}}(c,\gamma;\theta)\tilde{f}(\theta)$ holds, where $\tilde{f}(\theta) := \mathcal{U}_{\mathrm{el}}(\theta)\tilde{f}$. By holomorphic continuation we obtain for complex θ the equality $f(\theta) = U_{\mathrm{NR}}(c,\gamma;\theta)\tilde{f}(\theta)$. Thus Lemma 8 implies $\|f(\theta)\| \le \|U_{\mathrm{NR}}(c,\gamma;\theta)\|\|\tilde{f}(\theta)\| \le (1 + C_{\mathrm{NRP}}/c)C'\|\tilde{f}\| \le (1 + C_{\mathrm{NRP}}/c)C'\|f\|$ for some $C' > 0$ independent of c and θ. □

The following corollary shows that also the projections on the fine structure components are bounded uniformly in c. This follows from the fact the dilated projections are similar to the corresponding orthogonal projections belonging to the corresponding self-adjoint Dirac operators because of Lemma 9. Note that in general such projections are not uniformly bounded in the perturbation parameter (see [33, Chapter II-1.5]).

Corollary 5. *Let* $1 \le n \le \tilde{n}$ *and suppose that the assumptions of Lemma 9 hold. Then* $\|P_{n,l}(c,\gamma;\theta)\| \le C$ *for some* $C > 0$ *independent of* n, l, c *and* θ.

Proof. This follows from Lemma 5, since the projections $P_{n,l}(c,\gamma;0) = = \mathcal{U}(\theta)^{-1}P_{n,l}(c,\gamma;0)\mathcal{U}(\theta)$ are orthogonal. □

8.2 Application to expectation values of Dirac matrices

We are now in the position to investigate expectation values of the matrices $\boldsymbol{\alpha}$. Since these matrices are odd, such expectation values involve scalar products of the upper component of one spinor with the lower component of the other spinor. Therefore, one expects that such expectation values converge to zero like $1/c$ as $c \to \infty$ uniformly in a set of suitable spinors. We show in the following that this is true, if one of the spinors is in the set of eigenstates (in the positive part of the gap) and the other state is an arbitrary state from the positive spectral subspace. Note that this is not true, if both states are arbitrary states from the positive spectral subspace. At least for the free spectral subspaces this can be seen from the explicit form of the projections (see Section 5). We will apply this result in [30].

Lemma 10. *Suppose that the assumptions of Lemma 9 hold and let $\tilde n$ as in Section 7. Then there is a constant $C > 0$, independent of c and θ, such that for all $1 \le n, n' \le \tilde n$, $1 \le l \le n$, $1 \le l' \le n'$ and $k_1, k_2 \in \mathbb{R}^3$*

$$\|P_{n,l}(c,\gamma;\theta) k_1 \cdot \boldsymbol{\alpha} e^{ik_2\cdot x} P_{n',l'}(c,\gamma;\theta)\| \le \frac{C|k_1|}{c}.$$

Proof. This follows from Theorem 9 and Corollary 5, since $\boldsymbol{\alpha}$ is an odd operator. □

Lemma 11. *Suppose that $V = V_C$. Let $c \ge 1$ and $\gamma/c < \sqrt{3}/2$. Then there is a constant $C > 0$, independent of c, such that*

$$\||x| P_{n,l}(c,\gamma;0)\| \le C$$

holds, where x denotes the operator of multiplication with the space variable.

Proof. We define the unitary dilations $U_c\ f_c(x) := (U_c f)(x) := c^{-3/2} f(c^{-1}x)$ and note that $U_c D_{c,\gamma} U_c^{-1} = c^2 D_{1,\gamma/c}$. Thus, if $f \in H^1(\mathbb{R}^3;\mathbb{C}^4)$ is a normed eigenfunction of $D_{c,\gamma}$ with eigenvalue $E_{n,l}$, then f_c is a normed eigenfunction of $D_{1,\gamma/c}$ with eigenvalue $E_{n,l}/c^2$. The radial parts $f_{c,\pm}(r)$ of the upper and lower component, respectively, of f_c are (see [35, Abschnitt 36])

$$\begin{aligned} f_{c,\pm}(r) := \frac{\pm(2\lambda)^{3/2}}{\Gamma(2\tilde\gamma+1)} \frac{(1 \pm E_{n,l}/c^2)\Gamma(2\tilde\gamma + n_r + 1)}{4\frac{\gamma}{c\lambda}(\frac{\gamma}{c\lambda} - \kappa) n_r!} (2\lambda r)^{\tilde\gamma - 1} e^{-\lambda r} \\ \times \{(\frac{\gamma}{c\lambda} - \kappa) F(-n_r, 2\tilde\gamma+1, 2\lambda r) \mp n_r F(1-n_r, 2\tilde\gamma+1, 2\lambda r)\}. \end{aligned}$$

Here the radial quantum number fulfills $n_r \in \mathbb{N}_0$ if $\kappa < 0$ and $n_r \in \mathbb{N}$ if $\kappa > 0$, and $\kappa \in \pm\mathbb{N}$ is the eigenvalue of the spin-orbit operator (see [49, Chapter 4.6]). F denotes the confluent hypergeometric function, which reduces to a polynomial in $2\lambda r$ here (see [35, Abschnitt 36] and [34, Abschnitt d]). Moreover, $\tilde\gamma := \sqrt{\kappa^2 - \gamma^2/c^2}$ and $\lambda := \sqrt{1 - E_{n,l}^2/c^4}$. Thus, the radial parts $f_\pm(r)$ of the upper respectively lower components of f are

$$\begin{aligned} f_{\pm}(r) := \frac{\pm(2c\lambda)^{3/2}}{\Gamma(2\tilde\gamma+1)} \frac{(1 \pm E_{n,l}/c^2)\Gamma(2\tilde\gamma + n_r + 1)}{4\frac{\gamma}{c\lambda}(\frac{\gamma}{c\lambda} - \kappa) n_r!} (2\lambda r)^{\tilde\gamma - 1} e^{-c\lambda r} \\ \times \{(\frac{\gamma}{c\lambda} - \kappa) F(-n_r, 2\tilde\gamma+1, 2c\lambda r) \mp n_r F(1-n_r, 2\tilde\gamma+1, 2c\lambda r)\}. \end{aligned}$$

Using the explicit formula (see [35]) for the eigenvalues, we see that $c\lambda$ is a function bounded in c with $c\lambda \longrightarrow \gamma/n$ for $c \to \infty$. Moreover, obviously $\tilde\gamma \to |\kappa|$ holds. This shows the claim. □

Remark 7. *At this point we make use of the explicit from of the eigenfuntions of the Coulomb Dirac operator. There do not seem any results to be available in the literature about exponential decay of eifenfunctions of the Dirac operator uniformly in the velocity of light.*

LEMMA 12. *Suppose that the assumptions of Lemma 9 are fulfilled and let $\tilde{n}$ as in Section 7. Let moreover $f : \mathbb{C} \to \mathbb{C}$ with $|f(z)| \leq |z|$. Then there is a constant $C > 0$, independent of c, such that for all $1 \leq n \leq \tilde{n}$, $1 \leq l \leq n$ and $k_1, k_2 \in \mathbb{R}^3$*

$$\|P_{n,l}(c,\gamma;0)k_1 \cdot \boldsymbol{\alpha} f(k_2 \cdot x)P_{n',l'}(c,\gamma;0)\| \leq \frac{C|k_1||k_2|}{c}.$$

Proof. Lemma 11 implies that $\|xP_{n,l}(c,\gamma;0)\|$ is uniformly bounded in c, in particular (using the notation of Theorem 9) $x\phi_{n,l,+}(c,\gamma;0)$. Now the claim follows exactly as in Lemma 10. □

The following theorem generalizes Lemma 10. Note that the statement of Lemma 10 is not completely obvious, since not even the lower component of the free positive spectral projection converges to zero in norm as $c \to \infty$. This is, however, compensated by the fact that the H^1-norm of the upper component of bound states is bounded uniformly in c (Theorem 10).

THEOREM 11. *Suppose the assumptions of Lemma 9 hold and let $\tilde{n}$ as in Section 7. Then there is a constant $C > 0$, independent of c and θ such that for all $1 \leq n \leq \tilde{n}$, $1 \leq l \leq n$ and $k_1, k_2 \in \mathbb{R}^3$*

$$\|P_{n,l}(c,\gamma;\theta)k_1 \cdot \boldsymbol{\alpha} e^{ik_2\cdot x}\Lambda^{(\pm)}_{c,\gamma}(\theta)\| \leq \frac{C|k_1|(1+|k_2|)}{c}.$$

Proof. Corollary 1 and Corollary 5 imply $\|P_{n,l}(c,\gamma;\theta)k_1 \cdot \boldsymbol{\alpha} e^{ik_2\cdot x}\Lambda^{(\pm)}_{c,\gamma}(\theta)\| \leq \|P_{n,l}(c,\gamma;\theta)k_1 \cdot \boldsymbol{\alpha} e^{ik_2\cdot x}\Lambda^{(\pm)}_{c,0}(\theta)\| + C_{\mathrm{NR}}|k_1|\frac{\gamma}{c}$ with some $C > 0$ independent of θ and c. Thus, it suffices to show $\|P_{n,l}(c,\gamma;\theta)k \cdot \boldsymbol{\alpha} e^{ik_2\cdot x}\Lambda^{(\pm)}_{c,0}(\theta)\| \leq \frac{C|k_1|(1+|k_2|)}{c}$ for some $C > 0$. In a first step, we pick $f \in \operatorname{Ran} P_{n,l}(c,\gamma;\bar{\theta})$ and $g \in \operatorname{Ran}\Lambda^{(\pm)}_{c,0}(\theta)$ normed, but arbitrary otherwise. We have $g = V_{\mathrm{FW}}(c;\theta)(\tilde{g},0)^T$ for some $\tilde{g} \in L^2(\mathbb{R}^3;\mathbb{C}^2)$. It follows that $g = \mathcal{F}^{-1}(\frac{c^2+E_c(p;\theta)}{N_c(p;\theta)}\mathcal{F}\tilde{g}, \frac{ce^{-\theta}\boldsymbol{\sigma}\cdot p}{N_c(p;\theta)}\mathcal{F}\tilde{g})^T$, where $\mathcal{F}$ denotes both the Fourier transform on $L^2(\mathbb{R}^3;\mathbb{C}^2)$ and on $L^2(\mathbb{R}^3;\mathbb{C}^4)$. We decompose $f = (f_+, f_-)^T$ with $f_\pm \in L^2(\mathbb{R}^3;\mathbb{C}^2)$. It follows that

$$\begin{aligned}
&|(f, P_{n,l}(c,\gamma;\theta)k_1\boldsymbol{\alpha} e^{ik_2\cdot x}\Lambda^{(\pm)}_{c,0}(\theta)g)| \\
\leq &|(f_+, k_1\cdot\boldsymbol{\sigma}\mathcal{F}^{-1}\frac{-ce^{-\theta}\boldsymbol{\sigma}\cdot p}{N_c(p;\theta)}\mathcal{F}\tilde{g})| + |k_1|\|f_-\|\|\tilde{g}\| \sup_{p\in\mathbb{R}^3}|\frac{c^2+E_c(p;\theta)}{N_c(p;\theta)}|.
\end{aligned}$$

Similarly to the proof of Theorem 1 we see that the supremum $\sup_{p\in\mathbb{R}^3}|\frac{c^2+E_c(p;\theta)}{N_c(p;\theta)}|$ is bounded independently of c and θ. Thus, Theorem 9 implies the claim for the second summand.
For the first summand, observe that $\sup_{p\in\mathbb{R}^3}|\frac{ce^{-\theta}}{N_c(p;\theta)}| \leq e^{\pi/4}/c$. Thus

$$\begin{aligned}
&|(f_+, k_1\cdot\boldsymbol{\sigma} e^{ik_2\cdot x}\mathcal{F}\frac{-ce^{-\theta}\boldsymbol{\sigma}\cdot p}{N_c(p;\theta)}\mathcal{F}^{-1}\tilde{g})| \\
=&|(\boldsymbol{\sigma}\cdot(-\mathrm{i}\nabla)k_1\cdot\boldsymbol{\sigma} e^{-ik_2\cdot x}f_+, \mathcal{F}\frac{-ce^{-\theta}}{N_c(p;\theta)}\mathcal{F}^{-1}\tilde{g})| \leq \frac{|k_1|e^{\pi/4}}{c}\|\nabla e^{ik_2\cdot x}f_+\|\|\tilde{g}\|.
\end{aligned}$$

Theorem 1 yields $\|\tilde{g}\| \leq \sqrt{1 + C_{\mathrm{FW}}|\mathrm{Im}\,\theta|}\|g\|$, which shows

$$|(f, P_{n,l}(c,\gamma;\theta)k_1\boldsymbol{\alpha}e^{ik_2\cdot x}\Lambda_{c,0}^{(\pm)}(\theta)g)| \leq C\|f\|\|g\| \tag{87}$$

for some $C > 0$, if one takes Theorem 9 and Theorem 10 into account .
Now, pick $f, g \in L^2(\mathbb{R}^3;\mathbb{C}^4)$ arbitrarily and apply (87) to the functions $P_{n,l}(c,\gamma;\bar{\theta})f$ and $\Lambda_{c,0}^{(\pm)}(\theta)g$. This implies the claim together with Corollary 5 and Lemma 4. □

A Some Estimates taken from Bach, Fröhlich, and Sigal [5]

In this appendix we quote some results from [5] which we need for the investigation of the non-relativistic limit in Section 8. We quote the result only in the generality which we need here and would like to mention that it also holds for suitable multi-particle Schrödinger operators.
We define

$$H_\gamma(\theta) := -\frac{e^{-2\theta}}{2}\Delta - \gamma V(\theta) \tag{88}$$

as operator on $L^2(\mathbb{R}^3;\mathbb{C}^2)$ and pick some eigenvalue $E_{\tilde{n}}$. We define (with $r > 0$ small enough) $P_{\mathrm{el},n'}(\gamma;\theta) := -(2\pi\mathrm{i}\,)^{-1}\int_{|E_{n'}-z|=r}(H_\gamma(\theta) - z)^{-1}\mathrm{d}z$ as projection onto the eigenspace of $H_\gamma(\theta)$ with eigenvalue $E_{n'}$. We abbreviate $\overline{P}_{n'}(\gamma;\theta) := 1 - P_{n'}(\theta)$. For $\eta > 0$ with $E_{\tilde{n}} < -\eta < E_{\tilde{n}+1}$ we define $P_{\mathrm{disc}}(\gamma;\theta) := \sum_{i:E_i\leq-\eta} P_i(\gamma;\theta)$ and $\overline{P}_{\mathrm{disc}}(\gamma;\theta) := 1 - P_{\mathrm{disc}}(\gamma;\theta)$.
In the following, we pick a sufficiently small $\theta_0 > 0$.

Lemma A.1 ([5], Corollary 1.4.). *There is a constant $C > 0$ such that for all $|\theta| < \theta_0$ the estimate $\|[H_\gamma(\theta) - H_\gamma(0)](H_\gamma(0) \pm \mathrm{i}\,)^{-1}\| \leq C|\theta|$ holds.*

Lemma (A.1) implies

Corollary A.1 ([5], Equation (3.79)). *There is a $C > 0$ such that for all $|\theta| < \theta_0$ the estimate $\|P_n(\gamma;\theta) - P_n(\gamma;0)\| \leq C|\theta|$ holds. The same estimate is true if one replaces P_n with P_{disc}.*

Using Lemma A.1 and Corollary A.1 as well as a resolvent expansion one shows

Theorem A.1 ([5], Lemma 3.8.). *Let $z \in \mathbb{C}$ with $\mathrm{Re}\,z < \Sigma - \eta$. Then the operator $H_\gamma(\theta) - z$ is invertible on $\mathrm{Ran}\,\bar{P}_{\mathrm{disc}}(\gamma;\theta)$ for sufficiently small $|\theta|(1 + (-\eta - \mathrm{Re}\,z)^{-1})$ and the estimate*

$$\left\|(H_\gamma(\theta)|_{\bar{P}_{\mathrm{disc}}(\gamma;\theta)} - z)^{-1}\bar{P}_{\mathrm{disc}}(\gamma;\theta)\right\| \leq 2(-\eta - \mathrm{Re}\,z)^{-1}$$

holds.

Acknowledgements. *The author would like to thank his thesis advisor Heinz Siedentop for suggesting the investigation of the lifetime of relativistic excited states and constantly surveying the progress of the work, which is part of his*

PhD thesis. He would like to thank Ira Herbst, David Hasler, Hubert Kalf and Tomio Umeda for interesting conversations, Christiane Tretter for pointing out reference [16], and the referee of a previous version of the paper for useful comments. He acknowledges support by the Deutsche Forschungsgemeinschaft (DFG), grant no. SI 348/12-2, as well as by a "Doktorandenstipendium" from the German Academic Exchange Service (DAAD).

References

[1] W. K. Abou Salem, J. Faupin, J. Froehlich, and I. M. Sigal. On the Theory of Resonances in Non-Relativistic QED and Related models. *ArXiv e-prints*, 711, November 2007.

[2] J. Aguilar and J. M. Combes. A class of analytic perturbations for one-body Schrödinger Hamiltonians. *Comm. Math. Phys.*, 22:269–279, 1971.

[3] Asao Arai and Masao Hirokawa. Ground States of a General Class of Quantum Field Hamiltonians. *Rev. Math. Phys.*, 12(8):1085–1135, 2000.

[4] Volker Bach, Jürg Fröhlich, and Israel Michael Sigal. Quantum electrodynamics of confined nonrelativistic particles. *Adv. Math.*, 137(2):299–395, 1998.

[5] Volker Bach, Jürg Fröhlich, and Israel Michael Sigal. Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field. *Comm. Math. Phys.*, 207(2):249–290, 1999.

[6] E. Balslev and J. M. Combes. Spectral properties of many-body Schrödinger operators with dilatation-analytic interactions. *Comm. Math. Phys.*, 22:280–294, 1971.

[7] E. Balslev and B. Helffer. Limiting absorption principle and resonances for the Dirac operator. *Adv. Appl. Math.*, 13(2):186–215, 1992.

[8] Erik Balslev. Resonances, resonance functions and spectral deformations. In *Resonances—models and phenomena (Bielefeld, 1984)*, volume 211 of *Lecture Notes in Phys.*, pages 27–63. Springer, Berlin, 1984.

[9] Jan Dereziński and Christian Gérard. Scattering theory of infrared divergent Pauli-Fierz Hamiltonians. *Ann. Henri Poincaré*, 5(3):523–577, 2004.

[10] M. J. Esteban and E. Séré. Nonrelativistic limit of the Dirac-Fock equations. *Ann. Henri Poincaré*, 2(5):941–961, 2001.

[11] William Desmond Evans, Peter Perry, and Heinz Siedentop. The spectrum of relativistic one-electron atoms according to Bethe and Salpeter. *Comm. Math. Phys.*, 178(3):733–746, July 1996.

[12] Hiroshima F. and Spohn H. Two-fold degeneracy of the ground state band for the pauli-fierz model with spin. *Adv. Theor.Math. Phys.*, page to appear, 2002.

[13] Enrico Fermi. *Nuclear Physics. A course given at the University of Chicago. - Notes compiled by Jay Orear, A. H. Rosenfeld and R. A. Schluter. Revised edition.* Chicago: University of Chicago Press; London: Cambridge University Press., 1950.

[14] K. O. Friedrichs. On the perturbation of continuous spectra. *Communications on Appl. Math.*, 1:361–406, 1948.

[15] J. Fröhlich, M. Griesemer, and B. Schlein. Rayleigh scattering at atoms with dynamical nuclei. *Comm. Math. Phys.*, 271(2):387–430, 2007.

[16] Israel Gohberg, Seymour Goldberg, and Marinus A. Kaashoek. *Classes of Linear Operators. Vol. I*, volume 49 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1990.

[17] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, 4th edition, 1980.

[18] M. Griesemer, E. Lieb, and M. Loss. Ground states in non-relativistic quantum electrodynamics. *Inv. Math.*, 145:557–595, 2001.

[19] Marcel Griesemer, Roger T. Lewis, and Heinz Siedentop. A minimax principle for eigenvalues in spectral gaps: Dirac operators with Coulomb potential. *Doc. Math.*, 4:275–283, 1999.

[20] D. Hasler and I. Herbst. Absence of Ground States for a Class of Translation Invariant Models of Non-relativistic QED. *Comm. Math. Phys.*, 279(3):769–787, may 2008.

[21] D. Hasler and I. Herbst. On the self-adjointness and domain of Pauli-Fierz type Hamiltonians. *Rev. Math. Phys.*, 20(7):787–800, 2008.

[22] D. Hasler, I. Herbst, and M. Huber. On the Lifetime of Quasi-Stationary States in Non-Relativistic QED. *Ann. Henri Poincaré*, 9(5):1005 – 1028, 2008.

[23] I. Herbst. Temporal exponential decay for the Stark effect in atoms. *J. Funct. Anal.*, 48(2):224–251, 1982.

[24] Ira W. Herbst. Dilation analyticity in constant electric field. I. The two body problem. *Comm. Math. Phys.*, 64(3):279–298, 1979.

[25] Ira W. Herbst and B. Simon. Dilation analyticity in constant electric field. II. N-body problem, Borel summability. *Comm. Math. Phys.*, 80(2):181–216, 1981.

[26] Fumio Hiroshima. Ground states of a model in nonrelativistic quantum electrodynamics. II. *J. Math. Phys.*, 41(2):661–674, 2000.

[27] James S. Howland. Perturbation of embedded eigenvalues. *Bull. Amer. Math. Soc.*, 78:280–283, 1972.

[28] James S. Howland. Puiseux series for resonances at an embedded eigenvalue. *Pacific J. Math.*, 55:157–176, 1974.

[29] Matthias Huber. *Modelle relativistischer und nicht-relativistischer Coulomb-Systeme [Dissertation an der Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München]*. Logos-Verlag Berlin GmbH, 2008.

[30] Matthias Huber. Spectral analysis of relativistic atoms – interaction with the quantized radiation field. *In preparation.*, 2009.

[31] Matthias Huber and Edgardo Stockmeyer. Perturbative implementation of the Furry picture. *Lett. Math. Phys.*, 79(1):99–108, 2007.

[32] Walter Hunziker. Resonances, metastable states and exponential decay laws in perturbation theory. *Comm. Math. Phys.*, 132(1):177–188, 1990.

[33] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1 edition, 1966.

[34] L.D. Landau and E.M. Lifschitz. *Lehrbuch der theoretischen Physik. Band 3: Quantenmechanik. 9., bearb. Aufl.* Frankfurt am Main: H. Deutsch., 1992.

[35] L.D. Landau, E.M. Lifschitz, V.B. Berestetskij, and L.P. Pitaevskij. *Lehrbuch der theoretischen Physik. Band 4: Quantenelektrodynamik. 7., bearb. Aufl.* Frankfurt am Main: H. Deutsch., 1991.

[36] M. S. Livšic. The method of non-selfadjoint operators in scattering theory. *Uspehi Mat. Nauk (N.S.)*, 12(1(73)):212–218, 1957.

[37] Michael Loss, Tadahiro Miyao, and Herbert Spohn. Lowest energy states in nonrelativistic QED: atoms and ions in motion. *J. Funct. Anal.*, 243(2):353–393, 2007.

[38] Tadahiro Miyao and Herbert Spohn. Spectral analysis of the semi-relativistic pauli-fierz hamiltonian. *Journal of Functional Analysis*, In Press, Corrected Proof:–, 2008.

[39] Sergey Morozov. Extension of a minimax principle for Coulomb-Dirac operators. Master's thesis, Mathematisches Institut, Ludwig-Maximilians-Universität, Theresienstr. 39, 80333 München, Germany, August 2004.

[40] G. Nenciu. Self-adjointness and invariance of the essential spectrum for Dirac operators defined as quadratic forms. *Comm. Math. Phys.*, 48:235–247, 1976.

[41] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I. Functional analysis.* Academic Press, New York, 1972.

[42] Michael Reed and Barry Simon. *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness.* Academic Press, New York, 1975.

[43] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics*, volume 4: Analysis of Operators. Academic Press, New York, 1 edition, 1978.

[44] Michael Reed and Barry Simon. *Methods of modern mathematical physics. III.* Academic Press, New York, 1979. Scattering theory.

[45] Julian Schwinger. Field theory of unstable particles. *Ann. Phys.*, 9:169–193, 1960.

[46] Petr Šeba. The complex scaling method for Dirac resonances. *Lett. Math. Phys.*, 16(1):51–59, 1988.

[47] Heinz Siedentop and Edgardo Stockmeyer. The Douglas-Kroll-Heß method: Convergence and block-diagonalization of Dirac operators. *Ann. Henri Poincaré*, 7(1):45–58, 2006.

[48] Barry Simon. Resonances in n-body quantum systems with dilatation analytic potentials and the foundations of time-dependent perturbation theory. *Ann. of Math. (2)*, 97:247–274, 1973.

[49] Bernd Thaller. *The Dirac Equation.* Texts and Monographs in Physics. Springer-Verlag, Berlin, 1 edition, 1992.

[50] R. A. Weder. Spectral properties of the Dirac Hamiltonian. *Ann. Soc. Sci. Bruxelles Sér. I*, 87:341–355, 1973.

[51] Ricardo Weder. Spectral analysis of pseudodifferential operators. *J. Funct. Anal.*, 20:319–337, 1975.

[52] Joachim Weidmann. *Lineare Operatoren in Hilberträumen. Teil 1.* Mathematische Leitfäden. [Mathematical Textbooks]. B. G. Teubner, Stuttgart, 2000. Grundlagen. [Foundations].

[53] V. Weißkopf and E. Wigner. Über die natürliche Linienbreite in der Strahlung des harmonischen Oszillators. *Z. f. Physik*, 65:18–29, 1930.

Mathias Huber
Fakultät für Mathematik
und Informatik
FernUniversität in Hagen
Lützowstr. 125
58095 Hagen
Germany
Matthias.Huber@gmx.de

Homology of the Steinberg Variety and Weyl Group Coinvariants

J. Matthew Douglass and Gerhard Röhrle [1]

Received: September 1, 2008
Revised: July 31, 2009

Communicated by Wolfgang Soergel

Abstract. Let G be a complex, connected, reductive algebraic group with Weyl group W and Steinberg variety Z. We show that the graded Borel-Moore homology of Z is isomorphic to the smash product of the coinvariant algebra of W and the group algebra of W.

2000 Mathematics Subject Classification: Primary 20G05; Secondary 20F55
Keywords and Phrases: Borel-Moore Homology, Steinberg Variety, Coinvariant algebra, Weyl group

1. Introduction

Suppose G is a complex, reductive algebraic group, $\mathcal{B}$ is the variety of Borel subgroups of G. Let $\mathfrak{g}$ be the Lie algebra of G and $\mathcal{N}$ the cone of nilpotent elements in $\mathfrak{g}$. Let $T^*\mathcal{B}$ denote the cotangent bundle of $\mathcal{B}$. Then there is a *moment map*, $\mu_0\colon T^*\mathcal{B} \to \mathcal{N}$. The *Steinberg variety* of G is the fibered product $T^*\mathcal{B} \times_{\mathcal{N}} T^*\mathcal{B}$ which we will identify with the closed subvariety

$$Z = \{\, (x, B', B'') \in \mathcal{N} \times \mathcal{B} \times \mathcal{B} \mid x \in \operatorname{Lie}(B') \cap \operatorname{Lie}(B'') \,\}$$

of $\mathcal{N} \times \mathcal{B} \times \mathcal{B}$. Set $n = \dim \mathcal{B}$. Then Z is a $2n$-dimensional, complex algebraic variety.

If $V = \oplus_{i\geq 0} V_i$ is a graded vector space, we will frequently denote V by $V_\bullet$. Similarly, if X is a topological space, then $H_i(X)$ denotes the i^{th} rational Borel-Moore homology of X and $H_\bullet(X) = \oplus_{i\geq 0} H_i(X)$ denotes the total Borel-Moore homology of X.

Fix a maximal torus, T, of G, with Lie algebra $\mathfrak{t}$, and let $W = N_G(T)/T$ be the Weyl group of (G, T). In [6] Kazhdan and Lusztig defined an action of $W \times W$ on $H_\bullet(Z)$ and they showed that the representation of $W \times W$ on the

[1]The authors would like to thank their charming wives for their unwavering support during the preparation of this paper

top-dimensional homology of Z, $H_{4n}(Z)$, is equivalent to the two-sided regular representation of W. Tanisaki [11] and, more recently, Chriss and Ginzburg [3] have strengthened the connection between $H_\bullet(Z)$ and W by defining a $\mathbb{Q}$-algebra structure on $H_\bullet(Z)$ so that $H_i(Z) * H_j(Z) \subseteq H_{i+j-4n}(Z)$. Chriss and Ginzburg [3, §3.4] have also given an elementary construction of an isomorphism between $H_{4n}(Z)$ and the group algebra $\mathbb{Q}W$.

Let Z_1 denote the "diagonal" in Z:

$$Z_1 = \{\, (x, B', B') \in \mathcal{N} \times \mathcal{B} \times \mathcal{B} \mid x \in \mathrm{Lie}(B') \,\}.$$

In this paper we extend the results of Chriss and Ginzburg [3, §3.4] and show in Theorem 2.3 that for any i, the convolution product defines an isomorphism $H_i(Z_1) \otimes H_{4n}(Z) \xrightarrow{\sim} H_i(Z)$. It then follows easily that with the convolution product, $H_\bullet(Z)$ is isomorphic to the smash product of the coinvariant algebra of W and the group algebra of W.

Precisely, for $0 \leq i \leq n$ let $\mathrm{Coinv}_{2i}(W)$ denote the degree i subspace of the rational coinvariant algebra of W, so $\mathrm{Coinv}_{2i}(W)$ may be identified with the space of degree i, W-harmonic polynomials on $\mathfrak{t}$. If j is odd, define $\mathrm{Coinv}_j(W) = 0$. Recall that the smash product, $\mathrm{Coinv}(W)\#\mathbb{Q}W$, is the $\mathbb{Q}$-algebra whose underlying vector space is $\mathrm{Coinv}(W) \otimes_{\mathbb{Q}} \mathbb{Q}W$ with multiplication satisfying $(f_1 \otimes \phi_1)\cdot(f_2 \otimes \phi_2) = f_1\phi_1(f_2) \otimes \phi_1\phi_2$ where f_1 and f_2 are in $\mathrm{Coinv}(W)$, ϕ_1 and ϕ_2 are in $\mathbb{Q}W$, and $\mathbb{Q}W$ acts on $\mathrm{Coinv}(W)$ in the usual way. The algebra $\mathrm{Coinv}(W)\#\mathbb{Q}W$ is graded by $(\mathrm{Coinv}(W)\#\mathbb{Q}W)_i = \mathrm{Coinv}_i(W)\#\mathbb{Q}W$ and we will denote this graded algebra by $\mathrm{Coinv}_\bullet(W)\#\mathbb{Q}W$. In Theorem 2.5 we construct an explicit isomorphism of graded algebras $H_{4n-\bullet}(Z) \cong \mathrm{Coinv}_\bullet(W)\#\mathbb{Q}W$.

This paper was motivated by the observation, pointed out to the first author by Catharina Stroppel, that the argument in [3, 8.1.5] can be used to show that $H_\bullet(Z)$ is isomorphic to the smash product of $\mathbb{Q}W$ and $\mathrm{Coinv}_\bullet(W)$. The details of such an argument have been carried out in a recent preprint of Namhee Kwon [8]. This argument relies on some deep and technical results: the localization theorem in K-theory proved by Thomason [12], the bivariant Riemann-Roch Theorem [3, 5.11.11], and the Kazhdan-Lusztig isomorphism between the equivariant K-theory of Z and the extended, affine, Hecke algebra [7]. In contrast, and also in the spirit of Kazhdan and Lusztig's original analysis of $H_{4n}(Z)$, and the analysis of $H_{4n}(Z)$ in [3, 3.4], our argument uses more elementary notions and is accessible to readers who are not experts in equivariant K-theory and to readers who are not experts in the representation theory of reductive, algebraic groups.

Another approach to the Borel-Moore homology of the Steinberg variety uses intersection homology. Let $\mu\colon Z \to \mathcal{N}$ be projection on the first factor. Then, as in [3, §8.6], $H_\bullet(Z) \cong \mathrm{Ext}^{4n-\bullet}_{D(\mathcal{N})}(R\mu_*\mathbb{Q}_{\mathcal{N}}, R\mu_*\mathbb{Q}_{\mathcal{N}})$. The Decomposition Theorem of Beilinson, Bernstein, and Deligne can be used to decompose $R\mu_*\mathbb{Q}_{\mathcal{N}}$ into a direct sum of simple perverse sheaves $R\mu_*\mathbb{Q}_{\mathcal{N}} \cong \oplus_{x,\phi} \mathrm{IC}_{x,\phi}^{n_{x,\phi}}$ where x runs over a set of orbit representatives in $\mathcal{N}$, for each x, ϕ runs over a set of irreducible representations of the component group of $Z_G(x)$, and $\mathrm{IC}_{x,\phi}$ denotes an intersection complex (see [2] or [9, §4,5]). Chriss and Ginzburg have used this construction to describe an isomorphism $H_{4n}(Z) \cong \mathbb{Q}W$ and to in addition give a description of the projective, indecomposable $H_\bullet(Z)$-modules.

It follows from Theorem 2.3 that $H_i(Z) \cong \mathrm{Coinv}_{4n-i}(W) \otimes H_{4n}(Z)$ and so

$$\begin{aligned} \mathrm{Coinv}_i(W) \otimes H_{4n}(Z) &\cong \mathrm{Ext}^{4n-i}_{D(\mathcal{N})}\left(R\mu_*\mathbb{Q}_{\mathcal{N}}, R\mu_*\mathbb{Q}_{\mathcal{N}}\right) \\ &\cong \bigoplus_{x,\phi}\bigoplus_{y,\psi} \mathrm{Ext}^{4n-i}_{D(\mathcal{N})}\left(\mathrm{IC}^{n_{x,\phi}}_{x,\phi}, \mathrm{IC}^{n_{y,\psi}}_{y,\psi}\right). \end{aligned} \tag{1.1}$$

In the special case when $i = 0$ we have that

$$\mathrm{Coinv}_0(W) \otimes H_{4n}(Z) \cong \mathrm{End}_{D(\mathcal{N})}\left(R\mu_*\mathbb{Q}_{\mathcal{N}}\right) \cong \bigoplus_{x,\phi} \mathrm{End}_{D(\mathcal{N})}\left(\mathrm{IC}^{n_{x,\phi}}_{x,\phi}\right).$$

The image of the one-dimensional vector space $\mathrm{Coinv}_0(W)$ in $\mathrm{End}_{D(\mathcal{N})}\left(R\mu_*\mathbb{Q}_{\mathcal{N}}\right)$ is the line through the identity endomorphism and $\mathbb{Q}W \cong H_{4n}(Z) \cong \oplus_{x,\phi} \mathrm{End}_{D(\mathcal{N})}\left(\mathrm{IC}^{n_{x,\phi}}_{x,\phi}\right)$ is the Wedderburn decomposition of $\mathbb{Q}W$ as a direct sum of minimal two-sided ideals. For $i > 0$ we have not been able to find a nice description of the image of $\mathrm{Coinv}_i(W)$ in the right-hand side of (1.1).

The rest of this paper is organized as follows: in §2 we set up our notation and state the main results; in §3 we construct an isomorphism of graded vector spaces between $\mathrm{Coinv}_\bullet(W) \otimes \mathbb{Q}W$ and $H_{4n-\bullet}(Z)$; and in §4 we complete the proof that this isomorphism is in fact an algebra isomorphism when $\mathrm{Coinv}_\bullet(W) \otimes \mathbb{Q}W$ is given the smash product multiplication. Some very general results about graphs and convolution that we need for the proofs of the main theorems are proved in an appendix.

In this paper $\otimes = \otimes_{\mathbb{Q}}$, if X is a set, then δ_X, or just δ, will denote the diagonal embedding of X in $X \times X$, and for g in G and x in $\mathfrak{g}$, $g \cdot x$ denotes the adjoint action of g on x.

2. Preliminaries and Statement of Results

Fix a Borel subgroup, B, of G with $T \subseteq B$ and define U to be the unipotent radical of B. We will denote the Lie algebras of B and U by $\mathfrak{b}$ and $\mathfrak{u}$ respectively.

Our proof that $H_\bullet(Z)$ is isomorphic to $\mathrm{Coinv}_\bullet(W)\#\mathbb{Q}W$ makes use of the specialization construction used by Chriss and Ginzburg in [3, §3.4] to establish the isomorphism between $H_{4n}(Z)$ and $\mathbb{Q}W$. We begin by reviewing their construction.

The group G acts diagonally on $\mathcal{B} \times \mathcal{B}$. Let $\mathcal{O}_w$ denote the orbit containing (B, wBw^{-1}). Then the rule $w \mapsto \mathcal{O}_w$ defines a bijection between W and the set of G-orbits in $\mathcal{B} \times \mathcal{B}$.

Let $\pi_Z \colon Z \to \mathcal{B} \times \mathcal{B}$ denote the projection on the second and third factors and for w in W define $Z_w = \pi_Z^{-1}(\mathcal{O}_w)$. For w in W we also set $\mathfrak{u}_w = \mathfrak{u} \cap w \cdot \mathfrak{u}$. The following facts are well-known (see [10] and [9, §1.1]):

- $Z_w \cong G \times^{B \cap {}^wB} \mathfrak{u}_w$.

- $\dim Z_w = 2n$.
- The set $\{\, \overline{Z_w} \mid w \in W \,\}$ is the set of irreducible components of Z.

Define

$$\begin{aligned}\widetilde{\mathfrak{g}} &= \{\, (x, B') \in \mathfrak{g} \times \mathcal{B} \mid x \in \operatorname{Lie}(B') \,\},\\ \widetilde{\mathcal{N}} &= \{\, (x, B') \in \mathcal{N} \times \mathcal{B} \mid x \in \operatorname{Lie}(B') \,\}, \text{ and}\\ \widehat{Z} &= \{\, (x, B', B'') \in \mathfrak{g} \times \mathcal{B} \times \mathcal{B} \mid x \in \operatorname{Lie}(B') \cap \operatorname{Lie}(B'') \,\},\end{aligned}$$

and let $\mu\colon \widetilde{\mathfrak{g}} \to \mathfrak{g}$ denote the projection on the first factor. Then $\widetilde{\mathcal{N}} \cong T^*\mathcal{B}$, $\mu(\widetilde{\mathcal{N}}) = \mathcal{N}$, $Z \cong \widetilde{\mathcal{N}} \times_{\mathcal{N}} \widetilde{\mathcal{N}}$, and $\widehat{Z} \cong \widetilde{\mathfrak{g}} \times_{\mathfrak{g}} \widetilde{\mathfrak{g}}$.
Let $\hat{\pi}\colon \widehat{Z} \to \mathcal{B} \times \mathcal{B}$ denote the projection on the second and third factors and for w in W define $\widehat{Z}_w = \hat{\pi}^{-1}(\mathcal{O}_w)$. Then it is well-known that $\dim \widehat{Z}_w = \dim \mathfrak{g}$ and that the closures of the $\widehat{Z}_w$'s for w in W are the irreducible components of $\widehat{Z}$ (see [9, §1.1]).
Next, for (x, gBg^{-1}) in $\widetilde{\mathfrak{g}}$, define $\nu(x, gBg^{-1})$ to be the projection of $g^{-1} \cdot x$ in $\mathfrak{t}$. Then μ and ν are two of the maps in Grothendieck's simultaneous resolution:

$$\begin{tikzcd} \widetilde{\mathfrak{g}} \arrow[r, "\mu"] \arrow[d, "\nu"'] & \mathfrak{g} \arrow[d] \\ \mathfrak{t} \arrow[r] & \mathfrak{t}/W \end{tikzcd}$$

It is easily seen that if $\widehat{\mu}\colon \widehat{Z} \to \mathfrak{g}$ is the projection on the first factor, then the square

$$\begin{tikzcd} \widehat{Z} \arrow[r, "\widehat{\mu}"] \arrow[d] & \mathfrak{g} \arrow[d, "\delta_{\mathfrak{g}}"] \\ \widetilde{\mathfrak{g}} \times \widetilde{\mathfrak{g}} \arrow[r, "\mu\times\mu"'] & \mathfrak{g} \times \mathfrak{g} \end{tikzcd}$$

is cartesian, where the vertical map on the left is given by $(x, B', B'') \mapsto ((x, B'), (x, B''))$. We will frequently identify $\widehat{Z}$ with the subvariety of $\widetilde{\mathfrak{g}} \times \widetilde{\mathfrak{g}}$ consisting of all pairs $((x, B'), (x, B''))$ with x in $\operatorname{Lie}(B') \cap \operatorname{Lie}(B'')$.
For w in W, let $\Gamma_{w^{-1}} = \{\, (h, w^{-1} \cdot h) \mid h \in \mathfrak{t} \,\} \subseteq \mathfrak{t} \times \mathfrak{t}$ denote the graph of the action of w^{-1} on $\mathfrak{t}$ and define

$$\Lambda_w = \widehat{Z} \cap (\nu \times \nu)^{-1} (\Gamma_{w^{-1}}) = \{\, (x, B', B'') \in \widehat{Z} \mid \nu(x, B'') = w^{-1}\nu(x, B') \,\}.$$

In the special case when w is the identity element in W, we will denote Λ_w by Λ_1.

The spaces we have defined so far fit into a commutative diagram with cartesian squares:

(2.1)
$$\begin{tikzcd} \Lambda_w \arrow[r] \arrow[d] & \widehat{Z} \arrow[r, "\widehat{\mu}"] \arrow[d] & \mathfrak{g} \arrow[d, "\delta_{\mathfrak{g}}"] \\ (\nu\times\nu)^{-1}\left(\Gamma_{w^{-1}}\right) \arrow[r] \arrow[d] & \widetilde{\mathfrak{g}}\times\widetilde{\mathfrak{g}} \arrow[r, "\mu\times\mu"'] \arrow[d, "\nu\times\nu"'] & \mathfrak{g}\times\mathfrak{g} \\ \Gamma_{w^{-1}} \arrow[r] & \mathfrak{t}\times\mathfrak{t} & \end{tikzcd}$$

Let $\nu_w\colon \Lambda_w \to \Gamma_{w^{-1}}$ denote the composition of the leftmost vertical maps in (2.1), so ν_w is the restriction of $\nu\times\nu$ to Λ_w.
For the specialization construction, we consider subsets of $\widehat{Z}$ of the form $\nu_w^{-1}(S')$ for $S' \subseteq \Gamma_{w^{-1}}$. Thus, for h in $\mathfrak{t}$ we define $\Lambda_w^h = \nu_w^{-1}(h, w^{-1}h)$. Notice in particular that $\Lambda_w^0 = Z$. More generally, for a subset S of $\mathfrak{t}$ we define $\Lambda_w^S = \coprod_{h\in S}\Lambda_w^h$. Then, $\Lambda_w^S = \nu_w^{-1}(S')$, where S' is the graph of w^{-1} restricted to S. Let $\mathfrak{t}_{\mathrm{reg}}$ denote the set of regular elements in $\mathfrak{t}$.
Fix a one-dimensional subspace, ℓ, of $\mathfrak{t}$ so that $\ell\cap\mathfrak{t}_{\mathrm{reg}} = \ell\setminus\{0\}$ and set $\ell^* = \ell\setminus\{0\}$. Then $\Lambda_w^\ell = \Lambda_w^{\ell^*}\coprod\Lambda_w^0 = \Lambda_w^{\ell^*}\coprod Z$. We will see in Corollary 3.6 that the restriction of ν_w to $\Lambda_w^{\ell^*}$ is a locally trivial fibration with fibre G/T. Thus, using a construction due to Fulton and MacPherson ([4, §3.4], [3, §2.6.30]), there is a specialization map

$$\lim\colon H_{\bullet+2}(\Lambda_w^{\ell^*}) \longrightarrow H_\bullet(Z).$$

Since $\Lambda_w^{\ell^*}$ is an irreducible, $(2n+1)$-dimensional variety, if $[\Lambda_w^{\ell^*}]$ denotes the fundamental class of $\Lambda_w^{\ell^*}$, then $H_{4n+2}(\Lambda_w^{\ell^*})$ is one-dimensional with basis $\{[\Lambda_w^{\ell^*}]\}$. Define $\lambda_w = \lim([\Lambda_w^{\ell^*}])$ in $H_{4n}(Z)$. Chriss and Ginzburg [3, §3.4] have proved the following theorem.

Theorem 2.2. *Consider $H_\bullet(Z)$ endowed with the convolution product.*

(a) *For $0 \le i, j \le 4n$, $H_i(Z) * H_j(Z) \subseteq H_{i+j-4n}(Z)$. In particular, $H_{4n}(Z)$ is a subalgebra of $H_\bullet(Z)$.*

(b) *The element λ_w in $H_{4n}(Z)$ does not depend on the choice of ℓ.*

(c) *The assignment $w \mapsto \lambda_w$ extends to an algebra isomorphism $\alpha\colon \mathbb{Q}W \xrightarrow{\cong} H_{4n}(Z)$.*

Now consider

$$Z_1 = \{\, (x, B', B') \in \mathcal{N}\times\mathcal{B}\times\mathcal{B} \mid x \in \mathrm{Lie}(B') \,\}.$$

Then Z_1 may be identified with the diagonal in $\widetilde{\mathcal{N}}\times\widetilde{\mathcal{N}}$. It follows that Z_1 is closed in Z and isomorphic to $\widetilde{\mathcal{N}}$.

Since $\widetilde{\mathcal{N}} \cong T^*\mathcal{B}$, it follows from the Thom isomorphism in Borel-Moore homology [3, §2.6] that $H_{i+2n}(Z_1) \cong H_i(\mathcal{B})$ for all i. Since $\mathcal{B}$ is smooth and compact, $H_i(\mathcal{B}) \cong H^{2n-i}(\mathcal{B})$ by Poincaré duality. Therefore, $H_{4n-i}(Z_1) \cong H^i(\mathcal{B})$ for all i.

The cohomology of $\mathcal{B}$ is well-understood: there is an isomorphism of graded algebras, $H^\bullet(\mathcal{B}) \cong \mathrm{Coinv}_\bullet(W)$. It follows that $H_j(Z_1) = 0$ if j is odd and $H_{4n-2i}(Z_1) \cong \mathrm{Coinv}_{2i}(W)$ for $0 \leq i \leq n$.

In §3 below we will prove the following theorem.

Theorem 2.3. *Consider the Borel-Moore homology of the variety Z_1.*

(a) *There is a convolution product on $H_\bullet(Z_1)$. With this product, $H_\bullet(Z_1)$ is a commutative $\mathbb{Q}$-algebra and there is an isomorphism of graded $\mathbb{Q}$-algebras*

$$\beta\colon \mathrm{Coinv}_\bullet(W) \xrightarrow{\cong} H_{4n-\bullet}(Z_1).$$

(b) *If $r\colon Z_1 \to Z$ denotes the inclusion, then the direct image map in Borel-Moore homology, $r_*\colon H_\bullet(Z_1) \longrightarrow H_\bullet(Z)$, is an injective ring homomorphism.*

(c) *If we identify $H_\bullet(Z_1)$ with its image in $H_\bullet(Z)$ as in (b), then the linear transformation given by the convolution product*

$$H_i(Z_1) \otimes H_{4n}(Z) \xrightarrow{*} H_i(Z)$$

is an isomorphism of vector spaces for $0 \leq i \leq 4n$.

The algebra $\mathrm{Coinv}_\bullet(W)$ has a natural action of W by algebra automorphisms, and the isomorphism β in Theorem 2.3(a) is in fact an isomorphism of W-algebras. The W-algebra structure on $H_\bullet(Z_1)$ is described in the next theorem, which will be proved in §4.

Theorem 2.4. *If w is in W and $H_\bullet(Z_1)$ is identified with its image in $H_\bullet(Z)$, then*

$$\lambda_w * H_i(Z_1) * \lambda_{w^{-1}} = H_i(Z_1).$$

Thus, conjugation by λ_w defines a W-algebra structure on $H_\bullet(Z_1)$. With this W-algebra structure, the isomorphism $\beta\colon \mathrm{Coinv}_\bullet(W) \xrightarrow{\cong} H_{4n-\bullet}(Z_1)$ in Theorem 2.3(a) is an isomorphism of W-algebras.

Recall that the algebra $\mathrm{Coinv}(W)\#\mathbb{Q}W$ is graded by $(\mathrm{Coinv}(W)\#\mathbb{Q}W)_i = \mathrm{Coinv}_i(W) \otimes \mathbb{Q}W$. Then combining Theorem 2.2(c), Theorem 2.3(c), and Theorem 2.4 we get our main result.

Theorem 2.5. *The composition*

$$\mathrm{Coinv}_\bullet(W)\#\mathbb{Q}W \xrightarrow{\beta\otimes\alpha} H_{4n-\bullet}(Z_1) \otimes H_{4n}(Z) \xrightarrow{*} H_{4n-\bullet}(Z)$$

is an isomorphism of graded $\mathbb{Q}$-algebras.

3. Factorization of $H_\bullet(Z)$

Proof of Theorem 2.3(a). We need to prove that $H_\bullet(Z_1)$ is a commutative $\mathbb{Q}$-algebra and that $\mathrm{Coinv}_\bullet(W) \cong H_{4n-\bullet}(Z_1)$.

Let $\pi\colon \widetilde{\mathcal{N}} \to \mathcal{B}$ by $\pi(x, B') = B'$. Then π may be identified with the vector bundle projection $T^*\mathcal{B} \to \mathcal{B}$ and so the induced map in cohomology $\pi^*\colon H^i(\mathcal{B}) \to H^i(\widetilde{\mathcal{N}})$ is an isomorphism. The projection π determines an isomorphism in Borel-Moore homology that we will also denote by π^* (see [3, §2.6.42]). We have $\pi^*\colon H_i(\mathcal{B}) \xrightarrow{\cong} H_{i+2n}(\widetilde{\mathcal{N}})$.

For a smooth m-dimensional variety X, let $\mathrm{pd}\colon H^i(X) \longrightarrow H_{2m-i}(X)$ denote the Poincaré duality isomorphism. Then the composition

$$H_{2n-i}(\mathcal{B}) \xrightarrow{\mathrm{pd}^{-1}} H^i(\mathcal{B}) \xrightarrow{\pi^*} H^i(\widetilde{\mathcal{N}}) \xrightarrow{\mathrm{pd}} H_{4n-i}(\widetilde{\mathcal{N}})$$

is an isomorphism. It follows from the uniqueness construction in [3, §2.6.26] that

$$\mathrm{pd} \circ \pi^* \circ \mathrm{pd}^{-1} = \pi^*\colon H_{2n-i}(\mathcal{B}) \longrightarrow H_{4n-i}(\widetilde{\mathcal{N}})$$

and so $\pi^* \circ \mathrm{pd} = \mathrm{pd} \circ \pi^*\colon H^i(\mathcal{B}) \longrightarrow H_{4n-i}(\widetilde{\mathcal{N}})$.

Recall that $\mathrm{Coinv}_j(W) = 0$ if j is odd and $\mathrm{Coinv}_{2i}(W)$ is the degree i subspace of the coinvariant algebra of W. Let $\mathrm{bi}\colon \mathrm{Coinv}_\bullet(W) \longrightarrow H^\bullet(\mathcal{B})$ be the Borel isomorphism (see [1, §1.5] or [5]). Then with the cup product, $H^\bullet(\mathcal{B})$ is a graded algebra and bi is an isomorphism of graded algebras.

Define $\beta\colon \mathrm{Coinv}_i(W) \to H_{4n-i}(Z_1)$ to be the composition

$$\mathrm{Coinv}_i(W) \xrightarrow{\mathrm{bi}} H^i(\mathcal{B}) \xrightarrow{\pi^*} H^i(\widetilde{\mathcal{N}}) \xrightarrow{\mathrm{pd}} H_{4n-i}(\widetilde{\mathcal{N}}) \xrightarrow{\delta_*} H_{4n-i}(Z_1)$$

where $\delta = \delta_{\widetilde{\mathcal{N}}}$. Then β is an isomorphism of graded vector spaces and

$$\beta = \delta_* \circ \mathrm{pd} \circ \pi^* \circ \mathrm{bi} = \delta_* \circ \pi^* \circ \mathrm{pd} \circ \mathrm{bi}.$$

The algebra structure of $H^\bullet(\mathcal{B})$ and $H^\bullet(\widetilde{\mathcal{N}})$ is given by the cup product, and $\pi^*\colon H^\bullet(\mathcal{B}) \to H^\bullet(\widetilde{\mathcal{N}})$ is an isomorphism of graded algebras. Since $\widetilde{\mathcal{N}}$ is smooth, as in [3, §2.6.15], there is an intersection product defined on $H_\bullet(\widetilde{\mathcal{N}})$ using Poincaré duality and the cup product on $H^\bullet(\widetilde{\mathcal{N}})$. Thus, $\mathrm{pd}\colon H^\bullet(\widetilde{\mathcal{N}}) \to H_{4n-\bullet}(\widetilde{\mathcal{N}})$ is an algebra isomorphism. Finally, it is observed in [3, §2.7.10] that $\delta_*\colon H_\bullet(\widetilde{\mathcal{N}}) \to H_\bullet(Z_1)$ is a ring homomorphism and hence an algebra isomorphism. This shows that β is an isomorphism of graded algebras and proves Theorem 2.3(a).

Proof of Theorem 2.3(b). To prove the remaining parts of Theorem 2.3, we need a linear order on W. Suppose $|W| = N$. Fix a linear order on W that extends the Bruhat order. Say $W = \{w_1, \ldots, w_N\}$, where $w_1 = 1$ and w_N is the longest element in W.

For $1 \leq j \leq N$, define $Z_j = \coprod_{i=1}^{j} Z_{w_i}$. Then, for each j, Z_j is closed in Z, Z_{w_j} is open in Z_j, and $Z_j = Z_{j-1} \coprod Z_{w_j}$. Notice that $Z_N = Z$ and $Z_1 = Z_{w_1}$.

Similarly, define $\widehat{Z}_j = \coprod_{i=1}^{j} \widehat{Z}_{w_i}$. Then each $\widehat{Z}_j$ is closed in $\widehat{Z}$, $\widehat{Z}_{w_j}$ is open in $\widehat{Z}_j$, and $\widehat{Z}_j = \widehat{Z}_{j-1} \coprod \widehat{Z}_{w_j}$.

We need to show that $r_*\colon H_\bullet(Z_1) \longrightarrow H_\bullet(Z)$ is an injective ring homomorphism.
Let $\mathrm{res}_j\colon H_i(Z_j) \to H_i(Z_{w_j})$ denote the restriction map in Borel-Moore homology induced by the open embedding $Z_{w_j} \subseteq Z_j$ and let $r_j\colon H_i(Z_{j-1}) \longrightarrow H_i(Z_j)$ denote the direct image map in Borel-Moore homology induced by the closed embedding $Z_{j-1} \subseteq Z_j$. Then there is a long exact sequence in homology

$$\cdots \longrightarrow H_i(Z_{j-1}) \xrightarrow{r_j} H_i(Z_j) \xrightarrow{\mathrm{res}_j} H_i(Z_{w_j}) \xrightarrow{\partial} H_{i-1}(Z_{j-1}) \longrightarrow \cdots$$

It is shown in [3, §6.2] that $\partial = 0$ and so the sequence

$$0 \longrightarrow H_i(Z_{j-1}) \xrightarrow{r_j} H_i(Z_j) \xrightarrow{\mathrm{res}_j} H_i(Z_{w_j}) \longrightarrow 0 \tag{3.1}$$

is exact for every i and j. Therefore, if $r\colon Z_j \to Z$ denotes the inclusion, then the direct image $r_*\colon H_i(Z_j) \to H_i(Z)$ is an injection for all i. (The fact that r depends on j should not lead to any confusion.)
We will frequently identify $H_i(Z_j)$ with its image in $H_i(Z)$ and consider $H_i(Z_j)$ as a subset of $H_i(Z)$. Thus, we have a flag of subspaces $0 \subseteq H_i(Z_1) \subseteq \cdots \subseteq H_i(Z_{N-1}) \subseteq H_i(Z)$.
In particular, $r_*\colon H_i(Z_1) \to H_i(Z)$ is an injection for all i. It follows from [3, Lemma 5.2.23] that r_* is a ring homomorphism. This proves part (b) of Theorem 2.3.

Proof of Theorem 2.3(c). We need to show that the linear transformation given by the convolution product $H_i(Z_1) \otimes H_{4n}(Z) \to H_i(Z)$ is an isomorphism of vector spaces for $0 \leq i \leq 4n$.
The proof is a consequence of the following lemma.

Lemma 3.2. *The image of the convolution map* $*: H_i(Z_1) \otimes H_{4n}(Z_j) \longrightarrow H_i(Z)$ *is precisely* $H_i(Z_j)$ *for* $0 \leq i \leq 4n$ *and* $1 \leq j \leq N$.

Assuming that the lemma has been proved, taking $j = N$, we conclude that the convolution product in $H_\bullet(Z)$ induces a surjection $H_i(Z_1) \otimes H_{4n}(Z) \longrightarrow H_i(Z)$. It is shown in [3, §6.2] that $\dim H_\bullet(Z) = |W|^2$ and so $\dim H_\bullet(Z_1) \otimes H_{4n}(Z) = |W|^2 = \dim H_\bullet(Z)$. Thus, the convolution product induces an isomorphism $H_i(Z_1) \otimes H_{4n}(Z) \cong H_i(Z)$.
The rest of this section is devoted to the proof of Lemma 3.2.
To prove Lemma 3.2 we need to analyze the specialization map, $\lim\colon H_{\bullet+2}(\Lambda_w^{\ell^*}) \to H_\bullet(Z)$, beginning with the subvarieties Λ_w^{ℓ} and $\Lambda_w^{\ell^*}$ of Λ_w.

Subvarieties of Λ_w. Suppose that ℓ is a one-dimensional subspace of $\mathfrak{t}$ with $\ell^* = \ell \setminus \{0\} = \ell \cap \mathfrak{t}_{\mathrm{reg}}$. Recall that $\mathfrak{u}_w = \mathfrak{u} \cap w \cdot \mathfrak{u}$ for w in W.

Lemma 3.3. *The variety* $\Lambda_w^{\ell} \cap \widehat{Z}_w$ *is the* G*-saturation in* $\widehat{Z}$ *of the subset*

$$\{\, (h+n, B, wBw^{-1}) \mid h \in \ell,\, n \in \mathfrak{u}_w \,\}.$$

Proof. By definition,

$$\Lambda_w^\ell = \Lambda_w^{\ell^*} \coprod \Lambda_w^0 = \{\, (x, B', B'') \in \widehat{Z} \mid \nu(x, B'') = w^{-1}\nu(x, B') \in w^{-1}(\ell) \,\}.$$

Suppose that h is in $\mathfrak{t}_{\mathrm{reg}}$ and $(x, g_1Bg_1^{-1}, g_2Bg_2^{-1})$ is in Λ_w^h. Then $g_1^{-1}\cdot x = h+n_1$ and $g_2^{-1} \cdot x = w^{-1}h + n_2$ for some n_1 and n_2 in $\mathfrak{u}$. Since h is regular, there are elements u_1 and u_2 in U so that $u_1^{-1}g_1^{-1} \cdot h = h$ and $u_2^{-1}g_2^{-1} \cdot h = w^{-1}h$. Then $x = g_1u_1 \cdot h = g_2u_2w^{-1} \cdot h$ and so $g_1u_1 = g_2u_2w^{-1}t$ for some t in T. Therefore, $(x, g_1Bg_1^{-1}, g_2Bg_2^{-1}) = g_1u_1 \cdot (h, B, wBw^{-1})$. Thus, Λ_w^h is contained in the G-orbit of (h, B, wBw^{-1}). Since ν is G-equivariant, it follows that Λ_w^h is G-stable and so Λ_w^h is the full G-orbit of (h, B, wBw^{-1}). Therefore, $\Lambda_w^{\ell^*}$ is the G-saturation of $\{\, (h+n, B, wBw^{-1} \mid h \in \ell^*,\, n \in \mathfrak{u}_w \,\}$ and $\Lambda_w^h \subseteq \widehat{Z}_w$ for h in ℓ^*.
We have already observed that $\Lambda_w^0 = Z$ and so

$$\Lambda_w^\ell \cap \widehat{Z}_w = \left(\Lambda_w^{\ell^*} \cap \widehat{Z}_w\right) \coprod \left(\Lambda_w^0 \cap \widehat{Z}_w\right) = \Lambda_w^{\ell^*} \coprod Z_w.$$

It is easy to see that Z_w is the G-saturation of $\{\, (n, B, wBw^{-1}) \mid n \in \mathfrak{u}_w \,\}$ in Z. This proves the lemma. $\square$

Corollary 3.4. *The variety $\Lambda_w^\ell \cap \widehat{Z}_w$ is a locally trivial, affine space bundle over $\mathcal{O}_w$ with fibre isomorphic to $\ell + \mathfrak{u}_w$, and hence there is an isomorphism $\Lambda_w^\ell \cap \widehat{Z}_w \cong G \times^{B\cap {}^wB} (\ell + \mathfrak{u}_w)$.*

Proof. It follows from Lemma 3.3 that the map given by projection on the second and third factors is a G-equivariant morphism from Λ_w^ℓ onto $\mathcal{O}_w$ and that the fibre over (B, wBw^{-1}) is $\{\, (h+n, B, wBw^{-1}) \mid h \in \ell,\, n \in \mathfrak{u}_w \,\}$. Therefore, $\Lambda_w^\ell \cong G \times^{B\cap {}^wB} (\ell + \mathfrak{u}_w)$. $\square$

Let $\mathfrak{g}_{\mathrm{rs}}$ denote the set of regular semisimple elements in $\mathfrak{g}$ and define $\widetilde{\mathfrak{g}}_{\mathrm{rs}} = \{\, (x, B') \in \widetilde{\mathfrak{g}} \mid x \in \mathfrak{g}_{\mathrm{rs}} \,\}$. For an arbitrary subset S of $\mathfrak{t}$, define

$$\widetilde{\mathfrak{g}}^S = \nu^{-1}(S) = \{\, (x, B') \in \widetilde{\mathfrak{g}} \mid \nu(x, B') \in S \,\}.$$

For w in W, define $\widetilde{w}\colon G/T \times \mathfrak{t}_{\mathrm{reg}} \longrightarrow G/T \times \mathfrak{t}_{\mathrm{reg}}$ by $\widetilde{w}(gT, h) = (gwT, w^{-1}h)$. The rule $(gT, h) \mapsto (g \cdot h, gB)$ defines an isomorphism of varieties

$$f\colon G/T \times \mathfrak{t}_{\mathrm{reg}} \xrightarrow{\cong} \widetilde{\mathfrak{g}}_{\mathrm{rs}}$$

and we will denote the automorphism $f \circ \widetilde{w} \circ f^{-1}$ of $\widetilde{\mathfrak{g}}_{\mathrm{rs}}$ also by $\widetilde{w}$. Notice that if h is in $\mathfrak{t}_{\mathrm{reg}}$ and g is in G, then $\widetilde{w}(g \cdot h, gB) = (g \cdot h, gwBw^{-1}g^{-1})$.

Lemma 3.5. *The variety $\Lambda_w^{\ell^*}$ is the graph of $\widetilde{w}|_{\widetilde{\mathfrak{g}}^{\ell^*}} \colon \widetilde{\mathfrak{g}}^{\ell^*} \to \widetilde{\mathfrak{g}}^{w^{-1}(\ell^*)}$.*

Proof. It follows from Lemma 3.3 that

$$\begin{aligned} \Lambda_w^{\ell^*} &= \{\, (g \cdot h, gBg^{-1}, gwBw^{-1}g^{-1}) \in \mathfrak{g}_{\mathrm{rs}} \times \mathcal{B} \times \mathcal{B} \mid h \in \ell^*,\, g \in G \,\} \\ &= \{\, ((g \cdot h, gBg^{-1}), (g \cdot h, gwBw^{-1}g^{-1})) \in \widetilde{\mathfrak{g}} \times \widetilde{\mathfrak{g}} \mid h \in \ell^*,\, g \in G \,\}. \end{aligned}$$

The argument in the proof of Lemma 3.3 shows that

$$\widetilde{\mathfrak{g}}^{\ell^*} = \{\, (g \cdot h, gBg^{-1}) \mid h \in \ell^*,\, g \in G \,\}$$

and by definition $\widetilde{w}(g \cdot h, gB) = (g \cdot h, gwBw^{-1}g^{-1})$. Therefore, $\Lambda_w^{\ell^*}$ is the graph of $\widetilde{w}|_{\widetilde{\mathfrak{g}}^{\ell^*}}$. □

Corollary 3.6. *The map $\nu_w\colon \Lambda_w^{\ell^*} \to \ell^*$ is a locally trivial fibration with fibre isomorphic to G/T.*

Proof. This follows from the lemma and the fact that $\widetilde{\mathfrak{g}}^{\ell^*} \cong G/T \times \ell^*$. □

The specialization map. Suppose that w is in W and that ℓ is a one-dimensional subspace of $\mathfrak{t}$ with $\ell^* = \ell \setminus \{0\} = \ell \cap \mathfrak{t}_{\mathrm{reg}}$. As in [4] and [3, §2.6.30], $\lim\colon H_{i+2}(\Lambda_w^{\ell^*}) \to H_i(Z)$ is the composition of three maps, defined as follows.

As a vector space over $\mathbb{R}$, ℓ is two-dimensional. Fix an $\mathbb{R}$-basis of ℓ, say $\{v_1, v_2\}$. Define P to be the open half plane $\mathbb{R}_{>0}v_1 \oplus \mathbb{R}v_2$, define $I_{>0}$ to be the ray $\mathbb{R}_{>0}v_1$, and define I to be the closure of $I_{>0}$, so $I = \mathbb{R}_{\geq 0}v_1$.

Since P is an open subset of ℓ^*, Λ_w^P is an open subset of $\Lambda_w^{\ell^*}$ and so there is a restriction map in Borel-Moore homology $\mathrm{res}\colon H_{i+2}(\Lambda_w^{\ell^*}) \to H_{i+2}(\Lambda_w^P)$.

The projection map from P to $I_{>0}$ determines an isomorphism in Borel-Moore homology $\psi\colon H_{i+2}(\Lambda_w^P) \to H_{i+1}(\Lambda_w^{I_{>0}})$.

Since $I = I_{>0} \coprod \{0\}$, we have $\Lambda_w^I = \Lambda_w^{I_{>0}} \coprod \Lambda_w^0 = \Lambda_w^{I_{>0}} \coprod Z$, where Z is closed in Λ_w^I. The connecting homomorphism of the long exact sequence in Borel-Moore homology arising from the partition $\Lambda_w^I = \Lambda_w^{I_{>0}} \coprod Z$ is a map

$$\partial\colon H_{i+1}(\Lambda_w^{I_{>0}}) \to H_i(Z).$$

By definition, $\lim = \partial \circ \psi \circ \mathrm{res}$.

Now fix j with $1 \leq j \leq N$ and set $w = w_j$.

Consider the intersection $\Lambda_w^I \cap \widehat{Z}_j = (\Lambda_w^{I_{>0}} \cap \widehat{Z}_j) \coprod (Z \cap \widehat{Z}_j)$. Then $Z \cap \widehat{Z}_j$ is closed in $\Lambda_w^I \cap \widehat{Z}_j$ and by construction, $\Lambda_w^{I_{>0}} \subseteq \widehat{Z}_j$ and $Z \cap \widehat{Z}_j = Z_j$. Thus, $\Lambda_w^I \cap \widehat{Z}_j = \Lambda_w^{I_{>0}} \coprod Z_j$. Let $\partial_j\colon H_{i+1}(\Lambda_w^{I_{>0}}) \to H_i(Z_j)$ be the connecting homomorphism of the long exact sequence in Borel-Moore homology arising from this partition. Because the long exact sequence in Borel-Moore homology is natural, we have a commutative square:

$$\begin{tikzcd}
H_{i+1}(\Lambda_w^{I_{>0}}) \arrow[r, "\partial"] \arrow[d, equal] & H_i(Z) \\
H_{i+1}(\Lambda_w^{I_{>0}}) \arrow[r, "\partial_j"'] & H_i(Z_j) \arrow[u, "r_*"']
\end{tikzcd}$$

This proves the following lemma.

Lemma 3.7. *Fix j with $1 \leq j \leq N$ and set $w = w_j$. Then the map $\partial\colon H_{i+1}(\Lambda_w^{I_{>0}}) \longrightarrow H_i(Z)$ factors as $r_* \circ \partial_j$ where $\partial_j\colon H_{i+1}(\Lambda_w^{I_{>0}}) \longrightarrow H_i(Z_j)$ is the connecting homomorphism of the long exact sequence arising from the partition $\Lambda_w^I \cap \widehat{Z} = \Lambda_w^{I_{>0}} \coprod Z_j$.*

It follows from the lemma that $\lim\colon H_{i+2}(\Lambda_w^{\ell^*}) \longrightarrow H_i(Z)$ factors as

$$(3.8)\quad H_{i+2}(\Lambda_w^{\ell^*}) \xrightarrow{\text{res}} H_{i+2}(\Lambda_w^P) \xrightarrow{\psi} H_{i+1}(\Lambda_w^{I_{>0}}) \xrightarrow{\partial_j} H_i(Z_j) \xrightarrow{r_*} H_i(Z).$$

Define $\lim_j\colon H_{i+2}(\Lambda_w^{\ell^*}) \longrightarrow H_i(Z_j)$ by $\lim_j = \partial_j \circ \psi \circ \text{res}$.

Specialization and restriction. As above, fix j with $1 \leq j \leq N$ and a one-dimensional subspace ℓ of $\mathfrak{t}$ with $\ell^* = \ell \setminus \{0\} = \ell \cap \mathfrak{t}_{\text{reg}}$. Set $w = w_j$. Recall the restriction map $\text{res}_j\colon H_i(Z_j) \to H_i(Z_w)$ from (3.1).

Lemma 3.9. *The composition* $\text{res}_j \circ \lim_j\colon H_{i+2}(\Lambda_w^{\ell^*}) \longrightarrow H_i(Z_w)$ *is surjective for* $0 \leq i \leq 4n$.

Proof. Using (3.8), $\text{res}_j \circ \lim_j$ factors as

$$H_{i+2}(\Lambda_w^{\ell^*}) \xrightarrow{\text{res}} H_{i+2}(\Lambda_w^P) \xrightarrow{\psi} H_{i+1}(\Lambda_w^{I_{>0}}) \xrightarrow{\partial_j} H_i(Z_j) \xrightarrow{\text{res}_j} H_i(Z_w).$$

Lemma 3.11 below shows that res is always surjective and the map ψ is an isomorphism, so we need to show that the composition $\text{res}_j \circ \partial_j$ is surjective. Consider $\Lambda_w^I \cap \widehat{Z}_w = (\Lambda_w^I \cap \widehat{Z}_j) \cap \widehat{Z}_w = \Lambda_w^{I_{>0}} \coprod Z_w$. Then $\Lambda_w^{I_{>0}}$ is open in $\Lambda_w^I \cap \widehat{Z}_w$ and we have a commutative diagram of long exact sequences

$$\begin{tikzcd}
\cdots \arrow[r] & H_{i+1}(\Lambda_w^I \cap \widehat{Z}_w) \arrow[r] & H_{i+1}(\Lambda_w^{I_{>0}}) \arrow[r, "\partial_w"] & H_i(Z_w) \arrow[r] & \cdots \\
\cdots \arrow[r] & H_{i+1}(\Lambda_w^I \cap \widehat{Z}_j) \arrow[r] \arrow[u] & H_{i+1}(\Lambda_w^{I_{>0}}) \arrow[r, "\partial_j"] \arrow[u, equal] & H_i(Z_j) \arrow[r] \arrow[u, "\text{res}_j"'] & \cdots
\end{tikzcd}$$

where ∂_w is the connecting homomorphism of the long exact sequence arising from the partition $\Lambda_w^I \cap \widehat{Z}_w = \Lambda_w^{I_{>0}} \coprod Z_w$. We have seen at the beginning of this section that res_j is surjective and so it is enough to show that ∂_w is surjective. Recall that $\{v_1, v_2\}$ is an $\mathbb{R}$-basis of ℓ and $I = \mathbb{R}_{\geq 0} v_1$. Define

$$\begin{aligned}
E_I &= G \times^{B \cap {}^w B} (\mathbb{R}_{\geq 0} v_1 + \mathfrak{u}_w)\,, \\
E_{I_{>0}} &= G \times^{B \cap {}^w B} (\mathbb{R}_{>0} v_1 + \mathfrak{u}_w)\,, \text{ and} \\
E_0 &= G \times^{B \cap {}^w B} \mathfrak{u}_w.
\end{aligned}$$

It follows from Corollary 3.4 that $E_I \cong \Lambda_w^I$, $E_{I_{>0}} \cong \Lambda_w^{I_{>0}}$, and $E_0 \cong Z_w$, so the long exact sequence arising from the partition $\Lambda_w^I \cap \widehat{Z}_w = \Lambda_w^{I_{>0}} \coprod Z_w$ may be identified with the long exact sequence arising from the partition $E_I = E_{I_{>0}} \coprod E_0$:

$$\cdots \longrightarrow H_{i+1}(E_I) \longrightarrow H_{i+1}(E_{I_{>0}}) \xrightarrow{\partial_E} H_i(E_0) \longrightarrow \cdots$$

Therefore, it is enough to show that ∂_E is surjective. In fact, we show that $H_\bullet(E_I) = 0$ and so ∂_E is an isomorphism.

Define $E_{\mathbb{R}} = G \times^{B \cap {}^w B} (\mathbb{R} v_1 + \mathfrak{u}_w)$. Then $E_{\mathbb{R}}$ is a smooth, real vector bundle over $G/B \cap {}^w B$ and so $E_{\mathbb{R}}$ is a smooth manifold containing E_I as a closed subset. We may apply [3, 2.6.1] and conclude that $H_i(E_I) \cong H^{4n+1-i}(E_{\mathbb{R}}, E_{\mathbb{R}} \setminus E_I)$.

Consider the cohomology long exact sequence of the pair $(E_{\mathbb{R}}, E_{\mathbb{R}} \setminus E_I)$. Since $E_{\mathbb{R}}$ is a vector bundle over $G/B \cap {}^wB$, it is homotopy equivalent to $G/B \cap {}^wB$. Similarly, $E_{\mathbb{R}} \setminus E_I \cong G \times^{B\cap {}^wB} (\mathbb{R}_{<0} v_1 + \mathfrak{u}_w)$ and so is also homotopy equivalent to $G/B \cap {}^wB$. Therefore, $H^i(E_{\mathbb{R}}) \cong H^i(E_{\mathbb{R}} \setminus E_I)$ and it follows that the relative cohomology group $H^i(E_{\mathbb{R}}, E_{\mathbb{R}} \setminus E_I)$ is trivial for every i. Therefore, $H_\bullet(E_I) = 0$, as claimed.
This completes the proof of the lemma. □

Corollary 3.10. *The specialization map* $\lim_1 \colon H_{i+2}(\Lambda_1^{\ell^*}) \longrightarrow H_i(Z_1)$ *is surjective for* $0 \le i \le 4n$.

Proof. This follows from Lemma 3.9, because $Z_1 = Z_{w_1}$ and so res_1 is the identity map. □

The next lemma is true for any specialization map.

Lemma 3.11. *The restriction map* $\mathrm{res}\colon H_{i+2}(\Lambda_w^{\ell^*}) \longrightarrow H_{i+2}(\Lambda_w^P)$ *is surjective for every* w *in* W *and every* $i \ge 0$.

Proof. There are homeomorphisms $\Lambda_w^{\ell^*} \cong G/T \times \ell^*$ and $\Lambda_w^P \cong G/T \times P$. By definition, P is an open subset of ℓ^* and so there is a restriction map $\mathrm{res}\colon H_2(\ell^*) \to H_2(P)$. This map is a non-zero linear transformation between one-dimensional $\mathbb{Q}$-vector spaces so it is an isomorphism.
Using the Künneth formula we get a commutative square where the horizontal maps are isomorphisms and the right-hand vertical map is surjective:

$$\begin{tikzcd} H_{i+2}(\Lambda_w^{\ell^*}) \arrow[r, "\cong"] \arrow[d, "\mathrm{res}"'] & H_i(G/T)\otimes H_2(\ell^*) + H_{i+1}(G/T)\otimes H_1(\ell^*) \arrow[d, "id\otimes \mathrm{res}+0"] \\ H_{i+2}(\Lambda_w^P) \arrow[r, "\cong"'] & H_i(G/T)\otimes H_2(P) \end{tikzcd}$$

It follows that $\mathrm{res}\colon H_{i+2}(\Lambda_w^{\ell^*}) \to H_{i+2}(\Lambda_w^P)$ is surjective. □

Proof of Lemma 3.2. Fix i with $0 \le i \le 4n$. We show that the image of the convolution map $* : H_i(Z_1) \otimes H_{4n}(Z_j) \longrightarrow H_i(Z)$ is precisely $H_i(Z_j)$ for $1 \le j \le N$ using induction on j.
For $j = 1$, $H_{4n}(Z_1)$ is one-dimensional with basis $\{\lambda_1\}$. It follows from Theorem 2.2(c) that λ_1 is the identity in $H_\bullet(Z)$ and so clearly the image of the convolution map $H_i(Z_1) \otimes H_{4n}(Z_1) \longrightarrow H_i(Z)$ is precisely $H_i(Z_1)$.
Assume that $j > 1$ and set $w = w_j$. We will complete the proof using the commutative diagram with exact rows

$$\begin{tikzcd} H_i\otimes H_{4n}(Z_{j-1}) \arrow[r, hook, "id\otimes(r_j)_*"] \arrow[d, "*"'] & H_i\otimes H_{4n}(Z_j) \arrow[r, two heads, "id\otimes \mathrm{res}_j"] \arrow[d, "*"'] & H_i\otimes H_{4n}(Z_w) \arrow[d, "*"'] \\ H_i(Z_{j-1}) \arrow[r, hook, "(r_j)_*"] & H_i(Z_j) \arrow[r, two heads, "\mathrm{res}_j"] & H_i(Z_w) \end{tikzcd} \tag{3.12}$$

and the Five Lemma, where in the first line H_i means $H_i(Z_1)$. We saw in (3.1) that the bottom row is exact and it follows that the top row is also exact. By induction, the convolution product in $H_\bullet(Z)$ determines a surjective map $*: H_i(Z_1) \otimes H_{4n}(Z_{j-1}) \longrightarrow H_i(Z_{j-1})$. To conclude from the Five Lemma that the middle vertical map is a surjection, it remains to define the other vertical maps so that the diagram commutes and to show that the right-hand vertical map is a surjection.

First we show that the image of the map $H_i(Z_1) \otimes H_{4n}(Z_j) \longrightarrow H_i(Z_j)$ determined by the convolution product in $H_\bullet(Z)$ is contained in $H_i(Z_j)$. It then follows that the middle vertical map in (3.12) is defined and so by exactness there is an induced map from $H_i(Z_1) \otimes H_{4n}(Z_w)$ to $H_i(Z_w)$ so that the diagram (3.12) commutes. Second we show that the right-hand vertical map is a surjection.

By Lemma 3.5, $\Lambda_1^{\ell^*}$ is the graph of the identity map of $\widetilde{\mathfrak{g}}^{\ell^*}$, and $\Lambda_w^{\ell^*}$ is the graph of $\widetilde{w}|_{\widetilde{\mathfrak{g}}^{\ell^*}}$. Therefore, $\Lambda_1^{\ell^*} \circ \Lambda_w^{\ell^*} = \Lambda_w^{\ell^*}$ and there is a convolution product

$$H_{i+2}(\Lambda_1^{\ell^*}) \otimes H_{4n+2}(\Lambda_w^{\ell^*}) \xrightarrow{\;*\;} H_{i+2}(\Lambda_w^{\ell^*}).$$

Suppose a is in $H_i(Z_1)$. Then by Corollary 3.10, $a = \lim_1(a_1)$ for some a_1 in $H_{i+2}(\Lambda_1^{\ell^*})$. It is shown in [3, Proposition 2.7.23] that specialization commutes with convolution, so $\lim(a_1 * [\Lambda_w^{\ell^*}]) = \lim(a_1) * \lim([\Lambda_w^{\ell^*}]) = a * \lambda_w$. Also, $a_1 * [\Lambda_w^{\ell^*}]$ is in $H_{i+2}(\Lambda_w^{\ell^*})$ and $\lim = r_* \circ \lim_j$ and so $a * \lambda_w = r_* \circ \lim_j(a_1 * [\Lambda_w^{\ell^*}])$ is in $H_i(Z_j)$. By induction, if $k < j$, then $a * \lambda_{w_k}$ is in $H_i(Z_k)$ and so $a * \lambda_{w_k}$ is in $H_i(Z_k)$. Since the set $\{ \lambda_{w_k} \mid 1 \leq k \leq j \}$ is a basis of $H_{4n}(Z_j)$, it follows that $a * H_{4n}(Z_j) \subseteq H_i(Z_j)$. Therefore, the image of the convolution map $H_i(Z_1) \otimes H_{4n}(Z_j) \longrightarrow H_i(Z)$ is contained in $H_i(Z_j)$.

To complete the proof of Lemma 3.2, we need to show that the induced map from $H_i(Z_1) \otimes H_{4n}(Z_w)$ to $H_i(Z_w)$ is surjective.

Consider the following diagram:

$$\begin{tikzcd}
H_{i+2}(\Lambda_1^{\ell^*}) \otimes H_{4n+2}(\Lambda_w^{\ell^*}) \arrow[r, "*"] \arrow[d, "\lim_1 \otimes \lim_j"'] & H_{i+2}(\Lambda_w^{\ell^*}) \arrow[d, "\lim_j"] \\
H_i(Z_1) \otimes H_{4n}(Z_j) \arrow[r, "*"] \arrow[d, "id \otimes \mathrm{res}_j"'] & H_i(Z_j) \arrow[d, "\mathrm{res}_j"] \\
H_i(Z_1) \otimes H_{4n}(Z_w) \arrow[r, "*"] & H_i(Z_w)
\end{tikzcd}$$

We have seen that the bottom square is commutative. It follows from the fact that specialization commutes with convolution that the top square is also commutative. It is shown in Proposition A.2 that the convolution product $H_{i+2}(\Lambda_1^{\ell^*}) \otimes H_{4n+2}(\Lambda_w^{\ell^*}) \to H_{i+2}(\Lambda_w^{\ell^*})$ is an injection. Since $H_{i+2}(\Lambda_1^{\ell^*})$ is finite-dimensional and $H_{4n+2}(\Lambda_w^{\ell^*})$ is one-dimensional, it follows that this convolution mapping is an isomorphism. Also, we saw in Lemma 3.9 that $\mathrm{res}_j \circ \lim_j$ is surjective. Therefore, the composition $\mathrm{res}_j \circ \lim_j \circ *$ is surjective and it follows that the bottom convolution map $H_i(Z_1) \otimes H_{4n}(Z_w) \to H_i(Z_w)$ is also surjective. This completes the proof of Lemma 3.2.

4. Smash Product Structure

In this section we prove Theorem 2.4. We need to show that $\lambda_w * H_i(Z_1) * \lambda_{w^{-1}} = H_i(Z_1)$ and that $\beta\colon \operatorname{Coinv}_\bullet(W) \xrightarrow{\cong} H_{4n-\bullet}(Z_1)$ is an isomorphism of W-algebras.

Suppose that ℓ is a one-dimensional subspace of $\mathfrak{t}$ so that $\ell^* = \ell \setminus \{0\} = \ell \cap \mathfrak{t}_{\mathrm{reg}}$. Recall that for $S \subseteq \mathfrak{t}$, $\widetilde{\mathfrak{g}}^S = \nu^{-1}(S)$. By Lemma 3.5, if w is in W, then $\Lambda_w^{\ell^*}$ is the graph of the restriction of $\widetilde{w}$ to $\widetilde{\mathfrak{g}}^{\ell^*}$. It follows that there is a convolution product

$$H_{4n+2}(\Lambda_w^{\ell^*}) \otimes H_{i+2}(\Lambda_1^{w^{-1}(\ell^*)}) \otimes H_{4n+2}(\Lambda_w^{w^{-1}(\ell^*)}) \xrightarrow{\;*\;} H_{i+2}(\Lambda_1^{\ell^*}).$$

Because specialization commutes with convolution, the diagram

$$\begin{tikzcd}
H_{4n+2}(\Lambda_w^{\ell^*}) \otimes H_{i+2}(\Lambda_1^{w^{-1}(\ell^*)}) \otimes H_{4n+2}(\Lambda_w^{w^{-1}(\ell^*)}) \arrow[r, "*"] \arrow[d, "\lim\otimes\lim\otimes\lim"'] & H_{i+2}(\Lambda_1^{\ell^*}) \arrow[d, "\lim"] \\
H_{4n}(Z) \otimes H_i(Z_1) \otimes H_{4n}(Z) \arrow[r, "*"'] & H_i(Z)
\end{tikzcd}$$

commutes.

We saw in Corollary 3.10 that $\lim_1\colon H_{i+2}(\Lambda_1^{\ell^*}) \to H_i(Z_1)$ is surjective. Thus, if c is in $H_i(Z_1)$, then $c = \lim(c_1)$ for some c_1 in $H_{i+2}(\Lambda_{w_1}^{w^{-1}(\ell^*)})$. Therefore,

$$\lambda_w * c * \lambda_{w^{-1}} = \lim([\Lambda_w^{\ell^*}]) * \lim(c_1) * \lim([\Lambda_{w^{-1}}^{w^{-1}(\ell^*)}]) = \lim\left([\Lambda_w^{\ell^*}] * c_1 * [\Lambda_{w^{-1}}^{w^{-1}(\ell^*)}]\right).$$

Since $\Lambda_w^{\ell^*}$ and $\Lambda_{w^{-1}}^{w^{-1}(\ell^*)}$ are the graphs of $\widetilde{w}$ and $\widetilde{w}^{-1}$ respectively, and $\Lambda_1^{w^{-1}(\ell^*)}$ is the graph of the identity function, it follows that $[\Lambda_w^{\ell^*}] * c_1 * [\Lambda_{w^{-1}}^{w^{-1}(\ell^*)}]$ is in $H_{i+2}(\Lambda_1^{w^{-1}(\ell^*)})$ and so by (3.8), $\lambda_w * c * \lambda_{w^{-1}}$ is in $H_i(Z_1)$. This shows that $\lambda_w * H_i(Z_1) * \lambda_{w^{-1}} = H_i(Z_1)$ for all i.

To complete the proof of Theorem 2.4 we need to show that if w is in W and f is in $\operatorname{Coinv}_i(W)$, then $\beta(w \cdot f) = \lambda_w * \beta(f) * \lambda_{w^{-1}}$ where $w \cdot f$ denotes the natural action of w on f. To do this, we need some preliminary results.

First, since $\Lambda_1^{\ell^*}$ is the diagonal in $\widetilde{\mathfrak{g}}^{\ell^*} \times \widetilde{\mathfrak{g}}^{\ell^*}$, it is obvious that

$$\delta \circ \widetilde{w}^{-1} = (\widetilde{w}^{-1} \times \widetilde{w}^{-1}) \circ \delta\colon \widetilde{\mathfrak{g}}^{w^{-1}(\ell^*)} \longrightarrow \Lambda_1^{\ell^*}.$$

Therefore,

$$\delta_* \circ \widetilde{w}_*^{-1} = (\widetilde{w}^{-1} \times \widetilde{w}^{-1})_* \circ \delta_*\colon H_i(\widetilde{\mathfrak{g}}^{w^{-1}(\ell^*)}) \longrightarrow H_i(\Lambda_1^{\ell^*}) \tag{4.1}$$

for all i. (The first δ in (4.1) is the diagonal embedding $\widetilde{\mathfrak{g}}^{\ell^*} \cong \Lambda_1^{\ell^*}$ and the second δ is the diagonal embedding $\widetilde{\mathfrak{g}}^{w^{-1}(\ell^*)} \cong \Lambda_1^{w^{-1}(\ell^*)}$.)

Next, with $\ell \subseteq \mathfrak{t}$ as above, $\widetilde{\mathfrak{g}}^{\ell} = \widetilde{\mathfrak{g}}^{\ell^*} \coprod \nu^{-1}(0) = \widetilde{\mathfrak{g}}^{\ell^*} \coprod \widetilde{\mathcal{N}}$ and the restriction of $\nu\colon \widetilde{\mathfrak{g}}^{\ell} \to \ell$ to $\widetilde{\mathfrak{g}}^{\ell^*}$ is a locally trivial fibration. Therefore, there is a specialization map $\lim_0\colon H_{i+2}(\widetilde{\mathfrak{g}}^{\ell^*}) \to H_i(\widetilde{\mathcal{N}})$. Since $\delta_*\colon H_{i+2}(\widetilde{\mathfrak{g}}^{\ell^*}) \to H_{i+2}(\Lambda_1^{\ell^*})$ and $\delta_*\colon H_i(Z) \to H_i(Z_1)$ are isomorphisms, the next lemma is obvious.

Lemma 4.2. *Suppose that ℓ is a one-dimensional subspace of $\mathfrak{t}$ so that $\ell^* = \ell \setminus \{0\} \subseteq \mathfrak{t}_{\mathrm{reg}}$. Then the diagram*

$$\begin{tikzcd} H_{i+2}(\widetilde{\mathfrak{g}}^{\ell^*}) \arrow[r, "\delta_*"] \arrow[d, "\lim_0"'] & H_{i+2}(\Lambda_1^{\ell^*}) \arrow[d, "\lim_1"] \\ H_i(Z) \arrow[r, "\delta_*"'] & H_i(Z_1) \end{tikzcd}$$

commutes.

Finally, $\widetilde{\mathcal{N}} \times_{\mathcal{N}} \mathcal{N} = \widetilde{\mathcal{N}}$ and so $Z \circ \widetilde{\mathcal{N}} = (\widetilde{\mathcal{N}} \times_{\mathcal{N}} \widetilde{\mathcal{N}}) \circ (\widetilde{\mathcal{N}} \times_{\mathcal{N}} \mathcal{N}) = \widetilde{\mathcal{N}} \times_{\mathcal{N}} \mathcal{N}$. Thus, there is a convolution action, $H_{4n}(Z) \otimes H_i(\widetilde{\mathcal{N}}) \longrightarrow H_i(\widetilde{\mathcal{N}})$, of $H_{4n}(Z)$ on $H_i(\widetilde{\mathcal{N}})$.

Suppose that w is in W and z is in $H^i(\mathcal{B})$. Then $\pi^* \circ \mathrm{pd}(z)$ is in $H_{4n-i}(\widetilde{\mathcal{N}})$ and so $\lambda_w * (\pi^* \circ \mathrm{pd}(z))$ is in $H_{4n-i}(\widetilde{\mathcal{N}})$. It is shown in [3, Proposition 7.3.31] that for y in $H_\bullet(\mathcal{B})$, $\lambda_w * \pi^*(y) = \epsilon_w \pi^*(w \cdot y)$ where ϵ_w is the sign of w and $w \cdot y$ denotes the action of W on $H_\bullet(\mathcal{B})$ coming from the action of W on G/T and the homotopy equivalence $G/T \simeq \mathcal{B}$. It is also shown in [3, Proposition 7.3.31] that $\mathrm{pd}(w \cdot z) = \epsilon_w w \cdot \mathrm{pd}(z)$. Therefore,

$$\lambda_w * (\pi^* \circ \mathrm{pd}(z)) = \epsilon_w \pi^*(w \cdot \mathrm{pd}(z)) = \epsilon_w \epsilon_w \pi^* \circ \mathrm{pd}(w \cdot z) = \pi^* \circ \mathrm{pd}(w \cdot z).$$

This proves the next lemma.

Lemma 4.3. *If w is in W and z is in $H_i(\mathcal{B})$, then*

$$\lambda_w * (\pi^* \circ \mathrm{pd}(z)) = \pi^* \circ \mathrm{pd}(w \cdot z).$$

Proof of Theorem 2.4. Fix w in W and f in $\mathrm{Coinv}_i(W)$. We need to show that $\lambda_w * \beta(f) * \lambda_{w^{-1}} = \beta(w \cdot f)$. Set $C = \lambda_w * \beta(f) * \lambda_{w^{-1}}$. Using the fact that $\beta = \delta_* \circ \pi^* \circ \mathrm{pd} \circ \mathrm{bi}$ we compute

$$\begin{aligned}
C &= \lim\nolimits_1 \left([\Lambda_w^{\ell^*}] * \lim\nolimits_1^{-1}(\beta(f)) * [\Lambda_{w^{-1}}^{w^{-1}(\ell^*)}] \right) && [3, 2.7.23] \\
&= \lim\nolimits_1 \circ (\widetilde{w}^{-1} \times \widetilde{w}^{-1})_* \circ \lim\nolimits_1^{-1} \circ \beta(f) && \text{Proposition A.3} \\
&= \lim\nolimits_1 \circ \delta_* \circ \widetilde{w}_*^{-1} \circ \delta_*^{-1} \circ \lim\nolimits_1^{-1} \circ \beta(f) && (4.1) \\
&= \delta_* \circ \lim\nolimits_0 \circ \widetilde{w}_*^{-1} \circ \delta_*^{-1} \circ \lim\nolimits_1^{-1} \circ \delta_* \circ \delta_*^{-1} \circ \beta(f) && \text{Lemma 4.2} \\
&= \delta_* \circ \lim\nolimits_0 \circ \widetilde{w}_*^{-1} \circ \lim\nolimits_0^{-1} \circ \delta_*^{-1} \circ \beta(f) && \text{Lemma 4.2} \\
&= \delta_* \circ \lim\nolimits_0 \circ \widetilde{w}_*^{-1} \circ \lim\nolimits_0^{-1} \circ \pi^* \circ \mathrm{pd} \circ \mathrm{bi}(f) && \\
&= \delta_* \circ \lim\nolimits_0 \left((\lim\nolimits_0^{-1} \circ \pi^* \circ \mathrm{pd} \circ \mathrm{bi}(f)) * [\Lambda_{w^{-1}}^{w^{-1}(\ell^*)}] \right) && [3, 2.7.11] \\
&= \delta_* \left((\pi^* \circ \mathrm{pd} \circ \mathrm{bi}(f)) * \lambda_{w^{-1}} \right) && [3, 2.7.23] \\
&= \delta_* \left(\lambda_w * (\pi^* \circ \mathrm{pd} \circ \mathrm{bi}(f)) \right) && \text{Lemma A.1 and [3, 3.6.11]} \\
&= \delta_* \circ \pi^* \circ \mathrm{pd}(w \cdot \mathrm{bi}(f)) && \text{Lemma 4.3} \\
&= \delta_* \circ \pi^* \circ \mathrm{pd} \circ \mathrm{bi}(w \cdot f) && \mathrm{bi} \text{ is } W\text{-equivariant} \\
&= \beta(w \cdot f).
\end{aligned}$$

This completes the proof of Theorem 2.4.

Appendix A. Convolution and Graphs

In this appendix we prove some general properties of convolution and graphs. Suppose M_1, M_2, and M_3 are smooth varieties, $\dim M_2 = d$, and that $Z_{1,2} \subseteq M_1 \times M_2$ and $Z_{2,3} \subseteq M_2 \times M_3$ are two closed subvarieties so that the convolution product,

$$H_i(Z_{1,2}) \otimes H_j(Z_{2,3}) \xrightarrow{*} H_{i+j-2d}(Z_{1,2} \circ Z_{2,3}),$$

in [3, §2.7.5] is defined. For $1 \leq i, j \leq 3$, let $\tau_{i,j} \colon M_i \times M_j \to M_j \times M_i$ be the map that switches the factors. Define $Z_{2,1} = \tau_{1,2}(Z_{1,2}) \subseteq M_2 \times M_1$ and $Z_{3,2} = \tau_{2,3}(Z_{2,3}) \subseteq M_3 \times M_2$. Then the convolution product

$$H_j(Z_{3,2}) \otimes H_i(Z_{2,1}) \xrightarrow{*'} H_{i+j-2d}(Z_{3,2} \circ Z_{2,1})$$

is defined. We omit the easy proof of the following lemma.

Lemma A.1. *If c is in $H_i(Z_{1,2})$ and d is in $H_j(Z_{2,3})$, then $(\tau_{1,3})_*(c * d) = (\tau_{2,3})_*(d) *' (\tau_{1,2})_*(c)$.*

Now suppose X is an irreducible, smooth, m-dimensional variety, Y is a smooth variety, and $f \colon X \to Y$ is a morphism. Then if Γ_X and Γ_f denote the graphs of id_X and f respectively, using the notation in [3, §2.7], we have $\Gamma_X \circ \Gamma_f = \Gamma_f$ and there is a convolution product $* \colon H_i(\Gamma_X) \otimes H_{2m}(\Gamma_f) \longrightarrow H_i(\Gamma_f)$.

Proposition A.2. *The convolution product $* \colon H_i(\Gamma_X) \otimes H_{2m}(\Gamma_f) \longrightarrow H_i(\Gamma_f)$ is an injection.*

Proof. For $i, j = 1, 2, 3$, let $p_{i,j}$ denote the projection of $X \times X \times Y$ on the i^{th} and j^{th} factors. Then the restriction of $p_{1,3}$ to $(\Gamma_X \times Y) \cap (X \times \Gamma_f)$ is the map that sends $(x, x, f(x))$ to $(x, f(x))$. Thus, the restriction of $p_{1,3}$ to $(\Gamma_X \times Y) \cap (X \times \Gamma_f)$ is an isomorphism onto Γ_f and hence is proper. Therefore, the convolution product in homology is defined.

Since X is irreducible, so is Γ_f and so $H_{2m}(\Gamma_f)$ is one-dimensional with basis $[\Gamma_f]$. Suppose that c is in $H_i(\Gamma_X)$. We need to show that if $c * [\Gamma_f] = 0$, then $c = 0$.

Fix c in $H_i(\Gamma_X)$. Notice that the restriction of $p_{1,3}$ to $(\Gamma_X \times Y) \cap (X \times \Gamma_f)$ is the same as the restriction of $p_{2,3}$ to $(\Gamma_X \times Y) \cap (X \times \Gamma_f)$. Thus, using the projection formula, we have

$$\begin{aligned} c * [\Gamma_f] &= (p_{1,3})_* \left(p_{1,2}^* c \cap p_{2,3}^* [\Gamma_f]\right) \\ &= (p_{2,3})_* \left(p_{1,2}^* c \cap p_{2,3}^* [\Gamma_f]\right) \\ &= \left((p_{2,3})_* p_{1,2}^* c\right) \cap [\Gamma_f], \end{aligned}$$

where the intersection product in the last line is from the cartesian square:

$$\begin{tikzcd} \Gamma_f \arrow[r, "="] \arrow[d] & \Gamma_f \arrow[d] \\ X \times Y \arrow[r, "="'] & X \times Y \end{tikzcd}$$

Let $p\colon X \times Y \to X$ and $q\colon \Gamma_X \to X$ be the first and second projections, respectively. Then the square

$$\begin{tikzcd} \Gamma_X \times Y \arrow[r, "p_{2,3}"] \arrow[d, "p_{1,2}"'] & X \times Y \arrow[d, "p"] \\ \Gamma_X \arrow[r, "q"'] & X \end{tikzcd}$$

is cartesian. Thus,

$$\begin{aligned} p_*\left(c * [\Gamma_f]\right) &= p_*\left(\left((p_{2,3})_* p_{1,2}^* c\right) \cap [\Gamma_f]\right) \\ &= p_*\left((p^* q_* c) \cap [\Gamma_f]\right) \\ &= q_* c \cap (p|_{\Gamma_f})_* [\Gamma_f] \\ &= q_* c \cap [X] \\ &= q_* c, \end{aligned}$$

where we have used the projection formula and the fact that $(p|_{\Gamma_f})_*[\Gamma_f] = [X]$. Now if $c * [\Gamma_f] = 0$, then $q_* c = 0$ and so $c = 0$, because q is an isomorphism. $\square$

Let Γ_Y denote the graph of the identity functions id_Y. Then the following compositions and convolution products in Borel-Moore homology are defined:

- $\Gamma_f \circ \Gamma_X = \Gamma_f$ and so there is a convolution product
$$H_i(\Gamma_f) \otimes H_j(\Gamma_X) \longrightarrow H_{i+j-m}(\Gamma_f).$$
- $\Gamma_Y \circ \Gamma_{f^{-1}} = \Gamma_{f^{-1}}$ and so there is a convolution product
$$H_i(\Gamma_X) \otimes H_j(\Gamma_{f^{-1}}) \longrightarrow H_{i+j-m}(\Gamma_{f^{-1}}).$$
- $\Gamma_f \circ \Gamma_{f^{-1}} = \Gamma_X$ and so there is a convolution product
$$H_i(\Gamma_f) \otimes H_j(\Gamma_{f^{-1}}) \longrightarrow H_{i+j-m}(\Gamma_X).$$

Thus, if c is in $H_i(\Gamma_Y)$, then $[\Gamma_f] * c * [\Gamma_{f^{-1}}]$ is in $H_i(\Gamma_X)$. Note that the map $f^{-1} \times f^{-1}\colon \Gamma_Y \to \Gamma_X$ is an isomorphism, so in particular it is proper.

Proposition A.3. *If c is in $H_i(\Gamma_Y)$, then $[\Gamma_f] * c * [\Gamma_{f^{-1}}] = (f^{-1} \times f^{-1})_*(c)$.*

Proof. We compute $([\Gamma_f] * c) * [\Gamma_{f^{-1}}]$, starting with $[\Gamma_f] * c$.
For $1 \leq i, j \leq 3$ let $q_{i,j}$ be the projection of the subset

$$\Gamma_f \times Y \cap X \times \Gamma_Y = \{\, (x, f(x), f(x)) \mid x \in X \,\}$$

of $X \times Y \times Y$ onto the i,j-factors. Then $q_{1,3} = q_{1,2}$. Therefore, using the projection formula, we see that

$$\begin{aligned} [\Gamma_f] * c &= (q_{1,3})_* \left(q_{1,2}^* [\Gamma_f] \cap q_{2,3}^* c\right) \\ &= (q_{1,2})_* \left(q_{1,2}^* [\Gamma_f] \cap q_{2,3}^* c\right) \\ &= [\Gamma_f] \cap (q_{1,2})_* q_{2,3}^* c \\ &= (q_{1,2})_* q_{2,3}^* c. \end{aligned}$$

Next, for $1 \leq i, j \leq 3$ let $p_{i,j}$ be the projection of the subset

$$\Gamma_f \times X \cap X \times \Gamma_{f^{-1}} = \{\, (x, f(x), x) \mid x \in X \,\}$$

of $X \times Y \times X$ onto the i,j-factors. Then $p_{1,3} = (f^{-1} \times id) \circ p_{2,3}$. Therefore, using the fact that $[\Gamma_f] * c = (q_{1,2})_* q_{2,3}^* c$ and the projection formula, we have

$$\begin{aligned}
([\Gamma_f] * c) * [\Gamma_{f^{-1}}] &= (p_{1,3})_* \left(p_{1,2}^*((q_{1,2})_* q_{2,3}^* c) \cap p_{2,3}^*[\Gamma_{f^{-1}}] \right) \\
&= (f^{-1} \times id)_* (p_{2,3})_* \left(p_{1,2}^*((q_{1,2})_* q_{2,3}^* c) \cap p_{2,3}^*[\Gamma_{f^{-1}}] \right) \\
&= (f^{-1} \times id)_* \left((p_{2,3})_* p_{1,2}^* (q_{1,2})_* q_{2,3}^* c \cap [\Gamma_{f^{-1}}] \right) \\
&= (f^{-1} \times id)_* (p_{2,3})_* p_{1,2}^* (q_{1,2})_* q_{2,3}^* c.
\end{aligned}$$

The commutative square

$$\begin{tikzcd}
\Gamma_f \times X \cap X \times \Gamma_{f^{-1}} \arrow[r, "id \times id \times f"] \arrow[d, "id"'] & \Gamma_f \times Y \cap X \times \Gamma_Y \arrow[d, "q_{1,2}"] \\
\Gamma_f \times X \cap X \times \Gamma_{f^{-1}} \arrow[r, "p_{1,2}"'] & \Gamma_f
\end{tikzcd}$$

is cartesian, so $p_{1,2}^*(q_{1,2})_* = (id \times id \times f)^*$.
Also, the commutative square

$$\begin{tikzcd}
\Gamma_f \times X \cap X \times \Gamma_{f^{-1}} \arrow[r, "q_{2,3} \circ (id \times id \times f)"] \arrow[d, "(f^{-1} \times id) \circ p_{2,3}"'] & \Gamma_Y \arrow[d, "f^{-1} \times f^{-1}"] \\
\Gamma_X \arrow[r, "id"'] & \Gamma_X
\end{tikzcd}$$

is cartesian, so $(f^{-1} \times id)_*(p_{2,3})_*(id \times id \times f)^* q_{2,3}^* = (f^{-1} \times f^{-1})_*$.
Therefore,

$$\begin{aligned}
([\Gamma_f] * c) * [\Gamma_{f^{-1}}] &= (f^{-1} \times id)_* (p_{2,3})_* p_{1,2}^* (q_{1,2})_* q_{2,3}^* c \\
&= (f^{-1} \times id)_* (p_{2,3})_* (id \times id \times f)^* q_{2,3}^* c \\
&= (f^{-1} \times f^{-1})_* c.
\end{aligned}$$

This completes the proof of the proposition. $\square$

References

1. W. Borho, J.L. Brylinski, and R. MacPherson, *Nilpotent Orbits, Primitive Ideals, and Characteristic Classes*, Progress in Mathematics, vol. 78, Birkhäuser, Boston, MA, 1989.
2. W. Borho and R. MacPherson, *Représentations des groupes de Weyl et homologie d'intersection pour les variétés nilpotentes*, C. R. Acad. Sci. Paris 292 (1981), no. 15, 707–710.
3. N. Chriss and V. Ginzburg, *Representation theory and complex geometry*, Birkhäuser, Boston, 1997.

4. W. Fulton and R. MacPherson, *Categorical framework for the study of singular spaces*, Mem. Amer. Math. Soc. 31 (1981), no. 243.
5. H. Hiller, *Geometry of Coxeter groups*, Research Notes in Mathematics, vol. 54, Pitman (Advanced Publishing Program), Boston, Mass., 1982.
6. D. Kazhdan and G. Lusztig, *A topological approach to Springer's representations*, Adv. in Math. 38 (1980), 222–228.
7. ______, *Proof of the Deligne-Langlands conjecture for Hecke algebras*, Invent. Math. 87 (1987), 153–215.
8. N. Kwon, *Borel-Moore homology and K-theory on the Steinberg variety*, to appear in Michigan Math. J.
9. T. Shoji, *Geometry of orbits and Springer correspondence*, Astérisque (1988), no. 168, 9, 61–140, Orbites unipotentes et représentations, I.
10. R. Steinberg, *On the desingularization of the unipotent variety*, Invent. Math. 36 (1976), 209–224.
11. T. Tanisaki, *Twisted differential operators and affine Weyl groups*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. 34 (1987), no. 2, 203–221.
12. R.W. Thomason, *Une formule de Lefschetz en K-théorie équivariante algébrique*, Duke Math. J. 68 (1992), no. 3, 447–462.

J. Matthew Douglass
Department of Mathematics
University of North Texas
Denton TX
USA 76203
douglass@unt.edu

Gerhard Röhrle
Fakultät für Mathematik
Ruhr-Universität Bochum
D-44780 Bochum
Germany
gerhard.roehrle@rub.de

On the Motivic Spectra Representing Algebraic Cobordism and Algebraic K-Theory

David Gepner and Victor Snaith

Received: September 9, 2008

Communicated by Lars Hesselholt

Abstract. We show that the motivic spectrum representing algebraic K-theory is a localization of the suspension spectrum of $\mathbb{P}^\infty$, and similarly that the motivic spectrum representing periodic algebraic cobordism is a localization of the suspension spectrum of BGL. In particular, working over $\mathbb{C}$ and passing to spaces of $\mathbb{C}$-valued points, we obtain new proofs of the topological versions of these theorems, originally due to the second author. We conclude with a couple of applications: first, we give a short proof of the motivic Conner-Floyd theorem, and second, we show that algebraic K-theory and periodic algebraic cobordism are E_∞ motivic spectra.

2000 Mathematics Subject Classification: 55N15; 55N22

1. Introduction

1.1. Background and motivation. Let (X,μ) be an E_∞ monoid in the category of pointed spaces and let $\beta \in \pi_n(\Sigma^\infty X)$ be an element in the stable homotopy of X. Then $\Sigma^\infty X$ is an E_∞ ring spectrum, and we may invert the "multiplication by β" map

$$\mu(\beta) : \Sigma^\infty X \simeq \Sigma^\infty S^0 \wedge \Sigma^\infty X \xrightarrow{\Sigma^{-n}\beta\wedge 1} \Sigma^{-n}\Sigma^\infty X \wedge \Sigma^\infty X \xrightarrow{\Sigma^{-n}\Sigma^\infty \mu} \Sigma^{-n}\Sigma^\infty X.$$

to obtain an E_∞ ring spectrum

$$\Sigma^\infty X[1/\beta] := \operatorname{colim}\{\Sigma^\infty X \xrightarrow{\beta_*} \Sigma^{-n}\Sigma^\infty X \xrightarrow{\Sigma^{-n}\beta_*} \Sigma^{-2n}\Sigma^\infty X \longrightarrow \cdots\}$$

with the property that $\mu(\beta) : \Sigma^\infty X[1/\beta] \to \Sigma^{-n}\Sigma^\infty X[1/\beta]$ is an equivalence. In fact, as is well-known, $\Sigma^\infty X[1/\beta]$ is universal among E_∞ $\Sigma^\infty X$-algebras A in which β becomes a unit.

It was originally shown in [27] (see also [28] for a simpler proof) that the ring spectra $\Sigma^\infty_+ BU[1/\beta]$ and $\Sigma^\infty_+ \mathbb{CP}^\infty[1/\beta]$, obtained as above by taking X to be BU_+ or $\mathbb{P}^\infty_+$ and β a generator of $\pi_2 X$ (a copy of $\mathbb{Z}$ in both cases), represent periodic complex cobordism and topological K-theory, respectively.

This motivated an attempt in [27] to define algebraic cobordism by replacing $BGL(\mathbb{C})$ in this construction with Quillen's algebraic K-theory spaces [26]. The result was an algebraic cobordism theory, defined in the ordinary stable homotopy category, which was far too large.

By analogy with topological complex cobordism, algebraic cobordism ought to be the universal oriented algebraic cohomology theory. However, there are at least two algebraic reformulations of the topological theory; as a result, there are at least two distinct notions of algebraic cobordism popular in the literature today. One, due to Levine and Morel [11], [12], constructs a universal "oriented Borel-Moore" cohomology theory Ω by generators and relations in a way reminiscent of the construction of the Lazard ring, and indeed the value of Ω on the point is the Lazard ring. However, Ω is not a generalized motivic cohomology theory in the sense of Morel and Voevodsky [20], so it is not represented by a motivic ring spectrum.

The other notion, and the one relevant to this paper, is Voevodsky's spectrum MGL [34]. It is a bona fide motivic cohomology theory in the sense that it is defined directly on the level of motivic spectra. Although the coefficient ring of MGL is still not known (at least in all cases), the orientability of MGL implies that it is an algebra over the Lazard ring, as it carries a formal group law. Provided one defines an orientation as a compatible family of Thom classes for vector bundles, it is immediate that MGL represents the universal oriented motivic cohomology theory; moreover, as shown in [23], and just as in the classical case, the splitting principle implies that it is enough to specify a Thom class for the universal line bundle.

The infinite Grassmannian

$$BGL_n \simeq \mathrm{Grass}_{n,\infty} := \mathrm{colim}_k \, \mathrm{Grass}_{n,k}$$

represents, in the $\mathbb{A}^1$-local homotopy category, the functor which associates to a variety X the set of isomorphism classes of rank n vector bundles on X. In particular, tensor product of line bundles and Whitney sum of stable vector bundles endow $\mathbb{P}^\infty \simeq BGL_1$ and $BGL \simeq \mathrm{colim}_n BGL_n$ with the structure of abelian group objects in the $\mathbb{A}^1$-homotopy category. Note that, over $\mathbb{C}$, the spaces $\mathbb{P}^\infty(\mathbb{C})$ and $BGL(\mathbb{C})$ underlying the associated complex-analytic varieties are equivalent to the usual classifying spaces $\mathbb{CP}^\infty$ and BU.

We might therefore hypothesize, by analogy with topology, that there are equivalences of motivic ring spectra

$$\Sigma^\infty_+ BGL[1/\beta] \longrightarrow PMGL \qquad \text{and} \qquad \Sigma^\infty_+ \mathbb{P}^\infty[1/\beta] \longrightarrow K$$

where $PMGL$ denotes a periodic version of the algebraic cobordism spectrum MGL. The purpose of this paper is to prove this hypothesis. In fact, it holds over an arbitrary Noetherian base scheme S of finite Krull dimension, provided one interprets K properly: the Thomason-Trobaugh K-theory of schemes [33] is not homotopy invariant, and so it cannot possibly define a motivic cohomology theory. Rather, the motivic analogue of K-theory is Weibel's homotopy K-theory [38]; the two agree for any *regular* scheme.

1.2. Organization of the paper. We begin with an overview of the theory of oriented motivic ring spectra. The notion of an orientation is a powerful one, allowing us to compute first the oriented cohomology of flag varieties and Grassmannians. We use our calculations to identify the primitive elements in the Hopf algebra $R^0(\mathbb{Z} \times BGL)$ with $R^0(BGL_1)$, a key point in our analysis of the abelian group $R^0(K)$ of spectrum maps from K to R.
The second section is devoted to algebraic cobordism, in particular the proof that algebraic cobordism is represented by the motivic spectrum $\Sigma^\infty_+ BGL[1/\beta]$. We recall the construction of MGL as well as its periodic version $PMGL$ and note the functors they (co)represent as monoids in the homotopy category of motivic spectra. We show that $PMGL$ is equivalent to $\bigvee_n \Sigma^\infty MGL_n[1/\beta]$ and use the isomorphism $R^0(BGL) \cong \prod_n R^0(MGL_n)$ to identify the functors $\mathrm{Rings}(\Sigma^\infty_+ BGL[1/\beta], -)$ and $\mathrm{Rings}(\bigvee_n \Sigma^\infty MGL_n[1/\beta], -)$.
The third section provides the proof that algebraic K-theory is represented by the motivic spectrum $\Sigma^\infty_+ \mathbb{P}^\infty[1/\beta]$. First we construct a map; to see that it's an equivalence, we note that it's enough to show that the induced map $R^0(K) \to R^0(\Sigma^\infty_+ \mathbb{P}^\infty[1/\beta])$ is an isomorphism for any $PMGL$-algebra R. An element of $R^0(K)$ amounts to a homotopy class of an infinite loop map $\mathbb{Z} \times BGL \simeq \Omega^\infty K \to \Omega^\infty R$; since loop maps $\mathbb{Z} \times BGL \to \Omega^\infty R$ are necessarily additive, we are reduced to looking at maps $\mathbb{P}^\infty \to \Omega^\infty R$. We use this to show that the spaces $\mathrm{map}(K, R)$ and $\mathrm{map}(\Sigma^\infty_+ \mathbb{P}^\infty[1/\beta], R)$ both arise as the homotopy inverse limit of the tower associated to the endomorphism of the space $\mathrm{map}(\Sigma^\infty_+ \mathbb{P}^\infty, R)$ induced by the action of the Bott map $\mathbb{P}^1 \wedge \mathbb{P}^\infty \to \mathbb{P}^\infty$, and are therefore homotopy equivalent.
We conclude the paper with a couple of corollaries. The first is a quick proof of the motivic Conner-Floyd theorem, namely that the map

$$MGL^{*,*}(X) \otimes_{MGL^{*,*}} K^{*,*} \longrightarrow K^{*,*}(X),$$

induced by an MGL-algebra structure on K, is an isomorphism for any compact motivic spectrum X. This was first obtained by Panin-Pimenov-Röndigs [24] and follows from a motivic version of the Landweber exact functor theorem [21]. We include a proof because, using the aforementioned structure theorems, we obtain a simplification of the (somewhat similar) method in [24], but which is considerably more elementary than that of [21].
Second, it follows immediately from our theorems that both K and $PMGL$ are E_∞ *as motivic spectra*. An E_∞ motivic spectrum is a *coherently* commutative object in an appropriate symmetric monoidal model category of structured motivic spectra, such as P. Hu's motivic $\mathbb{S}$-modules [6] or J.F. Jardine's motivic symmetric spectra [7]; in particular, this is a much stronger than the assertion that algebraic K-theory defines a presheaf of (ordinary) E_∞ spectra on an appropriate site. This is already known to be the case for algebraic cobordism, where it is clear from the construction of MGL, but does not appear to be known either for *periodic* algebraic cobordism or algebraic K-theory.

This is important because the category of modules over an E_∞ motivic spectrum R inherits a symmetric monoidal structure, at least in the higher categorical sense of [16]. As a result, there is a version of derived algebraic geometry which uses E_∞ motivic spectra as its basic building blocks. In [13], J. Lurie shows that $\operatorname{spec}\Sigma^\infty_+\mathbb{P}^\infty[1/\beta]$ is the initial derived scheme over which the derived multiplicative group $\mathbb{G}_R := \operatorname{spec} R \wedge \Sigma^\infty_+\mathbb{Z}$ acquires an "orientation", in the sense that the formal group of $\mathbb{G}_R$ may be identified with the formal spectrum $\mathbb{P}^\infty \otimes \operatorname{spec} R$. Since $\Sigma^\infty_+\mathbb{CP}^\infty[1/\beta]$ represents topological K-theory, this is really a theorem about the relation between K-theory and the derived multiplicative group, and is the starting point for Lurie's program to similarly relate topological modular forms and derived elliptic curves. Hence the motivic version of the K-theory result may be seen as a small step towards an algebraic version of elliptic cohomology.

1.3. Acknowledgements. We are very grateful to Mike Hopkins and Rick Jardine for helpful lunch-break discussions during a workshop at the Fields Institute in May 2007, as well as to an anonymous referee for their careful reading of the paper and constructive criticism. We thank M. Spitzweck and P.A. Østvær for bringing to our attention their methods, which give an alternate proof of Theorem 4.17 (see [32]), and for suggesting a number of valuable comments. The first author would also like to thank Sarah Whitehouse for illuminating conversations about operations in K-theory and John Greenlees for his interest in this project and its equivariant analogues.

2. Oriented Cohomology Theories

2.1. Motivic spaces. Throughout this paper, we write S for a Noetherian base scheme of finite Krull dimension.

Definition 2.1. A *motivic space* is a simplicial sheaf on the Nisnevich site of smooth schemes over S.

We often write 0 for the initial motivic space $\emptyset$, the simplicial sheaf with constant value the set with zero elements, and 1 for the final motivic space S, the simplicial sheaf with constant value the set with one element.

We assume that the reader is familiar with the Morel-Voevodsky $\mathbb{A}^1$-local model structure on the category of motivic spaces used to define the unstable motivic homotopy category [20]. We adhere to this treatment with one exception: we adopt a different convention for indexing the simplicial and algebraic spheres. The *simplicial circle* is the pair associated to the constant simplicial sheaves

$$S^{1,0} := (\Delta^1, \partial\Delta^1);$$

its smash powers are the *simplicial spheres*

$$S^{n,0} := (\Delta^n, \partial\Delta^n).$$

The *algebraic circle* is the multiplicative group scheme $\mathbb{G} := \mathbb{G}_m := \mathbb{A}^1 - \mathbb{A}^0$, pointed by the identity section $1 \to \mathbb{G}$; its smash powers define the *algebraic* *spheres*

$$S^{0,n} := (\mathbb{G}, 1)^{\wedge n}.$$

Putting the two together, we obtain a bi-indexed family of spheres

$$S^{p,q} := S^{p,0} \wedge S^{0,q}.$$

It is straightforward to show that

$$(\mathbb{A}^n - \mathbb{A}^0, 1) \simeq S^{n-1,n}$$

and

$$(\mathbb{A}^n, \mathbb{A}^n - \mathbb{A}^0) \simeq (\mathbb{P}^n, \mathbb{P}^{n-1}) \simeq S^{n,n}.$$

We emphasize that, according to the more usual grading convention, $S^{p,q}$ is written $S^{p+q,q}$; we find it more intuitive to separate the simplicial and algebraic spheres notationally. Moreover, for this purposes of this paper, the diagonal spheres

$$S^{n,n} \simeq (\mathbb{A}^n, \mathbb{A}^n - \mathbb{A}^0) \simeq (\mathbb{P}^n, \mathbb{P}^{n-1})$$

are far and away the most important, so they will be abbreviated

$$S^n := S^{n,n}.$$

This allows us to get by with just a single index most of the time.
We extend this convention to suspension and loop functors. That is, $\Sigma(-)$ denotes the endofunctor on pointed motivic spaces (or spectra) defined by

$$\Sigma X := S^1 \wedge X := S^{1,1} \wedge X.$$

Similarly, its right adjoint $\Omega(-)$ is defined by

$$\Omega X := \mathrm{map}_+(S^1, X) := \mathrm{map}_+(S^{1,1}, X).$$

Note that Σ is therefore *not* the categorical suspension, which is to say that the cofiber of the unique map $X \to 1$ is given by $S^{1,0} \wedge X$ instead of $S^{1,1} \wedge X = \Sigma X$. While this may be confusing at first, we feel that the notational simplification that results makes it worthwhile in the end.

2.2. Motivic spectra. To form the stable motivic category, we formally add desuspensions with respect to the diagonal spheres $S^n = S^{n,n} = (\mathbb{A}^n, \mathbb{A}^n - \mathbb{A}^0)$.

Definition 2.2. A *motivic prespectrum* is a sequence of pointed motivic spaces

$$\{X(0), X(1), \ldots\},$$

equipped with maps $\Sigma^p X(q) \to X(p+q)$, such that the resulting squares

$$\begin{tikzcd} \Sigma^p \Sigma^q X(r) \arrow[r] \arrow[d] & \Sigma^{p+q} X(r) \arrow[d] \\ \Sigma^p X(q+r) \arrow[r] & X(p+q+r) \end{tikzcd}$$

commute.

Definition 2.3. A motivic prespectrum is a *motivic spectrum* if, for all natural numbers p, q, the adjoints $X(q) \to \Omega^p X(p+q)$ of the prespectrum structure maps $\Sigma^p X(q) \to X(p+q)$ are weak equivalences.

A pointed motivic space $X = (X, 1)$ gives rise to the suspension spectrum $\Sigma^\infty X$, the spectrum associated to the prespectrum with

$$(\Sigma^\infty X)(p) := \Sigma^p X$$

and structure maps

$$\Sigma^q \Sigma^p X \longrightarrow \Sigma^{p+q} X.$$

If X isn't already pointed, we usually write $\Sigma_+^\infty X$ for $\Sigma^\infty X_+$, where X_+ is the pointed space $(X_+, 1) \simeq (X, 0)$. If X happens to be the terminal object 1, we write $\mathbb{S} := \Sigma_+^\infty 1$ for the resulting suspension spectrum, the motivic sphere.

We will need that the category of motivic spectra is closed symmetric monoidal with respect to the smash product. However, we do not focus on the details of its construction, save to say that either P. Hu's theory of motivic $\mathbb{S}$-modules [6] or J.F. Jardine's motivic symmetric spectra [7] will do.

In particular, the category of motivic spectra is tensored and cotensored over itself via the smash product and the motivic function spectrum bifunctors. We may also regard it as being tensored and cotensored over pointed motivic spaces via the suspension spectrum functor. Given a motivic spectrum R and a pointed motivic space X, we write $X \wedge R$ for the motivic spectrum $\Sigma^\infty X \wedge R$ and R^X for the motivic spectrum of maps from $\Sigma^\infty X$ to R. Here $\Sigma^\infty X$ is the motivic spectrum associated to the motivic prespectrum whose value in degree n is the pointed motivic space $\Sigma^n X$. As a functor from pointed motivic spaces to motivic spectra, Σ^∞ admits a right adjoint Ω^∞ which associates to a motivic spectrum its underlying motivic "infinite-loop".

There are also a number of symmetric monoidal categories over which the category of motivic spectra is naturally enriched. We write Y^X for the motivic function spectrum of maps from the motivic spectrum X to the motivic spectrum Y, $\mathrm{map}(X, Y) = (\Omega^\infty Y^X)(S)$ for the (ordinary) space of maps from X to Y, and $[X, Y] = Y^0(X) = \pi_0 \,\mathrm{map}(X, Y)$ for the abelian group of homotopy classes of maps from X to Y.

2.3. Motivic ring spectra. In this paper, unless appropriately qualified, a motivic ring spectrum will always mean a (not necessarily commutative) monoid in the homotopy category of motivic spectra. We reiterate that a motivic spectrum is a $\mathbb{P}^1$-spectrum; that is, it admits desuspensions by algebraic spheres as well as simplicial spheres.

Definition 2.4. A motivic ring spectrum R is *periodic* if the graded ring $\pi_* R$ contains a unit $\mu \in \pi_1 R$ in degree one.

Remark 2.5. Since $\pi_1 R$ is by definition $\pi_0 \,\mathrm{map}_+(\mathbb{P}^1, \Omega^\infty R)$, and over $\mathrm{spec}\,\mathbb{C}$, $\mathbb{P}^1(\mathbb{C}) \simeq \mathbb{CP}^1$, the topological 2-sphere, this is compatible with the notion of an *even periodic* ring spectrum so common in ordinary stable homotopy theory.

Proposition 2.6. *If R is periodic then $R \simeq \Sigma^n R$ for all n.*

Proof. Let $\mu \in \pi_1 R$ be a unit with inverse $\mu^{-1} \in \pi_{-1} R$. Then for any n, the multiplication by μ^{-n} map

$$R \longrightarrow \Sigma^n R$$

is an equivalence, since multiplication by μ^n provides an inverse. □

Let $P\mathbb{S}$ denote the periodic sphere, the motivic spectrum

$$P\mathbb{S} := \bigvee_{n\in\mathbb{Z}} \Sigma^n \mathbb{S}.$$

With respect to the multiplication induced by the equivalences $\Sigma^p\mathbb{S} \wedge \Sigma^q\mathbb{S} \to \Sigma^{p+q}\mathbb{S}$, the unit in degree one given by the inclusion $\Sigma^1\mathbb{S} \to P\mathbb{S}$ makes $P\mathbb{S}$ into a periodic $\mathbb{S}$-algebra.

More generally, given an arbitrary motivic ring spectrum R,

$$PR := P\mathbb{S} \wedge R \simeq \bigvee_{n\in\mathbb{Z}} \Sigma^n R$$

is a periodic ring spectrum equipped with a ring map $R \to PR$.

Proposition 2.7. *Let R be a motivic ring spectrum. Then homotopy classes of ring maps $P\mathbb{S} \to R$ naturally biject with units in $\pi_1 R$.*

Proof. By definition, ring maps $P\mathbb{S} \to R$ are indexed by families of elements $r_n \in \pi_n R$ with $r_m r_n = r_{m+n}$ and $r_0 = 1$. Hence $r_n = r_1^n$, and in particular $r_{-1} = r_1^{-1}$. □

Said differently, the homotopy category of periodic motivic ring spectra is equivalent to the full subcategory of the homotopy category of motivic ring spectra which admit a ring map from $P\mathbb{S}$. This is not the same as the homotopy category of $P\mathbb{S}$-algebras, in which only those maps which preserve the distinguished unit are allowed.

Corollary 2.8. *Let Q be a motivic ring spectrum and R a periodic motivic ring spectrum. Then the set of homotopy classes of ring maps $PQ \to R$ is naturally isomorphic to the set of pairs consisting of a homotopy class of ring map $Q \to R$ and a distinguished unit $\mu \in \pi_1 R$.*

2.4. Orientations. Let R be a commutative motivic ring spectrum.

Definition 2.9. The *Thom space* of an n-plane bundle $V \to X$ is the pair $(V, V - X)$, where $V - X$ denotes the complement in V of the zero section $X \to V$.

Given two vector bundles $V \to X$ and $W \to Y$, the Thom space $(V \times W, V \times W - X \times Y)$ of the product bundle $V \times W \to X \times Y$ is equivalent (even isomorphic) to the smash product $(V, V - X) \wedge (W, W - Y)$ of the Thom spaces. Since the Thom space of the trivial 1-dimensional bundle $\mathbb{A}^1 \to \mathbb{A}^0$ is the motivic 1-sphere $S^1 \simeq (\mathbb{A}^1, \mathbb{A}^1 - \mathbb{A}^0)$, we see that the Thom space of the trivial n-dimensional bundle $\mathbb{A}^n \to \mathbb{A}^0$ is the motivic n-sphere $S^n \simeq (\mathbb{A}^n, \mathbb{A}^n - \mathbb{A}^0)$. Note that the complement of the zero section $\mathbb{L} - \mathbb{P}^\infty$ of the universal line bundle

$\mathbb{L} \to \mathbb{P}^\infty \simeq B\mathbb{G}$ is equivalent to the total space of the universal principal $\mathbb{G}$-bundle $E\mathbb{G} \to B\mathbb{G}$, which is contractible. Hence the Thom space of $\mathbb{L} \to \mathbb{P}^\infty$ is equivalent to $(\mathbb{P}^\infty, \mathbb{P}^0)$, and the Thom space of the restriction of $\mathbb{L} \to \mathbb{P}^\infty$ along the inclusion $\mathbb{P}^1 \to \mathbb{P}^\infty$ is equivalent to $(\mathbb{P}^1, \mathbb{P}^0) \simeq S^1$.

DEFINITION 2.10. An *orientation* of R is the assignment, to each m-plane bundle $V \to X$, of a class $\theta(V/X) \in R^m(V, V - X)$, in such a way that

(1) for any $f : Y \to X$, the class $\theta(f^*V/Y)$ of the restriction $f^*V \to Y$ of $V \to X$ is equal to the restriction $f^*\theta(V/X)$ of the class $\theta(V/X)$ in $R^m(f^*V, f^*V - Y)$,
(2) for any n-plane bundle $W \to Y$, the (external) product $\theta(V/X) \times \theta(W/Y)$ of the classes $\theta(V/X)$ and $\theta(W/Y)$ is equal to the class $\theta(V \times W/X \times Y)$ of the (external) product of $V \to X$ and $W \to Y$ in $R^{m+n}(V \times W, V \times W - X \times Y)$, and
(3) if $\mathbb{L} \to \mathbb{P}^\infty$ is the universal line bundle and $i : \mathbb{P}^1 \to \mathbb{P}^\infty$ denotes the inclusion, then $i^*\theta(\mathbb{L}/\mathbb{P}^\infty) \in R^1(f^*\mathbb{L}, f^*\mathbb{L} - \mathbb{P}^1)$ corresponds to $1 \in R^0(S^0)$ via the isomorphism $R^0(S^0) \cong R^1(S^1) \cong R^1(f^*\mathbb{L}, f^*\mathbb{L} - \mathbb{P}^1)$.

Given an orientation of R, the class $\theta(V/X) \in R^n(V, V - X)$ associated to a n-plane bundle $V \to X$ is called the *Thom class* of $V \to X$. The main utility of Thom classes is that they define $R^*(X)$-module isomorphisms $R^*(X) \to R^{*+n}(V, V - X)$ (cf. [23]).

REMARK 2.11. The naturality condition implies that it is enough to specify Thom classes for the universal vector bundles $V_n \to BGL_n$. We write MGL_n for the Thom space of $V_n \to BGL_n$ and θ_n for $\theta(V_n/BGL_n) \in R^n(MGL_n)$.

2.5. BASIC CALCULATIONS IN ORIENTED COHOMOLOGY. In this section we fix an oriented commutative motivic ring spectrum R equipped with a unit $\mu \in \pi_1 R$. Note that we can use μ to move the Thom classes $\theta_n \in R^n(MGL_n)$ to degree zero Thom classes $\vartheta_n := \mu^n\theta_n \in R^0(MGL_n)$. The following calculations are well known (cf. [1], [4], [23]). Note that all (co)homology is implicitly the (co)homology of a pair. In particular, if X is unpointed, then $R^0(X) := R^0(X, 0)$, where $0 \to X$ is the unique map from the inital object 0; if X is pointed, then $R^0(X) := R^0(X, 1)$, where $1 \to X$ is the designated map from the terminal object 1.

PROPOSITION 2.12. *The first Chern class of the tautological line bundle on $\mathbb{P}^n$ defines a ring isomorphism $R^0[\lambda]/(\lambda^{n+1}) \to R^0(\mathbb{P}^n)$.*

Proof. Inductively, one has a morphism of exact sequences

$$\begin{tikzcd}
\lambda^n R^0[\lambda]/(\lambda^{n+1}) \arrow[r] \arrow[d] & R^0[\lambda]/(\lambda^{n+1}) \arrow[r] \arrow[d] & R^0[\lambda]/(\lambda^n) \arrow[d] \\
R^0(\mathbb{P}^n, \mathbb{P}^{n-1}) \arrow[r] & R^0(\mathbb{P}^n) \arrow[r] & R^0(\mathbb{P}^{n-1})
\end{tikzcd}$$

in which the left and right, and hence also the middle, vertical maps are isomorphisms. □

Proposition 2.13. *The first Chern class of the tautological line bundle on* $\mathbb{P}^\infty$ *defines a ring isomorphism* $R^0[\![\lambda]\!] \cong \lim R^0[\lambda]/(\lambda^n) \to R^0(\mathbb{P}^\infty_+)$.

Proof. The $\lim^1$ term in the exact sequence

$$0 \longrightarrow \lim{}^1 R^{-1,0}(\mathbb{P}^n) \longrightarrow R^0(\mathbb{P}^\infty) \longrightarrow \lim R^0(\mathbb{P}^n)$$

vanishes because the maps $R^{-1,0}(\mathbb{P}^n) \longrightarrow R^{-1,0}(\mathbb{P}^{n-1})$ are surjective. □

Corollary 2.14. *For each n, the natural map*

$$R_0(\mathbb{P}^n) \longrightarrow \hom_{R^0}(R^0(\mathbb{P}^n), R^0)$$

is an isomorphism.

Proof. The dual of 2.12 shows that $R_0(\mathbb{P}^n)$ is free of rank $n+1$ over R_0. □

Proposition 2.15 (Atiyah [4]). *Let $p : Y \to X$ be a map of quasicompact S-schemes and let $y_1, \ldots, y_n$ be elements of $R^0(Y)$. Let M be the free abelian group on the $y_1, \ldots, y_n$, and suppose that X has a cover by open subschemes U such that for all open V in U, the natural map*

$$R^0(V) \otimes M \longrightarrow R^0(p^{-1}V)$$

is an isomorphism. Then, for any open W in X, the map

$$R^0(X, W) \otimes M \longrightarrow R^0(Y, p^{-1}W)$$

is an isomorphism.

Proof. Apply Atiyah's proof [4], *mutatis mutandis.* □

Proposition 2.16. *Let Z be an S-scheme such that, for any homotopy commutative R-algebra A, $A^0(Z) \cong R^0(Z) \otimes_{R^0} A^0$. Then, for any S-scheme X, $R^0(Z \times X) \cong R^0(Z) \otimes_{R^0} R^0(X)$.*

Proof. The diagonal of X induces a homotopy commutative R-algebra structure on $A = R^X$, the cotensor of the motivic space X with the motivic spectrum R. Hence

$$A^0(Z) \cong R^0(Z) \otimes_{R^0} A^0 \cong R^0(Z) \otimes_{R^0} R^0(X).$$

□

Corollary 2.17. *Let $p : V \to X$ be a rank n vector bundle over a quasicompact S-scheme X and let $L \to \mathbb{P}(V)$ be the tautological line bundle. Then the map which sends λ to the first Chern class of L induces an isomorphism*

$$R^0(X)[\lambda]/(\lambda^n - \lambda^{n-1}c_1 V + \cdots + (-1)^n c_n V) \longrightarrow R^0(\mathbb{P}(V))$$

of R^0-algebras.

Proof. If V is trivial then $\mathbb{P}(V) \cong \mathbb{P}^{n-1}_X$, and the result follows from Propositions 2.12 and 2.16. In general, the projection $\mathbb{P}(V) \to X$ is still locally trivial, so we may apply Proposition 2.15, with W empty and $\{y_i\}$ the image in $R^0(\mathbb{P}(V))$ of a basis for $R^0(X)[\lambda]/(\lambda^n - \ldots + (-1)^n c_n V)$ as a free $R^0(X)$-module. □

Proposition 2.18. *Let $V \to X$ be a rank n vector bundle over a quasicompact S-scheme X, $\mathrm{Flag}(V) \to X$ the associated flag bundle, and $\sigma_k(x_1, \ldots, x_n)$, $1 \le k \le n$, the k^{th} elementary symmetric function in the indeterminates λ_i. Then the map*

$$R^0(X)[\lambda_1, \ldots, \lambda_n]/(\{c_k(V) - \sigma_k(\lambda_1, \ldots, \lambda_n)\}_{k>0}) \longrightarrow R^0(\mathrm{Flag}(V))$$

which sends the λ_i to the first Chern classes of the n tautological line bundles on $\mathrm{Flag}(V)$*, is an isomorphism of R^0-algebras.*

Proof. The evident relations among the Chern classes imply that the map is well-defined. Using Proposition 2.15 and a basis for the free R^0-module $R^0(\mathrm{Flag}(\mathbb{A}^{n-1})$, it follows inductively from the fibration $\mathrm{Flag}(\mathbb{A}^{n-1}) \to \mathrm{Flag}(\mathbb{A}^n) \to \mathbb{P}^{n-1}$ that

$$R^0[\lambda_1, \ldots, \lambda_n]/(\{\sigma_k(\lambda_1, \ldots, \lambda_n)\}) \longrightarrow R^0(\mathrm{Flag}(\mathbb{A}^n))$$

is an isomorphism. Using Proposition 2.16, we deduce the desired result for trivial vector bundles V. For the general case, we apply Proposition 2.15 again, with a basis of the free $R^0(X)$-module $R^0(X)[\lambda_1, \ldots, \lambda_n]/(\{c_k(V) - \sigma_k\})$ giving the necessary elements of $R^0(\mathrm{Flag}(V))$. □

Proposition 2.19. *Let $p : V \to X$ be an rank n vector bundle over a quasicompact S-scheme X, let $q : \mathrm{Grass}_m(V) \to X$ be the Grassmannian bundle of m-dimensional subspaces of V, let $\xi_m(V) \to \mathrm{Grass}_m(V)$ be the tautological m-plane bundle over* $\mathrm{Grass}_m(V)$*, and write $q^*(V)/\xi_m(V)$ for the quotient $(n-m)$-plane bundle. Then the map*

$$R^0(X)[\sigma_1, \ldots, \sigma_m, \tau_1, \ldots, \tau_{n-m}] \longrightarrow R^0(\mathrm{Grass}_m(V))$$

which sends σ_i to $c_i(\xi_m(V))$ and τ_j to $c_j(q^*(V)/\xi_m(V))$ induces an isomorphism

$$R^0(X)[\sigma_1, \ldots, \sigma_m, \tau_1, \ldots, \tau_{n-m}]/(\{c_k(V) - \sum_{i+j=k} \sigma_i \tau_j\}) \longrightarrow R^0(\mathrm{Grass}_m(V))$$

of R^0-algebras (as usual, $c_k(V) = 0$ for $k > n$ and $c_0(V) = \sigma_0 = \tau_0 = 1$).

Proof. The identity $q^*c(V) = c(q^*V) = c(\xi_m(V))c(q^*V/\xi_m(V))$ implies that each $c_k(V) - \Sigma\sigma_i\tau_j$ is sent to zero, so the map is well-defined. Just as in the case of flag bundles, use induction together with the fibration $\mathrm{Flag}(\mathbb{A}^m) \times \mathrm{Flag}(\mathbb{A}^{n-m}) \to \mathrm{Flag}(\mathbb{A}^n) \to \mathrm{Grass}_m(\mathbb{A}^n)$ to see that

$$R^0[\sigma_1, \ldots, \sigma_m, \tau_1, \ldots, \tau_{n-m}]/(\{\sum_{i+j=k} \sigma_i \tau_j\}) \longrightarrow R^0(\mathrm{Grass}_m(\mathbb{A}^n))$$

is an isomorphism. This implies the result for trivial vector bundles by Proposition 2.16, and we deduce the general result from Proposition 2.15. □

Proposition 2.20. *There are isomorphisms*

$$R^0(BGL_n) \longrightarrow R^0(BGL_1^n)^{\Sigma_n} \cong R^0[\![\lambda_1, \ldots, \lambda_n]\!]^{\Sigma_n} \cong R^0[\![\sigma_1, \ldots, \sigma_n]\!].$$

Proof. Writing $V_n \to BGL_n$ for the tautological vector bundle, we have an equivalence $\mathrm{Flag}(V_n) \simeq BGL_1^n$. Inductively, we have isomorphisms

$$R^0[\![\lambda_1, \ldots, \lambda_n]\!] \longrightarrow R^0(BGL_1^n)$$

and the map

$$R^0(BGL_n) \longrightarrow R^0(BGL_1^n) \cong R^0[\![\lambda_1, \ldots, \lambda_n]\!]$$

factors through the invariant subring $R^0[\![\lambda_1, \ldots, \lambda_n]\!]^{\Sigma_n} \cong R^0[\![\sigma_1, \ldots, \sigma_n]\!]$. By Proposition 2.18, $R^0(BGL_1^n)$ is free of rank $n!$ over $R^0(BGL_n)$, so it follows that $R^0(BGL_n) \cong R^0(BGL_1^n)^{\Sigma_n}$. □

Corollary 2.21. *The natural map*

$$R^0(BGL_n) \longrightarrow \hom_{R_0}(\mathrm{Sym}^n_{R_0} R_0(\mathbb{P}^\infty), R_0)$$

is an isomorphism.

Proof. By Proposition 2.20, we need only check this for $n = 1$. But

$$R^0(\mathbb{P}^m) \cong \hom_{R_0}(R_0(\mathbb{P}^m), R_0),$$

both being free of rank $m+1$ over R^0, and

$$R^0(\mathbb{P}^\infty) \cong \lim R^0(\mathbb{P}^m) \cong \hom_{R_0}(\mathrm{colim}\, R_0(\mathbb{P}^m), R_0) \cong \hom_{R_0}(R_0(\mathbb{P}^\infty), R_0)$$

by Proposition 2.12. □

Corollary 2.22. *There are isomorphisms* $R^0(BGL) \cong \lim_n R^0(BGL_n) \cong R^0[\![\sigma_1, \sigma_2, \ldots]\!]$.

Proof. The $\lim^1$ term in the short exact sequence

$$0 \longrightarrow \lim\nolimits^1_n R^{1,0}(BGL_n) \longrightarrow R^0(BGL) \longrightarrow \lim\nolimits_n R^0(BGL_n)$$

vanishes since the maps $R^{1,0}(BGL_n) \to R^{1,0}(BGL_{n-1})$ are surjective. □

2.6. The oriented cohomology of $BGL_+ \wedge Z$. Let R be an oriented periodic commutative motivic ring spectrum and let Z be an arbitrary motivic spectrum. Recall (cf. [5]) that a motivic spectrum is *cellular* if belongs to the smallest full subcategory of motivic spectra which is closed under homotopy colimits and contains the spheres $S^{p,q}$ for all $p, q \in \mathbb{Z}$, and that a motivic space is *stably cellular* if its suspension spectrum is cellular.

Proposition 2.23. *Let $X := \operatorname{colim}_n X_n$ be a telescope of finite cellular motivic spectra such that each $R^{*,*}(X_n)$ is a finite free $R^{*,*}$-module and, for any motivic spectrum Z, the induced maps*

$$R^0(X_n \wedge Z) \longrightarrow R^0(X_{n-1} \wedge Z)$$

are surjective. Then the natural map

$$R^0(X) \underset{R^0}{\widehat{\otimes}} R^0(Z) \longrightarrow R^0(X \wedge Z)$$

is an isomorphism.

Proof. This is an immediate consequence of the motivic Künneth spectral sequence of Dugger-Isaksen [5]. Indeed, for each n, $R^{*,*}(X_n)$ is a free $R^{*,*}$-module, so the spectral sequence

$$\operatorname{Tor}_*^{R^{*,*}}(R^{*,*}(X_n), R^{*,*}(Z)) \Rightarrow R^{*,*}(X_n \wedge Z)$$

collapses to yield the isomorphism

$$R^{*,*}(X_n) \underset{R^{*,*}}{\widehat{\otimes}} R^{*,*}(Z) \cong R^{*,*}(X_n \wedge Z).$$

Moreover, by hypothesis, each of the relevant $\lim^1$ terms vanish, so that

$$\begin{aligned} R^{*,*}(X) \underset{R^{*,*}}{\widehat{\otimes}} R^{*,*}(Z) &\cong \lim_n R^{*,*}(X_n) \underset{R^{*,*}}{\otimes} R^{*,*}(Z) \\ &\cong \lim_n R^{*,*}(X_n \wedge Z) \cong R^{*,*}(X \wedge Z). \end{aligned}$$

□

Corollary 2.24. *Let Z be a motivic spectrum. Then there are natural isomorphisms*

$$R^0(\mathbb{P}^\infty) \underset{R^0}{\widehat{\otimes}} R^0(Z) \longrightarrow R^0(\mathbb{P}^\infty_+ \wedge Z)$$

and

$$R^0(BGL) \underset{R^0}{\widehat{\otimes}} R^0(Z) \longrightarrow R^0(BGL_+ \wedge Z)$$

Proof. For each m,

$$BGL_m \simeq \operatorname{colim}_n \operatorname{Grass}_m(\mathbb{A}^n)$$

is a colimit of finite stably cellular motivic spaces such that, for each n, $R^{*,*}(\operatorname{Grass}_m(\mathbb{A}^n))$ is a free $R^{*,*}$-module and

$$\begin{aligned} R^0(\operatorname{Grass}_m(\mathbb{A}^n)) \otimes_{R^0} R^0(Z) \quad &\cong \quad R^0(\operatorname{Grass}_m(\mathbb{A}^n)_+ \wedge Z) \longrightarrow \\ R^0(\operatorname{Grass}_m(\mathbb{A}^{n-1})_+ \wedge Z) \quad &\cong \quad R^0(\operatorname{Grass}_m(\mathbb{A}^{n-1})) \otimes_{R_0} R^0(Z) \end{aligned}$$

is (split) surjective. It therefore follows from Proposition 2.23 that, for each m,

$$R^0(\operatorname{Grass}_m(\mathbb{A}^\infty)_+) \underset{R^0}{\widehat{\otimes}} R^0(Z) \cong R^0(\operatorname{Grass}_m(\mathbb{A}^\infty)_+ \wedge Z).$$

Taking $m = 1$ yields the result for $\mathbb{P}^\infty$; for BGL, we must consider the sequence

$$BGL \simeq \operatorname{colim}_m BGL_m \simeq \operatorname{Grass}_m(\mathbb{A}^\infty)$$

in which the maps come from a fixed isomorphism $\mathbb{A}^1 \oplus \mathbb{A}^\infty \cong \mathbb{A}^\infty$. Note, however, that it follows from the above, together with the (split) surjection

$$R^0(\mathrm{Grass}_m(\mathbb{A}^\infty) \longrightarrow R^0(\mathrm{Grass}_{m-1}(\mathbb{A}^\infty))$$

of Proposition 2.20, that, for each m,

$$R^0(\mathrm{Grass}_m(\mathbb{A}^\infty)_+ \wedge Z) \longrightarrow R^0(\mathrm{Grass}_{m-1}(\mathbb{A}^\infty)_+ \wedge Z)$$

is (split) surjective, so that the $\lim^1$ term vanishes and

$$R^0(BGL)\widehat{\underset{R^0}{\otimes}} R^0(Z) \cong \lim_m R^0(\mathrm{Grass}_m(\mathbb{A}^\infty)_+ \wedge Z) \cong R^0(BGL_+ \wedge Z).$$

□

2.7. Primitives in the oriented cohomology of BGL. Let R be an oriented periodic commutative motivic ring spectrum. As is shown in Section 4.3 of [20], the group completion

$$BGL_{\mathbb{Z}} \simeq \Omega^{1,0} B(BGL_{\mathbb{N}})$$

(usually written $\mathbb{Z} \times BGL$) of the additive monoid $BGL_{\mathbb{N}} = \coprod_{n \in \mathbb{N}} BGL_n$ fits into a fibration sequence

$$BGL \longrightarrow BGL_{\mathbb{Z}} \longrightarrow \mathbb{Z},$$

where $BGL \simeq \mathrm{colim}_n\, BGL_n$. As $BGL_{\mathbb{N}}$ is commutative up to homotopy, $BGL_{\mathbb{Z}}$ is an abelian group object in the motivic homotopy category.

Lemma 2.25. *Let* $\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R)$ *denote the abelian group of homotopy classes of additive maps* $BGL_{\mathbb{Z}} \to \Omega^\infty R$. *Then the inclusion*

$$\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R) \longrightarrow R^0(BGL_{\mathbb{Z}})$$

identifies $\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R)$ *with the abelian group of primitive elements in the Hopf algebra* $R^0(BGL_{\mathbb{Z}})$.

Proof. By definition, there is an equalizer diagram

$$\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R) \longrightarrow R^0(BGL_{\mathbb{Z}}) \rightrightarrows (BGL_{\mathbb{Z}} \times BGL_{\mathbb{Z}})$$

associated to the square

$$\begin{tikzcd} BGL_{\mathbb{Z}} \times BGL_{\mathbb{Z}} \arrow[r] \arrow[d] & BGL_{\mathbb{Z}} \arrow[d] \\ \Omega^\infty R \times \Omega^\infty R \arrow[r] & \Omega^\infty R \end{tikzcd}$$

in which the horizontal maps are the addition maps. Let δ denote the Hopf algebra diagonal

$$\delta : R^0(BGL_{\mathbb{Z}}) \longrightarrow R^0(BGL_{\mathbb{Z}} \times BGL_{\mathbb{Z}}) \cong R^0(BGL_{\mathbb{Z}})\widehat{\underset{R^0}{\otimes}} R^0(BGL_{\mathbb{Z}}).$$

Then the equalizer consists of those $f \in R^0(BGL_{\mathbb{Z}})$ such that $\delta(f) = f \otimes 1 + 1 \otimes f$. This identifies $\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R)$ with the primitive elements in $R^0(BGL_{\mathbb{Z}})$. □

Lemma 2.26. *There are natural isomorphisms*

$$\begin{aligned}\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R) &\cong \mathrm{Add}(BGL, \Omega^\infty R) \times \mathrm{Add}(\mathbb{Z}, \Omega^\infty R)\\ &\cong \mathrm{Add}(BGL, \Omega^\infty R) \times R^0.\end{aligned}$$

Proof. The product of additive maps is additive, and, in any category with finite products and countable coproducts, $\mathbb{Z} = \coprod_{\mathbb{Z}} 1$ is the free abelian group on the terminal object 1. □

Proposition 2.27. *The map*

$$\mathrm{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R) \longrightarrow R^0(BGL_1),$$

obtained by restricting an additive map $BGL_{\mathbb{Z}} \to \Omega^\infty R$ along the inclusion $BGL_1 \to BGL_{\mathbb{Z}}$, is an isomorphism.

Proof. By Lemma 2.26, it's enough to show that the inclusion $(BGL_1, 1) \to (BGL, 1)$ induces an isomorphism

$$\mathrm{Add}(BGL, \Omega^\infty R) \longrightarrow R^0(BGL_1, 1).$$

Thus let $M = R_0(BGL_1, 1)$, and consider the R_0-algebra

$$A := \bigoplus_{n\geq 0} \mathrm{Sym}^n_{R_0} M$$

together with its augmentation ideal

$$I := \bigoplus_{n>0} \mathrm{Sym}^n_{R_0} M.$$

We have isomorphisms of split short exact sequences

$$\begin{tikzcd} 0 \arrow[r] & R^0(BGL,1) \arrow[r] \arrow[d] & R^0(BGL) \arrow[r] \arrow[d] & R^0 \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & \hom_{R_0}(I, R_0) \arrow[r] & \hom_{R_0}(A, R_0) \arrow[r] & \hom_{R_0}(R_0, R_0) \arrow[r] & 0 \end{tikzcd}$$

and

$$\begin{tikzcd} 0 \arrow[r] & R^0(BGL^{\times 2}, BGL^{\vee 2}) \arrow[r] \arrow[d] & R^0(BGL^{\times 2}) \arrow[r] \arrow[d] & \cdots \\ 0 \arrow[r] & \hom_{R_0}(I \otimes_{R_0} I, R_0) \arrow[r] & \hom_{R_0}(A \otimes_{R_0} A, R_0) \arrow[r] & \cdots \end{tikzcd}$$

$$\begin{tikzcd} \cdots \arrow[r] & R^0(BGL^{\vee 2}) \arrow[r] \arrow[d] & 0 \\ \cdots \arrow[r] & \hom_{R_0}(R_0 \oplus I^{\oplus 2}, R_0) \arrow[r] & 0 \end{tikzcd}$$

of R^0-modules. According to Lemmas 2.25 and 2.26, we have an exact sequence

$$0 \longrightarrow \mathrm{Add}(BGL, \Omega^\infty R) \longrightarrow R^0(BGL, 1) \longrightarrow R^0(BGL^{\times 2}, BGL^{\vee 2})$$

in which the map on the right is the cohomology of the map

$$\mu - p_1 - p_2 : (BGL^{\times 2}, BGL^{\vee 2}) \longrightarrow (BGL, 1)$$

(μ is the addition and the p_i are the projections); moreover, this map is the R_0-module dual of the multiplication $I \otimes_{R_0} I \to I$. Hence these short exact sequences assemble into a diagram

$$\begin{tikzcd}
0 \arrow[r] & \hom_{R_0}(I/I^2, R_0) \arrow[r] \arrow[d] & \mathrm{Add}(BGL, \Omega^\infty R) \arrow[r] \arrow[d] & \cdots \\
0 \arrow[r] & \hom_{R_0}(I, R_0) \arrow[r] \arrow[d] & R^0(BGL) \arrow[r] \arrow[d] & \cdots \\
0 \arrow[r] & \hom_{R_0}(I \otimes_{R_0} I, R_0) \arrow[r] & R^0(BGL^{\times 2}) \arrow[r] & \cdots
\end{tikzcd}$$

$$\begin{tikzcd}
\cdots \arrow[r] & 0 \arrow[r] \arrow[d] & 0 \\
\cdots \arrow[r] & R^0 \arrow[r] \arrow[d] & 0 \\
\cdots \arrow[r] & \hom_{R_0}(R_0 \oplus I^{\oplus 2}, R_0) \arrow[r] & 0
\end{tikzcd}$$

of short exact sequences by the snake lemma. In particular, we see that $\mathrm{Add}(BGL, \Omega^\infty R)$ is naturally identified with the dual $\hom_{R_0}(I/I^2, R_0)$ of the module of indecomposables I/I^2. But $I/I^2 \cong M = R_0(BGL_1, 1)$, the duality map

$$R^0(BGL_1, 1) \longrightarrow \hom_{R_0}(R_0(BGL_1, 1), R_0)$$

is an R^0-module isomorphism, and the restriction $R^0(BGL, 1) \to R^0(BGL_1, 1)$ is dual to the inclusion $M \to I$. $\square$

3. Algebraic Cobordism

3.1. The representing spectrum. For each natural number n, let $V_n \to BGL_n$ denote the universal n-plane bundle over BGL_n. Then the Thom spaces

$$MGL_n := (V_n, V_n - BGL_n)$$

come equipped with natural maps

$$MGL_p \wedge MGL_q \longrightarrow MGL_{p+q}$$

defined as the composite of the isomorphism

$$(V_p, V_p - BGL_p) \wedge (V_q, V_q - BGL_q) \longrightarrow (V_p \times V_q, V_p \times V_q - BGL_p \times BGL_q)$$

and the map on Thom spaces associated to the inclusion of vector bundles

$$\begin{tikzcd} V_p \times V_q \arrow[r] \arrow[d] & V_{p+q} \arrow[d] \\ BGL_p \times BGL_q \arrow[r] & BGL_{p+q}. \end{tikzcd}$$

Restricting this map of vector bundles along the inclusion $1 \times BGL_q \to BGL_p \times BGL_q$ gives a map of Thom spaces

$$(\mathbb{A}^p, \mathbb{A}^p - \mathbb{A}^0) \wedge MGL_q \to MGL_{p+q},$$

and these maps comprise the structure maps of the prespectrum MGL. The associated spectrum is defined by

$$MGL(p) := \operatorname{colim}_q \Omega^q MGL_{p+q},$$

as evidently the adjoints

$$MGL(q) \simeq \operatorname{colim}_r \Omega^r MGL_{q+r} \simeq \operatorname{colim}_r \Omega^{p+r} MGL_{p+q+r} \simeq \Omega^p MGL(p+q)$$

of the structure maps $\Sigma^p MGL(q) \to MGL(p+q)$ are equivalences. The last equivalence uses the fact that $\mathbb{P}^1$ is a compact object of the motivic homotopy category.

Definition 3.1 (Voevodsky [34]). Algebraic cobordism is the motivic cohomology theory represented by the motivic spectrum MGL.

3.2. Algebraic cobordism is the universal oriented motivic spectrum. Just as in ordinary stable homotopy theory, the Thom classes $\theta_n \in R^n(MGL_n)$ coming from an orientation on a commutative motivic ring spectrum R assemble to give a ring map $\theta : MGL \to R$. We begin with a brief review of this correspondence.

Proposition 3.2 (Panin, Pimenov, Röndigs [23]). *Let R be a commutative monoid in the homotopy category of motivic spectra. Then the set of monoidal maps $MGL \to R$ is naturally isomorphic to the set of orientations on R.*

Proof. The classical analysis of complex orientations on ring spectra R generalizes immediately. A spectrum map $\theta : MGL \to R$ is determined by a compatible family of maps $\theta_n : MGL_n \to R^n$, which is to say a family of universal Thom classes $\theta_n \in R^n(MGL_n)$. An arbitrary n-plane bundle $V \to X$, represented by a map $X \to BGL_n$, induces a map of Thom spaces $V/V - X \to MGL_n$, so θ_n restricts to a Thom class in $R^n(V/V - X)$. Moreover, these Thom classes are multiplicative and unital precisely when $\theta : MGL \to R$ is monoidal. Conversely, an orientation on R has, as part of its data, Thom classes $\theta_n \in R^n(MGL_n)$ for the universal bundles $V_n \to BGL_n$, and these assemble to form a ring map $\theta : MGL \to R$. □

Again, just as in topology, an orientation on R is equivalent to a compatible family of R-theory Chern classes for vector bundles $V \to X$. This follows from the Thom isomorphism $R^*(BGL_n) \cong R^*(MGL_n)$.

More difficult is the fact that an orientation on a ring spectrum R is uniquely determined by the first Thom class alone; that is, a class $\theta_1 \in R^1(BGL_1) = R$ whose restriction $i^*\theta_1 \in R^1(S^1)$ along the inclusion $S^1 \to MGL_1$ corresponds to $1 \in R^0(S^0)$ via the suspension isomorphism $R^1(S^1) \cong R^0(S^0)$. This is a result of the splitting principle, which allows one to construct Thom classes (or Chern classes) for general vector bundles by descent from a space over which they split. See Adams [1] and Panin-Pimenov-Röndigs [23] for details.

3.3. A ring spectrum equivalent to $PMGL$. The wedge

$$\bigvee_{n\in\mathbb{N}} \Sigma^\infty MGL_n$$

forms a ring spectrum with unit $\mathbb{S} \simeq \Sigma^\infty MGL_0$ and multiplication

$$\bigvee_p \Sigma^\infty MGL_p \wedge \bigvee_q \Sigma^\infty MGL_q \longrightarrow \bigvee_{p,q} \Sigma^\infty MGL_p \wedge MGL_q \longrightarrow \bigvee_n \Sigma^\infty MGL_n$$

induced by the maps $MGL_p \wedge MGL_q \to MGL_{p+q}$. Evidently, a (homotopy class of a) ring map $\bigvee_n \Sigma^\infty MGL_n \to R$ is equivalent to a family of *degree zero* Thom classes

$$\vartheta_n \in R^0(MGL_n)$$

with $\vartheta_0 = 1 \in R^0(MGL_0) = R^0$ such that ϑ_{p+q} restricts via $MGL_p \wedge MGL_q \to MGL_{p+q}$ to the product $\vartheta_p\vartheta_q$. This is *not* the same as an orientation on R, as there is nothing forcing $\vartheta_1 \in R^0(MGL_1)$ to restrict to a unit in $R^0(S^1)$. Clearly we should impose this condition, which amounts to inverting $\beta : \mathbb{P}^1 \to \mathbb{P}^\infty$.

Proposition 3.3. *A ring map $PMGL \to R$ induces a ring map $\bigvee_n \Sigma^\infty MGL_n[1/\beta] \to R$.*

Proof. A ring map $\theta : PMGL \to R$ consists of a ring map $MGL \to R$ and a unit $\mu \in \pi_1 R$. This specifies Thom classes $\theta_n \in R^n(MGL_n)$, and therefore Thom classes

$$\vartheta_n := \mu^n\theta_n \in R^0(MGL_n)$$

such that

$$\vartheta_p\vartheta_q = \mu^{p+q}\theta_p\theta_q = \mu^{p+q}i^*\theta_{p+q} = i^*\vartheta_{p+q} \in R^0(MGL_p \wedge MGL_q),$$

where i is the map $MGL_p \wedge MGL_q \to MGL_{p+q}$. This gives a ring map $\vartheta : \bigvee_n \Sigma^\infty MGL_n \to R$, and therefore the desired map, provided β is sent to a unit. But this is clear: as a class in $R^0(S^1)$,

$$\vartheta(\beta) = \beta^*\vartheta_1 = \mu\beta^*\theta_1,$$

and $\beta^*\theta_1 \in R^1(S^1)$ is the image of $1 \in R^0(S^0)$ under the isomorphism $R^0(S^0) \cong R^1(S^1)$. □

Proposition 3.4. *The ring map $\bigvee_{n\in\mathbb{N}} \Sigma^\infty MGL_n[1/\beta] \to PMGL$ is an equivalence.*

Proof. Write
$$M := \bigvee_{n \in \mathbb{N}} \Sigma^\infty MGL_n[1/\beta] \to PMGL,$$
and consider the natural transformation of set-valued functors
$$\mathrm{Rings}(PMGL, -) \longrightarrow \mathrm{Rings}(M, -).$$
Given a ring spectrum R, we have seen that the set $\mathrm{Rings}(M, R)$ is naturally isomorphic to the set of collections $\{\vartheta_n\}_{n \in \mathbb{N}}$ with $\vartheta_n \in R^0(MGL_n)$ such that ϑ_{p+q} restricts to $\vartheta_p \vartheta_q$, ϑ_1 restricts to a unit in R^{-1}, and $\vartheta_0 = 1 \in R^0(S^0)$. Similarly, the set $\mathrm{Rings}(PMGL, R)$ is naturally isomorphic to the product of the set of units in R^{-1} and the set of collections $\{\theta_n\}_{n \in \mathbb{N}}$ with $R^n(MGL_n)$ such that θ_{p+q} restricts to $\theta_p \theta_q$, θ_1 restricts to the image of $1 \in R^0(S^0)$ in $R^1(S^1)$, and $\theta_0 = 1 \in R^0(S^0)$.
The map $\mathrm{Rings}(PMGL, R) \to \mathrm{Rings}(M, R)$ sends $\mu \in R^{-1}$ and $\theta_n \in R^n(MGL_n)$ to $\vartheta_n = \mu^n \theta_n$. We get a natural map back which sends $\vartheta_n \in R^0(MGL_n)$ to $\theta_n = \mu^{-n} \vartheta_n$, where $\mu \in R^{-1}$ in the unit corresponding to $\beta^* \vartheta_1 \in R^0(S^1)$. Clearly the composites are the respective identities, and we conclude that $M \to PMGL$ is an equivalence. □

3.4. $\Sigma_+^\infty BGL[1/\beta]$ is orientable. Recall from [23] that, just as in the usual stable homotopy category, an orientation on a ring spectrum R is equivalent to a class in $R^1(MGL_1)$ which restricts, under the inclusion $i : S^1 \to MGL_1$ of the bottom cell, to the class in $R^1(S^1)$ corresponding to the unit $1 \in R^0(S^0)$ under the suspension isomorphism $R^0(S^0) \to R^1(S^1)$. Note also that in the case R is periodic with Bott element $\beta \in R^0(S^1)$, corresponding under the suspension isomorphism to the unit $\mu \in R^{-1}(S^0)$ with inverse $\mu^{-1} \in R^1(S^0)$, then the suspension isomorphism $R^0(S^0) \to R^1(S^1)$ sends 1 to $\mu^{-1}\beta$.
Now there's a canonical class $\theta_1 \in \Sigma_+^\infty BGL[1/\beta]^1(MGL_1)$ such that
$$\mu^{-1}\beta = i^*\theta_1 \in \Sigma_+^\infty BGL[1/\beta]^1(S^1).$$
Namely, set $\theta_1 := \mu^{-1}\vartheta_1$, where $\vartheta_1 \in \Sigma_+^\infty BGL[1/\beta]^0(MGL_1)$ is the class of the composite
$$\Sigma^\infty MGL_1 \simeq \Sigma^\infty BGL_1 \longrightarrow \Sigma_+^\infty BGL \longrightarrow \Sigma_+^\infty BGL[1/\beta].$$
Then $\beta = i^*\mu\theta$, so $\mu^{-1}\beta = i^*\theta$.

Proposition 3.5. *There is a canonical ring map $\theta : PMGL \to \Sigma_+^\infty BGL[1/\beta]$.*

Proof. The Thom class $\theta_1 \in \Sigma_+^\infty BGL[1/\beta]^0(MGL_1)$ extends, as in [1] or [23], to a ring map $MGL \to \Sigma_+^\infty BGL[1/\beta]$, and we have a canonical unit $\mu \in R^{-1}(S^0)$, the image of $\beta \in R^0(S^1)$ under the suspension isomorphism $R^0(S^1) \cong R^{-1}(S^0)$. □

Corollary 3.6. *There is a canonical ring map $\vartheta : \bigvee_n \Sigma^\infty MGL_n[1/\beta] \to \Sigma_+^\infty BGL[1/\beta]$.*

Proof. Precompose the map from the previous Proposition 3.5 with the equivalence

$$\bigvee_n \Sigma^\infty MGL_n[1/\beta] \to PMGL.$$

□

3.5. ϑ is an equivalence. We analyze the effect of $\vartheta : \bigvee_n \Sigma^\infty MGL_n[1/\beta] \to \Sigma^\infty_+ BGL[1/\beta]$ on cohomology. To this end, fix an oriented periodic commutative motivic ring spectrum R; we aim to show that the induced map

$$R^0(\Sigma^\infty_+ BGL[1/\beta]) \longrightarrow R^0(\bigvee_n \Sigma^\infty MGL_n[1/\beta])$$

is an isomorphism.

Lemma 3.7. *Let R be a commutative ring and let $A = \operatorname{colim}_n A_n$ be a filtered commutative R-algebra with the property that $A_p \otimes_R A_q \to A \otimes_R A \to A$ factors through the inclusion $A_{p+q} \to A$ (that is, the multiplication is compatible with the filtration). Suppose that, for each n, the maps $A_{n-1} \to A_n$ are split injections, so that the isomorphisms $A_{n-1} \oplus A_n/A_{n-1} \to A_n$ define an R-module isomorphism*

$$\mathrm{gr}A := \bigoplus_n A_n/A_{n-1} \xrightarrow{\simeq} \operatorname{colim}_n A_n = A$$

of A with its associated graded. Then the multiplication $A_p/A_{p-1} \otimes_R A_q/A_{q-1} \to A_{p+q}/A_{p+q-1}$ makes $\mathrm{gr}A = \bigoplus_n A_n/A_{n-1}$ into a commutative R-algebra in such a way that $\mathrm{gr}A \to A$ is an R-algebra isomorphism. □

Proposition 3.8. *There is a commuting square of R^0-module maps*

$$\begin{array}{ccc} R^0(BGL) & \longrightarrow & \prod_n R^0(MGL_n) \\ \downarrow & & \downarrow \\ R^0(BGL \times BGL) & \longrightarrow & \prod_{p,q} R^0(MGL_p \wedge MGL_q) \end{array},$$

in which the vertical maps are induced by the multiplication on BGL and $\bigvee_n MGL_n$, respectively, and the horizontal maps are isomorphisms.

Proof. Set $A := \operatorname{colim}_n \operatorname{Sym}^n_{R_0} R_0(\mathbb{P}^\infty)$, where the map $\operatorname{Sym}^{n-1}_{R_0} R_0(\mathbb{P}^\infty) \to \operatorname{Sym}^n_{R_0} R_0(\mathbb{P}^\infty)$ is induced by the the inclusion $R_0 \cong R_0(\mathbb{P}^0) \to R_0(\mathbb{P}^\infty)$. Applying R_0 to the cofiber sequence $BGL_{n-1} \to BGL_n \to MGL_n$ yields split short exact sequences

$$\begin{array}{ccccc} \operatorname{Sym}^{n-1} R_0(\mathbb{P}^\infty) & \longrightarrow & \operatorname{Sym}^n R_0(\mathbb{P}^\infty) & \longrightarrow & \operatorname{Sym}^n R_0(\mathbb{P}^\infty)/\operatorname{Sym}^{n-1} R_0(\mathbb{P}^\infty) \\ \downarrow \cong & & \downarrow \cong & & \downarrow \cong \\ R_0(BGL_{n-1}) & \longrightarrow & R_0(BGL_n) & \longrightarrow & R_0(MGL_n) \end{array}$$

with $A \cong \operatorname{colim}_n R_0(BGL_n) \cong R_0(BGL)$ a filtered commutative R-algebra. By the lemma, we have a commutative square

$$\begin{tikzcd} \bigoplus_{p,q} \operatorname{Sym}^p R_0(\mathbb{P}^\infty)/\operatorname{Sym}^{p-1} R_0(\mathbb{P}^\infty) \underset{R_0}{\otimes} \operatorname{Sym}^q R_0(\mathbb{P}^\infty)/\operatorname{Sym}^{q-1} R_0(\mathbb{P}^\infty) \arrow[r] \arrow[d] & A \underset{R_0}{\otimes} A \arrow[d] \\ \bigoplus_n \operatorname{Sym}^n R_0(\mathbb{P}^\infty)/\operatorname{Sym}^{n-1} R_0(\mathbb{P}^\infty) \arrow[r] & A \end{tikzcd}$$

in which the vertical maps are multiplication and the horizontal maps are R_0-algebra isomorphisms. The desired commutative square is obtained by taking R_0-module duals. □

Theorem 3.9. *The map of oriented periodic motivic ring spectra*

$$\vartheta : \bigvee_n \Sigma^\infty MGL_n[1/\beta] \to \Sigma^\infty_+ BGL[1/\beta]$$

is an equivalence.

Proof. We show that the induced natural transformation

$$\vartheta^* : \operatorname{Rings}(\Sigma^\infty_+ BGL[1/\beta], -) \longrightarrow \operatorname{Rings}(\bigvee_n \Sigma^\infty MGL_n[1/\beta], -)$$

is in fact a natural isomorphism. The result then follows immediately from Yoneda's Lemma.

Fix a ring spectrum R, and observe that, for another ring spectrum A, $\operatorname{Rings}(A, R)$ is the equalizer of the pair of maps from $R^0(A)$ to $R^0(A \wedge A) \times R^0(\mathbb{S})$ which assert the commutativity of the diagrams

$$\begin{tikzcd} A \wedge A \arrow[r] \arrow[d] & A \arrow[d] \\ R \wedge R \arrow[r] & R \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} \mathbb{S} \arrow[r] \arrow[dr] & A \arrow[d] \\ & R \end{tikzcd} .$$

Given a map $\beta : \Sigma^1 \mathbb{S} \to A$, the set $\operatorname{Rings}(A[1/\beta], R)$ is the equalizer of the pair of maps from $\operatorname{Rings}(A, R) \times R^0(\Sigma^{-1}\mathbb{S})$ to $R^0(\mathbb{S})$ which assert that the ring map $A \to R$ is such that there's a spectrum map $\Sigma^{-1}\mathbb{S} \to R$ for which the product

$$\mathbb{S} \simeq \Sigma^1 \mathbb{S} \wedge \Sigma^{-1}\mathbb{S} \longrightarrow A \wedge R \longrightarrow R \wedge R \longrightarrow R$$

is equivalent to the unit $\mathbb{S} \to R$. Putting these together, we may express $\operatorname{Rings}(A[1/\beta], R)$ as the equalizer of natural pair of maps from $R^0(A) \times R^0(\mathbb{S}^{-1})$ to $R^0(A \wedge A) \times R^0(\mathbb{S}) \times R^0(\mathbb{S})$.

We therefore get a map of equalizer diagrams

$$\begin{tikzcd} R^0(BGL) \times R^0(\Sigma^{-1}\mathbb{S}) \arrow[r, shift left] \arrow[r, shift right] \arrow[d] & R^0(BGL \times BGL) \times R^0(\mathbb{S}) \times R^0(\mathbb{S}) \arrow[d] \\ \prod_n R^0(MGL_n) \times R^0(\Sigma^{-1}\mathbb{S}) \arrow[r, shift left] \arrow[r, shift right] & \prod_{p,q} R^0(MGL_p \wedge MGL_q) \times R^0(\mathbb{S}) \times R^0(\mathbb{S}) \end{tikzcd} ,$$

the equalizer of which is ϑ^*. Now if R does *not* admit the structure of a $PMGL$-algebra, then clearly there cannot be any ring maps from either of the $PMGL$-algebras $\bigvee_n \Sigma^\infty MGL_n[1/\beta]$ of $\Sigma^\infty_+ BGL[1/\beta]$. Hence we may assume that R is also an oriented periodic ring spectrum, in which case Proposition 3.8 implies that the vertical maps are isomorphisms. □

Corollary 3.10. *The map of periodic oriented motivic ring spectra*

$$\theta : PMGL \to \Sigma^\infty_+ BGL[1/\beta]$$

is an equivalence.

Proof. $\bigvee_n \Sigma^\infty MGL_n[1/\beta] \to PMGL$ is an equivalence. □

4. Algebraic K-Theory

4.1. The representing spectrum. Let $BGL_{\mathbb{Z}} \simeq \mathbb{Z} \times BGL$ denote the group completion of the monoid

$$BGL_{\mathbb{N}} := \coprod_{n \in \mathbb{N}} BGL_n.$$

Given a motivic space X, write $K^0(X) := \pi_0 \operatorname{map}_S(X, BGL_{\mathbb{Z}})$. If $S = \operatorname{spec} \mathbb{Z}$ and X is a scheme, this agrees with the homotopy algebraic K-theory of X as defined by Weibel [38], and if in addition X is smooth, this also agrees with Thomason-Trobaugh algebraic K-theory of X [33]; see Proposition 4.3.9 of [20] for details. As the name suggests, homotopy algebraic K-theory is a homotopy invariant version of the Thomason-Trobaugh algebraic K-theory, and homotopy invariance is of course a prerequisite for any motivic cohomology theory.
It turns out that the motivic space $BGL_{\mathbb{Z}}$, pointed by the inclusion

$$1 \simeq BGL_0 \longrightarrow BGL_{\mathbb{N}} \longrightarrow BGL_{\mathbb{Z}},$$

is the zero space of the motivic spectrum K representing (homotopy) algebraic K-theory. This is a direct corollary of the following famous fact.

Proposition 4.1 (Motivic Bott Periodicity). *The adjoint*

$$(BGL_{\mathbb{Z}}, BGL_0) \longrightarrow \Omega(BGL_{\mathbb{Z}}, BGL_0)$$

of the map Bott map $\Sigma(BGL_{\mathbb{Z}}, BGL_0) \to (BGL_{\mathbb{Z}}, BGL_0)$ classifying the tensor product of $(L-1)$ and V, where $L \to \mathbb{P}^1$ is the restriction of the universal line bundle and $V \to BGL_{\mathbb{Z}}$ is the universal virtual vector bundle, is an equivalence.

Proof. Quillen's projective bundle theorem [26] implies that the tensor product of vector bundles induces an isomorphism

$$K^0(\mathbb{P}^1) \underset{K^0}{\otimes} K^0(X) \longrightarrow K^0(\mathbb{P}^1 \times X)$$

of abelian groups. It follows that there's an isomorphism of split short exact sequences of $K^0(X)$-modules

$$\begin{tikzcd} \lambda K^0(X)[\lambda]/(\lambda^2) \arrow[r] \arrow[d] & K^0(X)[\lambda]/(\lambda^2) \arrow[r] \arrow[d] & K^0(X) \arrow[d] \\ K^0((\mathbb{P}^1,\mathbb{P}^0)\wedge(X,0)) \arrow[r] & K^0(\mathbb{P}^1\times X) \arrow[r] & K^0(\mathbb{P}^0\times X)\ . \end{tikzcd}$$

In particular, $K^0((\mathbb{P}^1,\mathbb{P}^0)\wedge(X,0)) \cong K^0(X)$ and similarly $K^0((\mathbb{P}^1,\mathbb{P}^0)\wedge(X,1)) \cong K^0(X,1)$. □

Define a sequence of pointed spaces $K(n)$ by

$$K(n) := (BGL_{\mathbb{Z}}, BGL_0)$$

for all $n \in \mathbb{N}$. By Proposition 4.1, each $K(n)$ comes equipped with an equivalence

$$K(n) = (BGL_{\mathbb{Z}}, BGL_0) \longrightarrow \Omega(BGL_{\mathbb{Z}}, BGL_0) = \Omega K(n+1),$$

making $K := (K(0), K(1), \ldots)$ into a motivic spectrum.

4.2. A map $\Sigma^\infty_+\mathbb{P}^\infty[1/\beta] \to K$. Let $\beta : \mathbb{P}^1 \to \mathbb{P}^\infty$ be the map classifying the tautological line bundle on $\mathbb{P}^1$. We construct a ring map $\Sigma^\infty_+\mathbb{P}^\infty \to K$ which sends β to a unit in K, thus yielding a ring map $\Sigma^\infty_+\mathbb{P}^\infty[1/\beta] \to K$. There is a homotopy commutative ring structure on the motivic space $BGL_{\mathbb{Z}} \simeq \Omega^\infty K$ in which addition is induced by the sum of vector bundles and multiplication is induced by the tensor product of vector bundles.
Ring maps $\Sigma^\infty_+\mathbb{P}^\infty \to K$ are adjoint to monoidal maps $\mathbb{P}^\infty \to GL_1K$, the multiplicative monoid of units (up to homotopy) in the ring space $\Omega^\infty K \simeq BGL_{\mathbb{Z}}$. Since $\pi_0 BGL_{\mathbb{Z}}$ contains a copy of $\mathbb{Z}$, the multiplicative units contain the subgroup $\{\pm 1\} \to \mathbb{Z}$, giving a map

$$\{\pm 1\} \times BGL \longrightarrow GL_1K.$$

But the inclusion $BGL_1 \to BGL$ is monoidal with respect to the multiplicative structure on BGL, so we get a monoidal map

$$\mathbb{P}^\infty \simeq BGL_1 \longrightarrow \{+1\} \times BGL \longrightarrow GL_1K$$

and therefore a ring map $\Sigma^\infty_+\mathbb{P}^\infty \to K$.

Proposition 4.2. *The class of the composite*

$$\Sigma^\infty S^1 \simeq \Sigma^\infty(\mathbb{P}^1,\mathbb{P}^0) \longrightarrow \Sigma^\infty_+\mathbb{P}^\infty \longrightarrow K$$

is equal to that of the K-theory Bott element β, i.e. the class of the reduced tautological line bundle $L-1$ on $\mathbb{P}^1$.

Proof. The map $\Sigma^\infty_+\mathbb{P}^\infty \to K$ classifies the tautological line bundle on $\mathbb{P}^\infty$, so the pointed version $\Sigma^\infty\mathbb{P}^\infty \to K$ corresponds to the reduced tautological line bundle on $\mathbb{P}^\infty$. This restricts to the reduced tautological line bundle on $\mathbb{P}^1$. □

Corollary 4.3. *There's a canonical ring map* $\psi : \Sigma^\infty_+ \mathbb{P}^\infty[1/\beta] \to K$. $\square$

4.3. Comparing $R^0(K)$ and $R^0(L)$. Let L denote the localized motivic ring spectrum

$$L := \Sigma^\infty_+ \mathbb{P}^\infty[1/\beta].$$

That is, L is the colimit

$$L = \operatorname{colim}_n \Sigma^{\infty - n} \mathbb{P}^\infty_+$$

of a telescope of desuspended suspension spectra. We show that $\psi : L \to K$ is an equivalence by showing the induced map $R^0(L) \to R^0(K)$ is an isomorphism for a sufficiently large class of motivic spectra R.

Throughout this section, we will be considering the motivic space $BGL_{\mathbb{Z}}$ (pointed by $\{0\} \times BGL_0$) multiplicatively, as a homotopy commutative monoid with respect to the smash product. Accordingly, $\Sigma^\infty BGL_{\mathbb{Z}}$ is a ring spectrum, and the monoidal map

$$\mathbb{P}^\infty_+ \simeq BGL_0 + BGL_1 \longrightarrow BGL_{\mathbb{N}} \longrightarrow BGL_{\mathbb{Z}}$$

gives $\Sigma^\infty BGL_{\mathbb{Z}}$ the structure of a homotopy commutative $\Sigma^\infty \mathbb{P}^\infty_+$-algebra. In particular, the Bott element $\beta \in \pi_1 \Sigma^\infty \mathbb{P}^\infty_+$ determines a Bott element $\beta \in \pi_1 \Sigma^\infty BGL_{\mathbb{Z}}$ as well as a Bott element $\beta \in \pi_1 K$.

If R is a homotopy commutative ring spectrum equipped with homotopy element $\alpha \in \pi_n R$, we write

$$\mu(\alpha) = \Sigma^{-n}(\mu \circ (\alpha \wedge R)) : R \longrightarrow \Sigma^{-n} R$$

for the "multiplication by α" map, the n-fold desuspension of the composite

$$\Sigma^n R \simeq \Sigma^n \mathbb{S} \wedge R \xrightarrow{\alpha \wedge R} R \wedge R \xrightarrow{\mu} R.$$

If $\alpha \in \pi_n R$ is a unit, then this map has an inverse $\mu(\alpha)^{-1} : \Sigma^{-n} R \to R$, the n-fold desuspension of the multiplication by α^{-1} map $\mu(\alpha^{-1}) : R \to \Sigma^n R$. For our purposes, R will typically admit a periodic orientation, and α will be the image of the Bott element $\beta \in \pi_1 PMGL \cong \pi_1 \Sigma^\infty_+ BGL[1/\beta]$ under some ring map $PMGL \to R$.

Finally, we also write

$$\mu(\beta) : BGL_{\mathbb{Z}} \longrightarrow \Omega BGL_{\mathbb{Z}}$$

for multiplication by β in the homotopy commutative monoid $BGL_{\mathbb{Z}}$ (regarded multiplicatively). This is Ω^∞ applied to the multiplication by β map $\mu(\beta) : K \to \Sigma^{-1} K$ on K-theory, and thus it is the equivalence adjoint to the Bott map

$$\Sigma BGL_{\mathbb{Z}} \longrightarrow BGL_{\mathbb{Z}}$$

The following lemma is formal.

Lemma 4.4. *Let* $\varepsilon : \Sigma\Omega BGL_{\mathbb{Z}} \to BGL_{\mathbb{Z}}$ *denote the counit of the adjunction* (Σ, Ω) *applied to* $BGL_{\mathbb{Z}}$*. Then the composite*

$$\Sigma BGL_{\mathbb{Z}} \xrightarrow{\Sigma\mu(\beta)} \Sigma\Omega BGL_{\mathbb{Z}} \xrightarrow{\varepsilon} BGL_{\mathbb{Z}}$$

is the Bott map $\Sigma BGL_{\mathbb{Z}} \longrightarrow BGL_{\mathbb{Z}}$.

Proof. More generally, if (Σ, Ω) is any adjunction and $\beta^* : \Sigma X \to Y$ is a map left adjoint to $\beta_* : X \to \Omega Y$, then $\beta^* = \varepsilon_Y \circ \Sigma\beta_*$. $\square$

Proposition 4.5. *The square*

$$\begin{tikzcd} \Sigma^{\infty+1}\mathbb{P}^\infty_+ \arrow[r, "\Sigma\mu(\beta)"] \arrow[d] & \Sigma^\infty\mathbb{P}^\infty_+ \arrow[d] \\ \Sigma^{\infty+1}BGL_{\mathbb{Z}} \arrow[r, "\Sigma\mu(\beta)"] & \Sigma^\infty BGL_{\mathbb{Z}} , \end{tikzcd}$$

in which vertical maps come from the inclusion $i : \mathbb{P}^\infty_+ \simeq BGL_0 + BGL_1 \to BGL_{\mathbb{Z}}$ *and the horizontal maps are the Bott maps, commutes up to homotopy.*

Proof. The inclusion $\Sigma^\infty\mathbb{P}^\infty_+ \to \Sigma^\infty BGL_{\mathbb{Z}}$ is a map of homotopy commutative ring spectra, and the Bott element $\Sigma^\infty\mathbb{P}^1 \to \Sigma^\infty BGL_{\mathbb{Z}}$ factors through the Bott element $\Sigma^\infty\mathbb{P}^1 \to \Sigma^\infty\mathbb{P}^\infty_+$. $\square$

Proposition 4.6. *Let R be a homotopy commutative $PMGL$-algebra. Then the space* $\mathrm{map}(K, R)$ *of maps from K to R is equivalent to the homotopy inverse limit*

$$\mathrm{map}(K, R) \simeq \mathrm{holim}_n\{\cdots \xrightarrow{f} \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \xrightarrow{f} \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R)\},$$

where $f = \mu(\alpha)^{-1} \circ \Sigma^{-1} \circ \mu(\beta)$ *is the endomorphism of* $\mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R)$ *which sends a map* $x : \Sigma^\infty BGL_{\mathbb{Z}} \to R$ *to the composite*

$$\Sigma^\infty BGL_{\mathbb{Z}} \xrightarrow{\mu(\beta)} \Sigma^{\infty-1}BGL_{\mathbb{Z}} \xrightarrow{\Sigma^{-1}x} \Sigma^{-1}R \xrightarrow{\mu(\alpha)^{-1}} R.$$

Proof. In general, for motivic spectra M and N,

$$\mathrm{map}(M, N) \simeq \mathrm{holim}\{\cdots \longrightarrow \mathrm{map}(M(1), N(1)) \longrightarrow \mathrm{map}(M(0), N(0))\},$$

where the maps send a map $x : M(n) \to N(n)$ to $\Omega x : M(n-1) \simeq \Omega M(n) \to \Omega N(n) \simeq N(n-1)$. By adjunction, we may rewrite this as

$$\mathrm{map}(M, N) \simeq \mathrm{holim}\{\cdots \longrightarrow \mathrm{map}(\Sigma^\infty\Omega^\infty\Sigma M, \Sigma N) \longrightarrow \mathrm{map}(\Sigma^\infty\Omega^\infty M, N)\}.$$

Now K and R are periodic via equivalences $\mu(\beta) : K \to \Sigma^{-1}K$ and $\mu(\alpha) : R \to \Sigma^{-1}R$, the diagram

$$\begin{tikzcd} \mathrm{map}(BGL_{\mathbb{Z}}, \Omega^\infty R) \arrow[r, "\Omega"] \arrow[d, "\simeq"] & \mathrm{map}(\Omega BGL_{\mathbb{Z}}, \Omega^{\infty+1}R) \arrow[r, "\mu(\beta)^*"] \arrow[d, "\simeq"] & \mathrm{map}(BGL_{\mathbb{Z}}, \Omega^{\infty+1}R) \arrow[d, "\simeq"] \\ \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \arrow[r, "\Sigma^\infty\varepsilon^*"] & \mathrm{map}(\Sigma^{\infty+1}\Omega BGL_{\mathbb{Z}}, R) \arrow[r, "\Sigma^{\infty+1}\mu(\beta)^*"] & \mathrm{map}(\Sigma^{\infty+1}BGL_{\mathbb{Z}}, R) , \end{tikzcd}$$

in which the vertical arrows are adjunction equivalences, commutes, and according to Lemma 4.4 above, the Σ^∞ applied to the composite $\varepsilon \circ \Sigma\mu(\beta)$ is $\Sigma\mu(\beta) : \Sigma^{\infty+1}BGL_{\mathbb{Z}} \to \Sigma^\infty BGL_{\mathbb{Z}}$. Hence

$$\mathrm{map}(K, R) \simeq \mathrm{holim}\{\cdots \longrightarrow \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \longrightarrow \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R)\}$$

is the homotopy inverse limit of the tower determined by the composite

$$\mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \xrightarrow{\Sigma\mu(\beta)^*} \mathrm{map}(\Sigma^{\infty+1} BGL_{\mathbb{Z}}, R) \xrightarrow{\mu(\alpha)^{-1}_* \circ \Sigma^{-1}} \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R).$$

That is, the endomorphism f of $\mathrm{map}(\Sigma^\infty BGL, R)$ sends $x : \Sigma^\infty BGL \to R$ to the composite $\mu(\alpha)^{-1} \circ \Sigma^{-1}(x \circ \Sigma\mu(\beta)) = \mu(\alpha)^{-1} \circ \Sigma^{-1}(x) \circ \mu(\beta)$, which is to say that $f = \mu(\alpha)^{-1} \circ \Sigma^{-1} \circ \mu(\beta)$. □

Proposition 4.7. *Let R be a homotopy commutative $PMGL$-algebra. Then the space* $\mathrm{map}(L, R)$ *of maps from L to R is equivalent to the homotopy inverse limit*

$$\mathrm{map}(L, R) \simeq \mathrm{holim}_n\{\cdots \xrightarrow{g} \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R) \xrightarrow{g} \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)\},$$

where $g = \mu(\alpha)^{-1} \circ \Sigma^{-1} \circ \mu(\beta)$ is the endomorphism of $\mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)$ *which sends a map $y : \Sigma^\infty \mathbb{P}^\infty_+ \to R$ to the composite*

$$\Sigma^\infty \mathbb{P}^\infty_+ \xrightarrow{\mu(\beta)} \Sigma^{\infty-1} \mathbb{P}^\infty_+ \xrightarrow{\Sigma^{-1} y} \Sigma^{-1} R \xrightarrow{\mu(\alpha)^{-1}} R.$$

Proof. By definition, $L = \Sigma^\infty_+ \mathbb{P}^\infty[1/\beta] = \mathrm{hocolim}_n\, \Sigma^{\infty-n} \mathbb{P}^\infty_+$, where the map

$$\Sigma^{\infty-n} \mathbb{P}^\infty_+ \to \Sigma^{\infty-n-1} \mathbb{P}^\infty_+$$

is the n-fold desuspension of the multiplication by β map $\mu(\beta) : \Sigma^\infty \mathbb{P}^\infty_+ \to \Sigma^{\infty-1} \mathbb{P}^\infty_+$. Hence

$$\begin{aligned} &\mathrm{map}(L, R) \simeq \\ &\qquad \simeq \mathrm{holim}_n\{\cdots \xrightarrow{\Sigma^{-1}\circ\Sigma\mu(\beta)^*} \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, \Sigma R) \xrightarrow{\Sigma^{-1}\circ\Sigma\mu(\beta)^*} \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)\}. \end{aligned}$$

Again, since R is periodic via the multiplication by α map $\mu(\alpha) : R \to \Sigma^{-1} R$, we may compose with $\mu(\alpha)^{-1}$ in order to rewrite this as

$$\mathrm{map}(L, R) \simeq \mathrm{holim}_n\{\cdots \xrightarrow{g} \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R) \xrightarrow{g} \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)\},$$

where g is the endomorphism of $\mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)$ which sends the map $y : \Sigma^\infty \mathbb{P}^\infty_+ \to R$ to the map $g(y) = \mu(\alpha)^{-1} \circ \Sigma^{-1}(y) \circ \mu(\beta)$. □

Corollary 4.8. *Let R be a homotopy commutative $PMGL$-algebra. Then the square*

$$\begin{tikzcd} \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \arrow[r, "f"] \arrow[d, "i^*"] & \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \arrow[d, "i^*"] \\ \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R) \arrow[r, "g"] & \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)\ , \end{tikzcd}$$

in which the vertical map are induced by the inclusion $i : \mathbb{P}^\infty_+ \simeq BGL_0 + BGL_1 \to BGL_{\mathbb{Z}}$, commutes up to homotopy. In particular, $\psi^* : \mathrm{map}(K, R) \to$

map(L,R) *is the homotopy inverse limit of the map of towers*

$$\begin{tikzcd}
\mathrm{map}(K,R) \arrow[r, "\simeq"] \arrow[d, "\psi^*"] & \mathrm{holim}\{\cdots \arrow[r, "f"] & \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}},R) \arrow[r, "f"] \arrow[d, "i^*"] & \mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}},R)\} \arrow[d, "i^*"] \\
\mathrm{map}(L,R) \arrow[r, "\simeq"] & \mathrm{holim}\{\cdots \arrow[r, "g"] & \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+,R) \arrow[r, "g"] & \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+,R)\}
\end{tikzcd}$$

obtained from iterating this commuting square.

Proof. This is immediate from Lemmas 4.6 and 4.7. □

4.4. A useful splitting. To complete the analysis of $\psi^* : \mathrm{map}(K,R) \to \mathrm{map}(L,R)$, we must split the space of additive maps from $BGL_{\mathbb{Z}}$ to $\Omega^\infty R$ off of the space of all maps from $BGL_{\mathbb{Z}}$ to $\Omega^\infty R$.

Proposition 4.9. *Let R be a motivic spectrum equipped with an equivalence $\mu(\alpha) : R \to \Sigma^{-1}R$. Then given a map $x : \Omega^\infty K \to \Omega^\infty R$, the map*

$$\Omega^\infty\mu(\alpha)^{-1} \circ \Omega(x) \circ \Omega^\infty\mu(\beta) : \Omega^\infty K \longrightarrow \Omega^{\infty+1}K \longrightarrow \Omega^{\infty+1}R \longrightarrow \Omega^\infty R$$

is a homomorphism for the additive structures on $\Omega^\infty K$ and $\Omega^\infty R$.

Proof. If X is a motivic space equipped with an equivalence $X \to \Omega Y$, then X is a group object in the homotopy category of motivic spaces; if in addition $Y \simeq \Omega Z$, then X is an abelian group object. In particular, the additions on $\Omega^\infty K$ and $\Omega^\infty R$ are induced by the equivalences $\Omega^\infty\mu(\beta) : \Omega^\infty K \to \Omega^{\infty+1}K$ and $\Omega^\infty\mu(\alpha) : \Omega^\infty R \to \Omega^{\infty+1}R$, respectively, and $\Omega^\infty\mu(\alpha)^{-1} \circ \Omega(x) \circ \Omega^\infty\mu(\beta)$ is a map of loop spaces and therefore respects this addition. □

Proposition 4.10. *Let R be a homotopy commutative $PMGL$-algebra. Then there exists a canonical section $s : R^{\mathbb{P}^\infty_+} \to R^{BGL_{\mathbb{Z}}}$ of the restriction $r = i^* : R^{BGL_{\mathbb{Z}}} \to R^{\mathbb{P}^\infty_+}$ induced by the inclusion $i : \mathbb{P}^\infty_+ \simeq BGL_0 + BGL_1 \to BGL_{\mathbb{Z}}$.*

Proof. Set $Y = R^{BGL_{\mathbb{Z}}}$ and $Z = R^{\mathbb{P}^\infty_+}$. By Proposition 4.11, the additive maps from $BGL_{\mathbb{Z}}$ to $\Omega^\infty R$ define a canonical section $\pi_0 Z \to \pi_0 Y$ of the surjection $\pi_0 Y \to \pi_0 Z$. We must lift this to a map of spectra $s : Z \to Y$.
By Proposition 2.24, we have isomorphisms $R^0(\mathbb{P}^\infty_+ \otimes Z) \cong R^0(Z) \otimes_{\pi_0 R} \pi_0 Z \cong Z^0(Z)$. Combined with the section $\pi_0 Y \to \pi_0 Z$, this induces a map

$$Z^0(Z) \to R^0(Z) \otimes_{\pi_0 R} \pi_0 Z \to R^0(Z) \otimes_{\pi_0 R} \pi_0 Y \to Y^0(Z).$$

Take $s \in Y^0(Z)$ to be the image of $1 \in Z^0(Z)$ under this map. □

Corollary 4.11. *Let R be a homotopy commutative $PMGL$-algebra. Then*

$$\mathrm{map}(\Sigma^\infty BGL_{\mathbb{Z}},R) \simeq \mathrm{map}(\Sigma^\infty \mathbb{P}^\infty_+,R) \times X$$

for some space X.

Proof. Take X to be the global points of the motivic space obtained by applying Ω^∞ to the fiber of $r : R^{BGL_{\mathbb{Z}}} \to R^{\mathbb{P}^\infty_+}$. $\square$

We write

$$r : \operatorname{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \longrightarrow \operatorname{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)$$

for the restriction and

$$s : \operatorname{map}(\Sigma^\infty \mathbb{P}^\infty_+, R) \longrightarrow \operatorname{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R)$$

for the section corresponding to the inclusion of the additive maps from $BGL_{\mathbb{Z}}$ to $\Omega^\infty R$ into all maps from $BGL_{\mathbb{Z}}$ to $\Omega^\infty R$. By Corollary 4.8, we have that $r \circ f = g \circ r$; however, the next proposition shows that in fact $f \simeq s \circ g \circ r$, which is stronger since $r \circ f \simeq r \circ s \circ g \circ r \simeq g \circ r$ as s is a section of r.

Proposition 4.12. *Let R be a homotopy commutative $PMGL$-algebra. Then*

$$f \simeq s \circ g \circ r : \operatorname{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R) \longrightarrow \operatorname{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R).$$

Proof. Note that f and g are induced from corresponding maps $f : R^{BGL_{\mathbb{Z}}} \to R^{BGL_{\mathbb{Z}}}$ and $g : R^{\mathbb{P}^\infty_+} \to R^{\mathbb{P}^\infty_+}$, respectively, and that r and s come similarly from maps $r : R^{BGL_{\mathbb{Z}}} \to R^{\mathbb{P}^\infty_+}$ and $s : R^{\mathbb{P}^\infty_+} \to R^{BGL_{\mathbb{Z}}}$. Thus it's enough to check that

$$f = s \circ g \circ r \in \pi_0 \operatorname{map}(R^{BGL_{\mathbb{Z}}}, R^{BGL_{\mathbb{Z}}}).$$

Since $r \circ f \simeq g \circ r$, we may instead show that $f \simeq s \circ r \circ f$. By definition,

$$f \in \pi_0 \operatorname{map}(R^{BGL_{\mathbb{Z}}}, R^{BGL_{\mathbb{Z}}}) \cong R^0(BGL_{\mathbb{Z}}) \widehat{\underset{R^0}{\otimes}} R^0(R^{BGL_{\mathbb{Z}}})$$

is the image of

$$1 \in \pi_0 \operatorname{map}(R^{BGL_{\mathbb{Z}}}, R^{BGL_{\mathbb{Z}}}) \cong R^0(BGL_{\mathbb{Z}}) \widehat{\underset{R^0}{\otimes}} R^0(R^{BGL_{\mathbb{Z}}})$$

under the map obtained by applying $(-)\widehat{\otimes}_{R^0} R^0(R^{BGL_{\mathbb{Z}}})$ to the map

$$\Omega^\infty \mu(\alpha)_*^{-1} \Omega^\infty \mu(\beta)^* \Omega : \pi_0 \operatorname{map}(BGL_{\mathbb{Z}}, \Omega^\infty R) \longrightarrow \pi_0 \operatorname{map}(BGL_{\mathbb{Z}}, \Omega^\infty R).$$

One similarly checks that r and s are obtained by applying $(-)\widehat{\otimes}_{R_0} R^0(R^{BGL_{\mathbb{Z}}})$ to the restriction $R^0(BGL_{\mathbb{Z}}) \to R^0(\mathbb{P}^\infty)$ and its section $R^0(\mathbb{P}^\infty) \to R^0(BGL_{\mathbb{Z}})$, respectively. Now, according to Proposition 4.9, as a map of motivic loop spaces, $\Omega^\infty \mu(\alpha)_*^{-1} \Omega^\infty \mu(\beta)^* \Omega$ sends $x : BGL_{\mathbb{Z}} \to \Omega^\infty R$ to the additive map

$$\Omega^\infty \mu(\alpha)^{-1} \circ \Omega(x) \circ \Omega^\infty \mu(\beta) : BGL_{\mathbb{Z}} \longrightarrow \Omega BGL_{\mathbb{Z}} \longrightarrow \Omega^{\infty+1} R \longrightarrow \Omega^\infty R,$$

which is to say that it factors through the inclusion $\operatorname{Add}(BGL_{\mathbb{Z}}, \Omega^\infty R) \cong R^0(\mathbb{P}^\infty_+) \to R^0(BGL_{\mathbb{Z}})$. Hence f is in the image of

$$\begin{tikzcd}
R^0(\mathbb{P}^\infty_+)\widehat{\otimes}_{R^0} R^0(R^{BGL_{\mathbb{Z}}}) \arrow[r, "s"] \arrow[d, "\cong"] & R^0(BGL_{\mathbb{Z}})\widehat{\otimes}_{R^0} R^0(R^{BGL_{\mathbb{Z}}}) \arrow[d, "\cong"] \\
\pi_0 \operatorname{map}(R^{BGL_{\mathbb{Z}}}, R^{\mathbb{P}^\infty_+}) \arrow[r, "s"] & \pi_0 \operatorname{map}(R^{BGL_{\mathbb{Z}}}, R^{BGL_{\mathbb{Z}}}) ,
\end{tikzcd}$$

so $f = s \circ \tilde{f}$ for some $\tilde{f} : R^{BGL_{\mathbb{Z}}} \to R^{\mathbb{P}^\infty_+}$. But then $s \circ r \circ f \simeq s \circ r \circ s \circ \tilde{f} \simeq s \circ \tilde{f} \simeq f$. $\square$

Lemma 4.13. *Suppose given a homotopy commutative diagram of (ordinary) spectra*

$$\begin{tikzcd} Y \arrow[r, "f"] \arrow[d, "r"'] & Y \arrow[d, "r"] \\ Z \arrow[r, "g"] & Z \end{tikzcd}$$

such that $r : Y \to Z$ admits a section $s : Z \to Y$ with

$$f \simeq s \circ g \circ r : Y \to Y.$$

Then the natural map from the homotopy limit of the tower $\{\cdots \to Y \to Y\}$, obtained by iterating f, to the homotopy limit of the tower $\{\cdots \to Z \to Z\}$, obtained by iterating g, is an equivalence.

Proof. Since Z is a retract of Y, we may write $Y \simeq Z \times X$ for some spectrum X such that the fiber of f over g is the trivial map $X \to X$. Now consider the diagram

$$\begin{tikzcd} W \arrow[r] \arrow[d] & \prod_n X \arrow[r] \arrow[d] & \prod_n X \arrow[d] \\ \operatorname{holim}\{\cdots \to Y \to Y\} \arrow[r] \arrow[d] & \prod_n Y \arrow[r, "1-f"] \arrow[d] & \prod_n Y \arrow[d] \\ \operatorname{holim}\{\cdots \to Z \to Z\} \arrow[r] & \prod_n Z \arrow[r, "1-g"] & \prod_n Z \end{tikzcd}$$

in which the rows and columns are fiber sequences. Then the fiber of $1 - f$ over $1 - g$ is the identity $\prod_n X \to \prod_n X$, so W is trivial and $\operatorname{holim}\{\cdots \to Y \to Y\} \simeq \operatorname{holim}\{\cdots \to Z \to Z\}$. $\square$

Proposition 4.14. *Let R be an orientable commutative motivic ring spectrum and let $\alpha \in \pi_1 R$ be a unit. Then*

$$\psi^* : \operatorname{map}(K, R) \longrightarrow \operatorname{map}(L, R)$$

is an equivalence.

Proof. Set $Y = \operatorname{map}(\Sigma^\infty BGL_{\mathbb{Z}}, R)$ and $Z = \operatorname{map}(\Sigma^\infty \mathbb{P}^\infty_+, R)$, so that $Y \simeq Z \times X$ via $r : Y \to Z$ and its section $s : Z \to Y$. By Proposition 4.12, $f \simeq s \circ g \circ r : Y \to Y$, and since r and s are infinite loop maps we may regard them as maps of (ordinary) connective spectra. The result then follows from Lemma 4.13. $\square$

Corollary 4.15. *The map $\psi : L \to K$ induces an isomorphism $\psi^* : R^0(K) \to R^0(L)$ for any orientable periodic commutative motivic ring spectrum R.*

Proof. This is immediate from Proposition 4.14 above, since the spectrum of motivic spectrum maps from L to R admits precisely the same description as that of the spectrum of motivic spectrum maps from K to R. Indeed,

$$L \simeq \mathrm{colim}_n\{\Sigma^\infty_+\mathbb{P}^\infty \xrightarrow{\beta_*} \Sigma^{-1}\Sigma^\infty_+\mathbb{P}^\infty \xrightarrow{\beta_*} \cdots\},$$

and we see that

$$R^L \simeq \mathrm{holim}_n\{\cdots \xrightarrow{g} R^{\mathbb{P}^\infty_+} \xrightarrow{g} R^{\mathbb{P}^\infty_+}\},$$

where the map $g : R^{\mathbb{P}^\infty_+} \to R^{\mathbb{P}^\infty_+}$ sends $\lambda : \Sigma^\infty_+\mathbb{P}^\infty \to R$ to $\alpha_*^{-1} \circ \Sigma^{-1}\lambda \circ \beta_*$, the composite

$$\Sigma^\infty_+\mathbb{P}^\infty \to \Sigma^{-1}\Sigma^\infty_+\mathbb{P}^\infty \to \Sigma^{-1}R \to R,$$

just as above. $\square$

By the homotopy category of orientable periodic spectra, we mean the full subcategory of the homotopy category of spectra on the orientable and periodic objects. In other words, R is an orientable periodic spectrum if there exists a homotopy commutative ring structure on R which admits a ring map $PMGL \to R$. Note that, according to this definition, maps between orientable periodic spectra need not preserve potential orientations or even ring structures.

Proposition 4.16. *ψ induces an isomorphism $\psi^* : [K, -] \to [L, -]$ of functors from the homotopy category of orientable periodic spectra to abelian groups.*

Proof. Let R be an orientable periodic spectrum. Then

$$\psi^* : [K, R] = R^0(K) \to R^0(L) = [L, R]$$

is an isomorphism by Corollary 4.15, and this isomorphism is natural in spectrum maps $R \to R'$, provided of course that R' is also orientable and periodic. $\square$

Theorem 4.17. *The ring map $\psi : L \to K$ is an equivalence.*

Proof. Let $\varphi^* : [L, -] \to [K, -]$ be the inverse of the isomorphism $\psi^* : [K, -] \to [L, -]$ of Proposition 4.16, and let $\varphi : K \to L$ be the map obtained by applying φ^* to the identity $1 \in [L, L]$. It follows from the Yoneda lemma that φ^* is precomposition with φ. The equations $\varphi^* \circ \psi^* = 1_K^*$ and $\psi^* \circ \varphi^* = 1_L^*$ imply that $\psi \circ \varphi = 1_K$ and $\varphi \circ \psi = 1_L$ in the homotopy category of orientable periodic spectra, and therefore that $\psi \circ \varphi = 1_K$ and $\varphi \circ \psi = 1_L$ in the homotopy category of spectra. Hence $\psi : L \to K$ is an equivalence with inverse $\varphi : K \to L$. $\square$

5. Applications

5.1. The motivic Conner-Floyd theorem. The classical theorem of Conner and Floyd shows that complex cobordism determines complex K-theory by base change. More precisely, writing PMU for periodic complex cobordism and KU for complex K-theory, then, for any finite spectrum X, the natural map

$$PMU^0(X) \otimes_{PMU^0} KU^0 \to KU^0(X)$$

is an isomorphism of KU^0-modules.

Remark 5.1. This is the precursor of the more general notion of Landweber exactness. In [8], P. Landweber gives a necessary and sufficient condition on an MU_*-module G so that the functor $(-) \otimes_{MU_*} G$, from (MU_*, MU_*MU)-comodules to graded abelian groups, is exact. For $G = K_*$, it follows that the natural map

$$MU^*(-) \otimes_{MU^*} K^* \to K^*(-)$$

is an isomorphism. See [21] for the motivic analogue of Landweber exactness.

We now turn to the motivic version of the theorem of Conner and Floyd. A motivic spectrum X is said to be *compact* if $[X, -]$, viewed as a functor from motivic spectra to abelian groups, commutes with filtered colimits.

Proposition 5.2. *Let X be a compact motivic spectrum. Then the natural map*

$$PMGL^0(X) \underset{PMGL^0}{\otimes} K^0 \longrightarrow K^0(X)$$

is surjective.

Proof. Set $B := \Sigma^\infty_+ BGL$ and $A := \Sigma^\infty_+ \mathbb{P}^\infty$. Then the determinant map $r : B \to A$ admits a section $s : A \to B$, so, for each n, $\Sigma^{-n}A$ is a retract of $\Sigma^{-n}B$ and $\Sigma^{-n}B^0(X) \to \Sigma^{-n}A^0(X)$ is surjective. Since X is compact, the colimit

$$B[1/\beta]^0(X) \cong \operatorname{colim}_n \Sigma^{-n}B^0(X) \longrightarrow \operatorname{colim}_n \Sigma^{-n}A^0(X) \cong A[1/\beta]^0(X)$$

is also surjective, and we see that

$$B[1/\beta]^0(X) \underset{B[1/\beta]^0}{\otimes} A[1/\beta]^0 \longrightarrow A[1/\beta]^0(X) \underset{B[1/\beta]^0}{\otimes} A[1/\beta]^0 \cong A[1/\beta]^0(X)$$

is surjective as well. $\square$

Theorem 5.3. *Let X be a compact motivic spectrum. Then the natural map*

$$PMGL^0(X) \underset{PMGL^0}{\otimes} K^0 \longrightarrow K^0(X)$$

is an isomorphism.

Proof. According to Proposition 5.2, it's enough to show that the map is injective. For simplicity of notation, set $R := PMGL$, define a contravariant functor $J^0(-)$ from compact motivic spectra to R^0-modules by the rule

$$J^0(X) := \ker\{R^0(X) \to K^0(X)\},$$

and write J^0 for $J^0(\mathbb{S})$. Since the tensor product is right exact, the map

$$J^0(X) \otimes_{R^0} K^0 \longrightarrow \ker\{R^0(X) \otimes_{R^0} K^0 \to K^0(X) \otimes_{R^0} K^0\}$$

is surjective, so in light of the isomorphism

$$K^0(X) \otimes_{R^0} K^0 \cong K^0(X) \otimes_{K^0} K^0 \cong K^0(X)$$

it's enough to show that $J^0(X) \otimes_{R^0} K^0$ is zero, or, equivalently, that

$$J^0(X) \otimes_{R^0} J^0 \longrightarrow J^0(X) \otimes_{R^0} R^0 \cong J^0(X)$$

is surjective. To this end, set

$$I^0(X) := \operatorname{im}\{J^0(X) \otimes_{R^0} J^0 \to J^0(X) \otimes_{R^0} R^0 \cong J^0(X)\};$$

we must show that $I^0(X) \cong J^0(X)$.
Now, writing $B := \Sigma^\infty_+ BGL$ and $A := \Sigma^\infty_+ \mathbb{P}^\infty$ as above, and using the compactness of X, we see that any element of

$$\begin{aligned} J^0(X) &\cong \ker\{\operatorname{colim}_n [X, \Sigma^{-n} B] \to \operatorname{colim}_n [X, \Sigma^{-n} A]\} \\ &\cong \operatorname{colim}_n \ker\{[X, \Sigma^{-n} B] \to [X, \Sigma^{-n} A]\} \end{aligned}$$

is represented by a map

$$x : \Sigma^n X \to B \simeq \operatorname{colim}_p \operatorname{colim}_q \Sigma^\infty_+ \operatorname{Grass}_{p,q},$$

which, by compactness, factors as $f_x : \Sigma^n X \to Y_x$ followed by $y : Y_x \to B$ for $Y_x \simeq \Sigma^\infty_+ \operatorname{Grass}_{p,q}$ the suspension spectrum of a finite Grassmannian. This yields a commuting diagram

$$\begin{tikzcd} \Sigma^n X \arrow[r, "x"] \arrow[d, "f_x"'] & B \arrow[d, "r"] \\ Y_x \arrow[ur, "y"] \arrow[r] & A \end{tikzcd}$$

in which $r \circ x$ is trivial and the determinant map $r : B \to A$ admits a section $s : A \to B$. Of course, as $r \circ y$ need not be trivial, set $y' := y - s \circ r \circ y$, so that

$$y' \circ f \simeq (y - s \circ r \circ y) \circ f \simeq y \circ f - s \circ r \circ y \circ f \simeq x - s \circ r \circ x \simeq x$$

and $r \circ y' \simeq 0$, which is to say that $y' \in J^0(X)$ and $f_x^* y' = x$.
Finally, according to Proposition 2.19, $R^0(Y_x) \otimes_{R^0} K^0 \cong K^0(Y_x)$ for each $x \in J^0(X)$, so we must have surjections

$$J^0(Y_x) \otimes_{R^0} J^0 \longrightarrow J^0(Y_x) \otimes_{R^0} R^0 \cong J^0(Y_x).$$

Adding these together, we obtain a morphism of short exact sequences

$$\begin{tikzcd}
0 \arrow[r] & \bigoplus_x I^0(Y_x) \arrow[r] \arrow[d] & \bigoplus_x J^0(Y_x) \arrow[r] \arrow[d, "\bigoplus_x f_x^*"] & 0 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & I^0(X) \arrow[r] & J^0(X) \arrow[r] & J^0(X) \otimes_{R^0} K^0 \arrow[r] & 0
\end{tikzcd}$$

such that $\bigoplus_x f_x^* : \bigoplus_x J^0(Y_x) \to J^0(X)$ is surjective. It follows that $I^0(X) \cong J^0(X)$. □

5.2. *PMGL* and *K* are E_∞ motivic spectra. As a final application, we show that $PMGL$ and K are E_∞ motivic spectra. As we shall see, this is an immediate consequence of the fact that $PMGL$ and K are obtained through a localization of the category of E_∞ R-algebras for some E_∞ motivic spectrum R. Roughly, given an element $\beta \in \pi_{p,q}R$, the functor which sends the R-module M to $M[1/\beta] := M \wedge_R R[1/\beta]$ defines a monoidal localization of the category of R-modules, so it extends to a localization of the category of E_∞ R-algebras. Taking $R = \Sigma_+^\infty BGL$ and β the Bott element, we see that $PMGL$ is the localization of the initial E_∞ R-algebra and K is the localization of the determinant E_∞ R-algebra.

In order to make this precise, we fix a suitable symmetric monoidal model category $(\mathrm{Mod}_{\mathbb{S}}, \wedge_{\mathbb{S}})$ of motivic spectra, such as motivic $\mathbb{S}$-modules [6] or motivic symmetric spectra [7]. For sake of definiteness, we adopt the formalism of the latter; nevertheless, we refer to motivic symmetric spectra as $\mathbb{S}$-modules, as they are indeed modules over the symmetric motivic sphere $\mathbb{S}$.

Recall that a *motivic symmetric sequence* is a functor from the groupoid Σ of finite sets and isomorphisms to *pointed* motivic spaces. It is sometimes convenient to use a skeleton of Σ, so we simply write n for a finite set with n elements and $\Sigma(n)$ for its automorphism group. Motivic symmetric sequences form a symmetric monoidal category under the smash product defined by

$$(X \wedge Y)(n) := \bigvee_{n=p+q} \Sigma(n)_+ \wedge_{\Sigma(p)\times\Sigma(q)} X(p) \wedge Y(q).$$

The motivic sphere $\mathbb{S}$ has a natural interpretation as the motivic symmetric sequence in which $\mathbb{S}(n)$ is the pointed $\Sigma(n)$-space associated to the pair $(\mathbb{A}^n, \mathbb{A}^n - \mathbb{A}^0)$, where $\Sigma(n)$ acts by permutation of coordinates. The $\Sigma(p)\times\Sigma(q)$-equivariant maps

$$(\mathbb{A}^p, \mathbb{A}^p - \mathbb{A}^0) \wedge (\mathbb{A}^q, \mathbb{A}^q - \mathbb{A}^0) \longrightarrow (\mathbb{A}^{p+q}, \mathbb{A}^{p+q} - \mathbb{A}^0)$$

give $\mathbb{S}$ the structure of a commutative monoid for this smash product. An $\mathbb{S}$-*module* is then a motivic symmetric sequence equipped with an action of $\mathbb{S}$, which is to say a sequence $X(p)$ of pointed $\Sigma(p)$-equivariant motivic spaces equipped with $\Sigma(p) \times \Sigma(q)$-equivariant maps

$$(\mathbb{A}^p, \mathbb{A}^p - \mathbb{A}^0) \wedge X(q) \longrightarrow X(p+q);$$

the fact that $\mathbb{S}$ is a commutative monoid implies that $\wedge$ extends to a smash product $\wedge_{\mathbb{S}}$ on the category $\mathrm{Mod}_{\mathbb{S}}$ of $\mathbb{S}$-modules. There are monoidal functors

$$\{\text{Motivic spaces}\} \to \{\text{Motivic symmetric spaces}\} \to \{\text{Motivic symmetric spectra}\}$$

in which the righthand map is the free $\mathbb{S}$-module functor, the left hand map sends the motivic space X to the constant motivic symmetric space X_+, and the composite is a structured version of the suspension spectrum functor Σ^∞_+.

Proposition 5.4. *The $\mathbb{S}$-modules $\Sigma^\infty_+ BGL$ and $\Sigma^\infty_+ \mathbb{P}^\infty$ are equivalent to strictly commutative $\mathbb{S}$-algebras in such a way that the determinant map $\Sigma^\infty_+ BGL \to \Sigma^\infty_+ \mathbb{P}^\infty$ is equivalent to a map of strictly commutative $\mathbb{S}$-algebras.*

Proof. For each n, write $GL(n)$ for the group S-scheme of linear automorphisms of $\mathbb{A}^n$. Then $\Sigma(n)$ acts on $GL(n)$ by conjugation via the embedding $\Sigma(n) \to GL(n)$, so that $GL(n)$ is the value at n of a symmetric sequence GL in group S-schemes such that the determinant map $GL \to GL_1$ is a morphism of symmetric sequences in group S-schemes, where we regard GL_1 as a constant symmetric sequence. Taking classifying spaces, we obtain a morphism of commutative monoid symmetric sequences $BGL \to BGL_1$ in unpointed motivic spaces. Now let $\mathbb{S}[BGL]$ and $\mathbb{S}[BGL_1]$ denote the $\mathbb{S}$-modules defined by

$$\mathbb{S}[BGL](n) := \mathbb{S}(n) \wedge BGL(n)_+ \qquad \text{and} \qquad \mathbb{S}[BGL_1](n) := \mathbb{S}(n) \wedge BGL_{1+},$$

respectively, where $\Sigma(n)$ acts diagonally; note that $\mathbb{S}[BGL_1]$ is the free $\mathbb{S}$-module on the motivic symmetric sequence BGL_{1+}, whereas the action of $\mathbb{S}$ on $\mathbb{S}[BGL]$ is induced by the canonical $\Sigma(p) \times \Sigma(q)$-equivariant inclusions

$$BGL(q) \longrightarrow BGL(p) \times BGL(q) \to BGL(p+q)$$

coming from the fact that $BGL(p)$ has a canonical basepoint which is fixed by the action of $\Sigma(p)$. The monoidal structure on BGL_1 extends to a strictly commutative $\mathbb{S}$-algebra structure on $\mathbb{S}[BGL_1]$, and the strictly commutative $\mathbb{S}$-algebra structure on $\mathbb{S}[BGL]$ comes from $\Sigma(p) \times \Sigma(q)$-equivariant pairing

$$\mathbb{S}(p) \wedge BGL(p)_+ \wedge \mathbb{S}(q) \wedge BGL(q)_+ \longrightarrow \mathbb{S}(p+q) \wedge BGL(p+q)_+;$$

moreover, it is clear that the determinant map $\mathbb{S}[BGL] \to \mathbb{S}[BGL_1]$ is monoidal with respect to these multiplicative structures. Hence we are done, provided the underlying ordinary motivic spectra (obtained by forgetting the actions of the symmetric groups) of $\mathbb{S}[BGL]$ and $\mathbb{S}[BGL_1]$ are equivalent to the motivic spectra $\Sigma^\infty_+ BGL$ and $\Sigma^\infty_+ \mathbb{P}^\infty$, respectively. This is immediate for $\mathbb{S}[BGL_1]$, whose underlying spectrum is the suspension spectrum $\Sigma^\infty_+ BGL_1$; on the other hand, the underlying spectrum of $\mathbb{S}[BGL]$ is the prespectrum $\{S^n \wedge BGL_{n+}\}$. But the motivic spectrum associated to $\Sigma^\infty_+ BGL$ is given by

$$\mathrm{colim}_p\, \Omega^p \Sigma^p\, \mathrm{colim}_q\, BGL_q \simeq \mathrm{colim}_p\, \mathrm{colim}_q\, \Omega^p \Sigma^p BGL_q \simeq \mathrm{colim}_n\, \Omega^n \Sigma^n BGL_n,$$

so the two motivic prespectra are stably equivalent. $\square$

Instead of considering localization in the context of symmetric monoidal model categories, it will be enough to consider localization in the context of symmetric

monoidal ∞-categories, in the sense of Lurie [16]. Indeed, if $(\mathscr{M}, \otimes)$ is a symmetric monoidal model category, then, in the notation of [16], the commutative algebra objects of the associated symmetric monoidal ∞-category $\mathrm{N}(\mathscr{M}, \otimes)^\circ$ correspond to *coherently homotopy commutative*, or E_∞, objects of $(\mathscr{M}, \otimes)$. Here N denotes the simplicial nerve of a simplicial category; if the simplicial category comes from a symmetric monoidal model category, then its simplicial nerve is symmetric monoidal as an ∞-category. See [17] for facts about ∞-categories and simplicial nerves, and [16] for a treatment of commutative algebra in the ∞-categorical context.
Recall (cf. [17]) that a map $F : \mathscr{C} \to \mathscr{D}$ of ∞-categories is said to be a *localization* if F admits a fully faithful right adjoint G. In this case, it is common to identify $\mathscr{D}$ with the full subcategory of $\mathscr{C}$ consisting of those objects in the essential image of G (the "local objects"), and suppress $\mathscr{D}$ and G from the notation by writing L for the composite $G \circ F : \mathscr{C} \to \mathscr{D} \to \mathscr{C}$. If $\mathscr{C}$ is the underlying ∞-category of a symmetric monoidal ∞-category $(\mathscr{C}, \otimes)$, then we may ask when a localization $L : \mathscr{C} \to \mathscr{C}$ extends to a lax symmetric monoidal functor on $(\mathscr{C}, \otimes)$. Given a localization $L : \mathscr{C} \to \mathscr{C}$, an L-equivalence is a map which becomes an equivalence after applying L.

DEFINITION 5.5 ([16], Example 1.7.5). Let $(\mathscr{C}, \otimes)$ be a symmetric monoidal ∞-category and let $L : \mathscr{C} \to \mathscr{C}$ be a localization of the underlying ∞-category. Then L is said to be *compatible with* $\otimes$ if, for all L-equivalences $A \to A'$ and all objects B of $\mathscr{C}$, $A \otimes B \to A' \otimes B$ is an L-equivalence.

PROPOSITION 5.6 ([16], Proposition 1.7.6). *Let $(\mathscr{C}, \otimes)$ be a symmetric monoidal ∞-category, let $L : \mathscr{C} \to \mathscr{C}$ be a localization of the underlying ∞-category, and suppose that L is compatible with $\otimes$. Then L extends to a lax symmetric monoidal functor*

$$(L, \otimes) : (\mathscr{C}, \otimes) \to (\mathscr{C}, \otimes).$$

In particular, L preserves algebra and commutative algebra objects of $(\mathscr{C}, \otimes)$.

Let $\mathrm{N}(\mathrm{Mod}_{\mathbb{S}}, \wedge_{\mathbb{S}})^\circ$ denote the symmetric monoidal ∞-category which arises as the simplicial nerve of the symmetric monoidal simplicial model category $(\mathrm{Mod}_{\mathbb{S}}, \wedge_{\mathbb{S}})$ of $\mathbb{S}$-modules. Since commutative algebra objects of $\mathrm{N}(\mathrm{Mod}_{\mathbb{S}}, \wedge_{\mathbb{S}})^\circ$ are modeled by algebras over a suitable E_∞ operad, we refer to commutative algebra objects of $\mathrm{N}(\mathrm{Mod}_{\mathbb{S}}, \wedge_{\mathbb{S}})^\circ$ as E_∞ $\mathbb{S}$-*algebras*. Given an E_∞ $\mathbb{S}$-algebra R, we write $(\mathrm{Mod}_R, \wedge_R)$ for the resulting symmetric monoidal ∞-category of R-modules, and refer to commutative algebra objects of $(\mathrm{Mod}_R, \wedge_R)$ as E_∞ R-*algebras.*

PROPOSITION 5.7. *Let R be an E_∞ $\mathbb{S}$-algebra, let $f \in \pi_{p,q}R$ be an arbitrary element, and write $L_f : \mathrm{Mod}_R \to \mathrm{Mod}_R$ for the functor which sends the R-module M to the R-module*

$$M[1/f] := M \wedge_R R[1/f].$$

Then L_f is a localization functor which is compatible with the symmetric monoidal structure $\wedge_R$ on Mod_R; in particular, L_f extends to a lax monoidal functor $L_f : (\mathrm{Mod}_R, \wedge_R) \to (\mathrm{Mod}_R, \wedge_R)$.

Proof. Say that an R-module M is f-local if the multiplication by f map $f_* : M \to \Sigma^{-p,-q} M$ is an equivalence. Given an f-local R-module M, the induced map

$$\mathrm{Map}(R[1/f], M) \simeq \lim\{M \leftarrow \Sigma^{p,q} M \leftarrow \cdots\} \simeq M$$

is an equivalence, so

$$\mathrm{map}(N[1/f], M) \simeq \mathrm{map}(N, \mathrm{Map}(R[1/f], M)) \simeq \mathrm{map}(N, M)$$

is an equivalence for any R-module N. Hence L_f is left adjoint to the inclusion of the full subcategory of f-local R-modules, and is therefore a localization. Moreover, it is compatible with $\wedge_R$, since if $M \to M'$ is an L_f-equivalence then so is $M \wedge_R N \to M' \wedge_R N$ for any R-module N, for if $M[1/f] \to M'[1/f]$ is an equivalence then

$$(M \wedge_R N)[1/f] \simeq M[1/f] \wedge_R N \longrightarrow M'[1/f] \wedge_R N \simeq (M' \wedge_R N)[1/f]$$

is as well. Hence, by Proposition 5.6, L_f extends to a lax symmetric monoidal endofunctor $(L_f, \wedge_R)$ of $(\mathrm{Mod}_R, \wedge_R)$. □

Corollary 5.8. *Let R be an E_∞ $\mathbb{S}$-algebra and let $f \in \pi_{p,q}$ be a fixed element. Then $R[1/f]$ is an E_∞ R-algebra, and therefore also an E_∞ $\mathbb{S}$-algebra.*

Proof. By Proposition 5.7, L_f is a lax symmetric monoidal functor with $L_f R \simeq R[1/f]$. Since lax symmetric monoidal functors preserve commutative algebra objects, we see that $R[1/f]$ is a commutative algebra object in $(\mathrm{Mod}_R, \wedge_R)$. Lastly, as the forgetful functor from R-modules to $\mathbb{S}$-modules is lax symmetric monoidal, it follows that $R[1/f]$ is also an E_∞ $\mathbb{S}$-algebra. □

Corollary 5.9. *MGL, $PMGL$ and K are E_∞ $\mathbb{S}$-algebras.*

Proof. The MGL case is already well-known (cf. [6], for example). By Proposition 5.4, $\Sigma^\infty_+ BGL$ and $\Sigma^\infty_+ \mathbb{P}^\infty$ are equivalent to strictly commutative $\mathbb{S}$-algebras, so they are naturally commutative algebra objects in the symmetric monoidal ∞-category $\mathrm{N}(\mathrm{Mod}_{\mathbb{S}}, \wedge_{\mathbb{S}})$. Applying Proposition 5.8, we see that $PMGL \simeq \Sigma^\infty_+ BGL[1/\beta]$ is an E_∞ $\Sigma^\infty_+ BGL$-algebra, and likewise that $K \simeq \Sigma^\infty_+ \mathbb{P}^\infty[1/\beta]$ is an E_∞ $\Sigma^\infty_+ \mathbb{P}^\infty$-algebra. In particular, $PMGL$ and K are E_∞ $\mathbb{S}$-algebras. □

Proposition 5.10. *K is an E_∞ $PMGL$-algebra.*

Proof. Note that the Bott element $\Sigma\mathbb{S} \to \Sigma^\infty_+ \mathbb{P}^\infty$ factors as the composite of the Bott element $\Sigma\mathbb{S} \to \Sigma^\infty_+ BGL$ followed by the determinant map $\Sigma^\infty_+ BGL \to \Sigma^\infty_+ \mathbb{P}^\infty$. By Proposition 5.4, the determinant map $\Sigma^\infty_+ BGL \to \Sigma^\infty_+ \mathbb{P}^\infty$ is a map of E_∞ $\Sigma^\infty_+ BGL$-algebras, so by Proposition 5.8, the localization $K \simeq \Sigma^\infty_+ \mathbb{P}^\infty[1/\beta]$ is an E_∞ algebra over $PMGL \simeq \Sigma^\infty_+ BGL[1/\beta]$. □

References

[1] J. F. Adams. *Stable Homotopy and Generalised Homology.* University of Chicago Press (1974).

[2] J. F. Adams. Primitive elements in the K-theory of BSU. Quart. J. Math. Oxford (2) 27 (1976) no. 106, 253-262.

[3] R. D. Arthan. Localization of stable homotopy rings. Math. Proc. Camb. Phil. Soc. 93 (1983) 295-302.

[4] M. F. Atiyah. *K-Theory.* Benjamin (1968).

[5] D. Dugger and D. Isaksen. Motivic cell structures. Algebraic and Geometric Topology, Vol. 5 (2005) 615–652.

[6] P. Hu. S-modules in the category of schemes. Mem. Amer. Math. Soc. 161 (2003), no. 767.

[7] J. F. Jardine. Motivic symmetric spectra. Documenta Math. 5 (2000) 445-552.

[8] P. Landweber. Homological properties of comodules over MU*(MU) and BP*(BP). American Journal of Mathematics, Vol. 98, No. 3 (1976) 591-610.

[9] M. Levine. A survey of algebraic cobordism. Proc. International Conf. on Algebra; Algebra Colloq. 11 (2004) no.1 79-90.

[10] M. Levine. Algebraic cobordism II. Preprint. http://www.math.uiuc.edu/K-theory/0577

[11] M. Levine and F. Morel. Cobordisme algébrique I, II. C.R. Acad. Sci. Paris Sér. I Math. (8) 332 (2001) 723-728 and (9) 332 (2001) 815-820.

[12] M. Levine and F. Morel. Algebraic cobordism I. Preprint. http://www.math.uiuc.edu/K-theory/0547

[13] J. Lurie. Survey article on elliptic cohomology. Preprint. http://www-math.mit.edu/ ~lurie/papers/survey.pdf

[14] J. Lurie. Derived algebraic geometry I: stable ∞-categories. Preprint. http://arxiv.org/abs/math/0608228

[15] J. Lurie. Derived algebraic geometry II: noncommutative algebra. Preprint. http://arxiv.org/abs/math/0702299

[16] J. Lurie. Derived algebraic geometry III: commutative algebra. Preprint. http://arxiv.org/abs/math/0703204

[17] J. Lurie. Higher topos theory. Preprint. http://arxiv.org/abs/math/0608040

[18] F. Morel. On the motivic π_0 of the sphere spectrum. *Axiomatic, Enriched and Motivic Homotopy Theory* (ed. J.P.C. Greenlees) NATO Science Series II #131 (2004) 219-260

[19] F. Morel. *Homotopy Theory of Schemes*, (trans. J.D. Lewis). A.M. Soc. Texts and Monographs #12 (2006).

[20] F. Morel and V. Voevodsky. $\mathbb{A}^1$ homotopy theory of schemes. Publ. IHES 90 (1999) 45-143.

[21] N. Naumann, M. Spitzweck and P.A. Østvær. Motivic Landweber exactness. Preprint. http://arxiv.org/abs/0806.0274

[22] I. Panin, K. Pimenov and O. Röndigs. On Voevodsky's algebraic K-theory spectrum BGL. Preprint. http://www.math.uiuc.edu/K-theory/0838

[23] I. Panin, K. Pimenov and O. Röndigs. A universality theorem for Voevodsky's algebraic cobordism spectrum. Homology, Homotopy and Applications, Vol. 10 (2008), No.2, pp. 211-226.

[24] I. Panin, K. Pimenov and O. Röndigs. On the relation of Voevodsky's algebraic cobordism to Quillen's K-theory. Invent. Math. 175 (2009), No. 2, 435-451.
Preprint. http://www.math.uiuc.edu/K-theory/0847

[25] D. G. Quillen. On the formal group laws of unoriented and complex cobordism theory. Bull. A.M.Soc. 75 (1969) 1293-1298.

[26] D. G. Quillen. Higher Algebraic K-theory I. Lecture Notes in Math. 341 (1973) 85-147.

[27] V. P. Snaith. *Algebraic Cobordism and K-theory.* Mem. Amer. Math. Soc. #221 (1979).

[28] V. P. Snaith. Localised stable homotopy of some classifying spaces. Math. Proc. Camb. Phil. Soc. 89 (1981) 325-330.

[29] V. P. Snaith. A model for equivariant K-theory. C.R. Acad. Sci. Canada 3(3) (1981) 143–147.

[30] V. P. Snaith. Localised stable homotopy and algebraic K-theory. Memoirs Amer. Math.Soc. #280 (1983).

[31] V. P. Snaith. *Stable Homotopy Theory around the Arf-Kervaire invariant.* Preprint.

[32] M. Spitzweck and P. A. Østvær. The Bott inverted infinite projective space is homotopy algebraic K-theory. Bulletin of the London Mathematical Society 41(2) (2009) 281-292.

[33] R. Thomason and T. Trobaugh. Higher algebraic K-theory of schemes and derived categories. *The Grothendieck festschrift*, vol. 3 (1990), 247-436, Boston, Birkhauser.

[34] V. Voevodsky. $\mathbb{A}^1$-homotopy theory. Doc. Math. Extra Vol. ICM I (1998) 579-604.

[35] V. Voevodsky. Open problems in the motivic stable homotopy theory I. *Motives, Polylogarithms and Hodge Theory* Part I (Irvine CA 1998) 3-34, Int. Press Lect. Ser. 3 I Int. Press Somerville MA (2002).

[36] V. Voevodsky. On the zero slice of the sphere spectrum. Proc. Steklov Inst. Math. (translation) (2004) no.3 (246) 93-102.

[37] V. Voevodsky, A. Suslin and E.M. Friedlander. *Cycles, Transfers and Motivic Homology Theories.* Annals of Math. Studies #143, Princeton Univ. Press (2000).

[38] C. A. Weibel. Homotopy algebraic K-theory. Contemp. Math. #83 (1988) 461-488.

David Gepner
Department of Mathematics
University of Illinois at Chicago
851 S. Morgan Street
Chicago, IL 60607
USA

Victor Snaith
Department of Pure Mathematics
University of Sheffield
Hounsfield Road
Sheffield S3 7RH
United Kingdom

p-Adic Monodromy of the Universal Deformation of a HW-Cyclic Barsotti-Tate Group

Yichao Tian

Received: September 5, 2009

Communicated by Peter Schneider

Abstract. Let k be an algebraically closed field of characteristic $p > 0$, and G be a Barsotti-Tate over k. We denote by $\mathbf{S}$ the "algebraic" local moduli in characteristic p of G, by $\mathbf{G}$ the universal deformation of G over $\mathbf{S}$, and by $\mathbf{U} \subset \mathbf{S}$ the ordinary locus of $\mathbf{G}$. The étale part of $\mathbf{G}$ over $\mathbf{U}$ gives rise to a monodromy representation $\rho_{\mathbf{G}}$ of the fundamental group of $\mathbf{U}$ on the Tate module of $\mathbf{G}$. Motivated by a famous theorem of Igusa, we prove in this article that $\rho_{\mathbf{G}}$ is surjective if G is connected and HW-cyclic. This latter condition is equivalent to saying that Oort's a-number of G equals 1, and it is satisfied by all connected one-dimensional Barsotti-Tate groups over k.

2000 Mathematics Subject Classification: 13D10, 14L05, 14H30, 14B12, 14D15, 14L15
Keywords and Phrases: Barsotti-Tate groups (p-divisible groups), p-adic monodromy representation, universal deformation, Hasse-Witt maps.

1. Introduction

1.1. A classical theorem of Igusa says that the monodromy representation associated with a versal family of ordinary elliptic curves in characteristic $p > 0$ is surjective [Igu, Ka2]. This important result has deep consequences in the theory of p-adic modular forms, and inpsired various generalizations. Faltings and Chai [Ch2, FC] extended it to the universal family over the moduli space of higher dimensional principally polarized ordinary abelian varieties in characteristic p, and Ekedahl [Eke] generalized it to the jacobian of the universal n-pointed curve in characteristic p, equipped with a symplectic level structure. Recently, Chai and Oort [CO] proved the maximality of the p-adic monodromy over each "central leaf" in the moduli space of abelian varieties which is not contained in the supersingular locus. We refer to Deligne-Ribet [DR] and Hida [Hid] for other generalizations to some moduli spaces of PEL-type and their arithmetic applications. Though it has been formulated in a global setting, the proof of Igusa's theorem is purely local, and it has got also local generalizations. Gross [Gro] generalized it to one-dimensional formal $\mathscr{O}$-modules over a complete discrete valuation ring of characteristic p, where $\mathscr{O}$ is the integral closure of $\mathbb{Z}_p$ in a finite extension of $\mathbb{Q}_p$. We refer to Chai [Ch2] and Achter-Norman [AN] for more results on local monodromy of Barsotti-Tate groups. Motivated by these results, it has been longly expected/conjectured that the monodromy of a *versal* family of ordinary Barsotti-Tate groups in characteristic $p > 0$ is maximal. The aim of this paper is to prove the surjectivity of the monodromy representation associated with the universal deformation in characteristic p of a certain class of Barsotti-Tate groups.

1.2. To describe our main result, we introduce first the notion of HW-cyclic Barsotti-Tate groups. Let k be an algebraically closed field of characteristic $p > 0$, and G be a Barsotti-Tate group over k. We denote by $G^\vee$ the Serre dual of G, and by $\mathrm{Lie}(G^\vee)$ its Lie algebra. The Frobenius homomorphism of G (or dually the Verschiebung of $G^\vee$) induces a semi-linear endomorphism φ_G on $\mathrm{Lie}(G^\vee)$, called the Hasse-Witt map of G (2.6.1). We say that G is *HW-cyclic*, if $c = \dim(G^\vee) \geq 1$ and there is a $v \in \mathrm{Lie}(G^\vee)$ such that $v, \varphi_G(v), \cdots, \varphi_G^{c-1}(v)$ form a basis of $\mathrm{Lie}(G^\vee)$ over k (4.1). We prove in 4.7 that G is HW-cyclic and non-ordinary if and only if the a-number of G, defined previously by Oort, equals 1. Basic examples of HW-cyclic Barsotti-Tate groups are given as follows. Let r, s be relatively prime integers such that $0 \leq s \leq r$ and $r \neq 0$, $\lambda = s/r$, G^λ be the Barsotti-Tate group over k whose (contravariant) Dieudonné module is generated by an element e over the non-commutative Dieudonné ring with the relation $(F^{r-s} - V^s) \cdot e = 0$ (4.10). It is easy to see that G^λ is HW-cyclic for any $0 < \lambda < 1$. Any connected Barsotti-Tate group over k of dimension 1 and height h is isomorphic to $G^{1/h}$ [Dem, Chap.IV §8].

Let G be a Barsotti-Tate group of dimension d and height $c+d$ over k; assume $c \geq 1$. We denote by $\mathbf{S}$ the "algebraic" local moduli of G in characteristic p, and by $\mathbf{G}$ be the universal deformation of G over $\mathbf{S}$ (cf. 3.8). The scheme $\mathbf{S}$ is affine of ring $R \simeq k[[(t_{i,j})_{1\leq i\leq c, 1\leq j\leq d}]]$, and the Barsotti-Tate group $\mathbf{G}$ is obtained by algebraizing the formal universal deformation of G over $\mathrm{Spf}(R)$ (3.7). Let $\mathbf{U}$ be the ordinary locus of $\mathbf{G}$ (*i.e.* the open subscheme of $\mathbf{S}$ parametrizing the ordinary fibers of $\mathbf{G}$), and $\overline{\eta}$ a geometric point over the generic point of $\mathbf{U}$. For any integer $n \geq 1$, we denote by $\mathbf{G}(n)$ the kernel of the multiplication by p^n on $\mathbf{G}$, and by

$$\mathrm{T}_p(\mathbf{G}, \overline{\eta}) = \varprojlim_n \mathbf{G}(n)(\overline{\eta})$$

the Tate module of $\mathbf{G}$ at $\overline{\eta}$. This is a free $\mathbb{Z}_p$-module of rank c. We consider the monodromy representation attached to the étale part of $\mathbf{G}$ over $\mathbf{U}$

$$\rho_{\mathbf{G}} : \pi_1(\mathbf{U}, \overline{\eta}) \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(\mathbf{G}, \overline{\eta})) \simeq \mathrm{GL}_c(\mathbb{Z}_p). \tag{1.2.1}$$

The aim of this paper is to prove the following :

Theorem 1.3. *If G is connected and HW-cyclic, then the monodromy representation $\rho_{\mathbf{G}}$ is surjective.*

Igusa's theorem mentioned above corresponds to Theorem 1.3 for $G = G^{1/2}$ (cf. 5.7). My interest in the p-adic monodromy problem started with the second part of my PhD thesis [Ti1], where I guessed 1.3 for $G = G^{\lambda}$ with $0 < \lambda < 1$ and proved it for $G^{1/3}$. After I posted the manuscript on ArXiv [Ti2], Strauch proved the one-dimensional case of 1.3 by using Drinfeld's level structures [Str, Theorem 2.1]. Later on, Lau [Lau] proved 1.3 without the assumption that G is HW-cyclic. By using the Newton stratification of the universal deformation space of G due to Oort, Lau reduced the higher dimensional case to the one-dimensional case treated by Strauch. In fact, Strauch and Lau considered more generally the monodromy representation over each p-rank stratum of the universal deformation space. In this paper, we provide first a different proof of the one-dimensional case of 1.3. Our approach is purely characteristic p, while Strauch used Drinfeld's level structure in characteristic 0. Then by following Lau's strategy, we give a new (and easier) argument to reduce the general case of 1.3 to the one-dimensional case for HW-cyclic groups. The essential part of our argument is a versality criterion by Hasse-Witt maps of deformations of a connected one-dimensional Barsotti-Tate group (Prop. 4.11). This criterion can be considered as a generalization of another theorem of Igusa which claims that the Hasse invariant of a versal family of elliptic curves in characteristic p has simple zeros. Compared with Strauch's approach, our characteristic p approach has the advantage of giving also results on the monodromy of Barsotti-Tate groups over a discrete valuation ring of characteristic p.

1.4. Let $A = k[[\pi]]$ be the ring of formal power series over k in the variable π, K its fraction field, and $\mathtt{v}$ the valuation on K normalized by $\mathtt{v}(\pi) = 1$. We fix an algebraic closure $\overline{K}$ of K, and let K^{sep} be the separable closure of K contained in $\overline{K}$, I be the Galois group of K^{sep} over K, $I_p \subset I$ be the wild inertia subgroup, and $I_t = I/I_p$ the tame inertia group. For every integer $n \geq 1$, there is a canonical surjective character $\theta_{p^n-1} : I_t \to \mathbb{F}_{p^n}^{\times}$ (5.2), where $\mathbb{F}_{p^n}$ is the finite subfield of k with p^n elements.
We put $S = \mathrm{Spec}(A)$. Let G be a Barsotti-Tate group over S, $G^{\vee}$ be its Serre dual, $\mathrm{Lie}(G^{\vee})$ the Lie algebra of $G^{\vee}$, and φ_G the Hasse-Witt map of G, *i.e.* the semi-linear endomorphism of $\mathrm{Lie}(G^{\vee})$ induced by the Frobenius of G. We define $h(G)$ to be the valuation of the determinant of a matrix of φ_G, and call it the *Hasse invariant* of G (5.4). We see easily that $h(G) = 0$ if and only if G is ordinary over S, and $h(G) < \infty$ if and only if G is generically ordinary. If G is connected of height 2 and dimension 1, then $h(G) = 1$ is equivalent to that G is versal (5.7).

Proposition 1.5. *Let $S = \mathrm{Spec}(A)$ be as above, G be a connected HW-cyclic Barsotti-Tate group with Hasse invariant $h(G) = 1$, and $G(1)$ the kernel of the multiplication by p on G. Then the action of I on $G(1)(\overline{K})$ is tame; moverover,*

$G(1)(\overline{K})$ is an $\mathbb{F}_{p^c}$-vector space of dimension 1 on which the induced action of I_t is given by the surjective character $\theta_{p^c-1} : I_t \to \mathbb{F}_{p^c}^{\times}$.

This proposition is an analog in characteristic p of Serre's result [Se3, Prop. 9] on the tameness of the monodromy associated with one-dimensional formal groups over a trait of mixed characteristic. We refer to 5.8 for the proof of this proposition and more results on the p-adic monodromy of HW-cyclic Barsotti-Tate groups over a trait in characteristic p.

1.6. This paper is organized as follows. In Section 2, we review some well known facts on ordinary Barsotti-Tate groups. Section 3 contains some preliminaries on the Dieudonné theory and the deformation theory of Barsotti-Tate groups. In Section 4, after establishing some basic properties of HW-cyclic groups, we give the fundamental relation between the versality of a Barsotti-Tate group and the coefficients of its Hasse-Witt matrix (Prop. 4.11). Section 5 is devoted to the study of the monodromy of a HW-cyclic Barsotti-Tate group over a complete trait of characteristic p. Section 6 is totally elementary, and contains a criterion (6.3) for the surjectivity of a homomorphism from a profinite group to $\mathrm{GL}_n(\mathbb{Z}_p)$. Section 7 is the heart of this work, and it contains a proof of Theorem 1.3 in the one-dimensional case. Finally in Section 8, we follow Lau's strategy and complete the proof of 1.3 by reducing the general case to the one-dimensional case treated in Section 7.
The proof in Section 7 of 1.3 in the one-dimensional case is based on an induction on the height $n+1 \geq 2$ of G. The case $n = 1$ is just the classical Igusa's theorem (5.7). For $n \geq 2$, by lemmas 6.3 and 6.5, it suffices to prove the following two statements: (a) the image of reduction modulo p of $\rho_{\mathbf{G}}$ contains a non-split Cartan subgroup; (b) under a suitable basis, the image of $\rho_{\mathbf{G}}$ contains all matrix of the form $\begin{pmatrix} B & b \\ 0 & 1 \end{pmatrix}$ with $B \in \mathrm{GL}_{n-1}(\mathbb{Z}_p)$ and $b \in \mathrm{M}_{(n-1)\times 1}(\mathbb{Z}_p)$. The first statement follows easily from 1.5 by considering a certain base change of $\mathbf{G}$ to a complete discrete valuation ring. To prove (b), we consider the formal completion $\mathrm{Spec}(R')$ of the localization of the local moduli $\mathbf{S} = \mathrm{Spec}(R)$ of G at the generic point of the locus where the universal deformation $\mathbf{G}$ has p-rank ≤ 1 (7.4). The ring R' is a complete regular ring of dimension $n-1$, and the Barsotti-Tate group $\mathscr{G}' = \mathbf{G} \otimes_R R'$ has a connected part of height n and an étale part of height 1. Let K_0 be the residue field of R', and $\overline{K}_0$ an algebraic closure of K_0. In order to apply the induction hypothesis, we consider the set of k-algebra homomorphisms $\sigma : R' \to \widetilde{R'} = \overline{K}_0[[t_1, \cdots, t_{n-1}]]$ lifting the natural inclusion $K_0 \to \overline{K}_0$. The key point is that, the natural map $\sigma \mapsto \mathscr{G}_{\widetilde{R'},\sigma} = \mathscr{G}' \otimes_{R',\sigma} \widetilde{R'}$ gives a bijection between the set of such σ's and the set of deformations of $\mathscr{G}_{\overline{K}_0} = \mathscr{G}' \otimes_{R'} \overline{K}_0$ to $\widetilde{R'}$; moreover, we can compute explicitly the Hasse-Witt map of the connected component $\mathscr{G}^{\circ}_{\widetilde{R'},\sigma}$ of $\mathscr{G}_{\widetilde{R'},\sigma}$ (Lemma 7.8). From the versality criterion for one-dimensional Barsotti-Tate groups in terms of the Hasse-Witt map established in Section 4 (Prop. 4.11), it follows immediately that there exists a σ such that the Barsotti-Tate group $\mathscr{G}^{\circ}_{\widetilde{R'},\sigma}$, which

is connected and one-dimensional of height n, is the universal deformation of its closed fiber. We fix such a σ. Then the set of all σ' with $\mathscr{G}^{\circ}_{\widetilde{R'},\sigma'} \simeq \mathscr{G}^{\circ}_{\widetilde{R'},\sigma}$ as deformations of their common closed fiber is actually a group isomorphic to $\mathrm{Ext}^1_{\widetilde{R'}}(\mathbb{Q}_p/\mathbb{Z}_p, \mathscr{G}^{\circ}_{\widetilde{R'},\sigma})$ (Prop. 3.10). Let σ_1 be the element corresponding to neutral element in $\mathrm{Ext}^1_{\widetilde{R'}}(\mathbb{Q}_p/\mathbb{Z}_p, \mathscr{G}^{\circ}_{\widetilde{R'},\sigma})$. Applying the induction hypothesis to $\mathscr{G}^{\circ}_{\widetilde{R'},\sigma_1}$, we see that the monodromy group of $\mathscr{G}_{\widetilde{R'},\sigma_1}$, hence that of $\mathbf{G}$, contains the subgroup $\begin{pmatrix} \mathrm{GL}_{n-1}(\mathbb{Z}_p) & 0 \\ 0 & 1 \end{pmatrix}$ under a suitable basis of the Tate module (7.5.3). In order to conclude the proof, we need another σ_2 such that $\mathscr{G}_{\widetilde{R'},\sigma_2}$ has the same connected component as $\mathscr{G}_{\widetilde{R'},\sigma_1}$, and that the induced extension between the Tate module of the étale part of $\mathscr{G}_{\widetilde{R'},\sigma_2}$ and that of $\mathscr{G}^{\circ}_{R',\sigma_2}$ is non-trivial after reduction modulo p (see 7.5 and 7.5.4). To verify the existence of such a σ_2, we reduce the problem to a similar situation over a complete trait of characteristic p (see 7.9), and we use a criterion of non-triviality of extensions by Hasse-Witt maps (5.12).

1.7. Acknowledgement. This paper is an expanded version of the second part of my Ph.D. thesis at University Paris 13. I would like to express my great gratitude to my thesis advisor Prof. A. Abbes for his encouragement during this work, and also for his various helpful comments on earlier versions of this paper. I also thank heartily E. Lau, F. Oort and M. Strauch for interesting discussions and valuable suggestions.

1.8. Notations. Let S be a scheme of characteristic $p > 0$. A *BT-group* over S stands for a Barsotti-Tate group over S. Let G be a commutative finite group scheme (*resp.* a BT-group) over S. We denote by $G^\vee$ its Cartier dual (*resp.* its Serre dual), by ω_G the sheaf of invariant differentials of G over S, and by $\mathrm{Lie}(G)$ the sheaf of Lie algebras of G. If $S = \mathrm{Spec}(A)$ is affine and there is no risk of confusions, we also use ω_G and $\mathrm{Lie}(G)$ to denote the corresponding A-modules of global sections. We put $G^{(p)}$ the pull-back of G by the absolute Frobenius of S, $F_G \colon G \to G^{(p)}$ the Frobenius homomorphism and $V_G \colon G^{(p)} \to G$ the Verschiebung homomorphism. If G is a BT-group and n an integer ≥ 1, we denote by $G(n)$ the kernel of the multiplication by p^n on G; we have $G^\vee(n) = (G^\vee)(n)$ by definition. For an $\mathscr{O}_S$-module M, we denote by $M^{(p)} = \mathscr{O}_S \otimes_{F_S} M$ the scalar extension of M by the absolute Frobenius of $\mathscr{O}_S$. If $\varphi : M \to N$ be a semi-linear homomorphism of $\mathscr{O}_S$-modules, we denote by $\widetilde{\varphi} : M^{(p)} \to N$ the linearization of φ, *i.e.* we have $\widetilde{\varphi}(\lambda \otimes x) = \lambda \cdot \varphi(x)$, where λ (*resp.* x) is a local section of $\mathscr{O}_S$ (*resp.* of M).
Starting from Section 5, k will denote an algebraically closed field of characteristic $p > 0$.

2. Review of ordinary Barsotti-Tate groups

In this section, S denotes a scheme of characteristic $p > 0$.

2.1. Let G be a commutative group scheme, locally free of finite type over S. We have a canonical isomorphism of coherent $\mathscr{O}_S$-modules [Ill, 2.1]

$$\mathrm{Lie}(G^\vee) \simeq \mathscr{H}om_{S_{\mathrm{fppf}}}(G, \mathbb{G}_a), \tag{2.1.1}$$

where $\mathscr{H}om_{S_{\mathrm{fppf}}}$ is the sheaf of homomorphisms in the category of abelian fppf-sheaves over S, and $\mathbb{G}_a$ is the additive group scheme. Since $\mathbb{G}_a^{(p)} \simeq \mathbb{G}_a$, the Frobenius homomorphism of $\mathbb{G}_a$ induces an endomorphism

$$\varphi_G : \mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G^\vee), \tag{2.1.2}$$

semi-linear with respect to the absolute Frobenius map $F_S : \mathscr{O}_S \to \mathscr{O}_S$; we call it the *Hasse-Witt* map of G. By the functoriality of Frobenius, φ_G is also the canonical map induced by the Frobenius of G, or dually by the Verschiebung of $G^\vee$.

2.2. By a *commutative p-Lie algebra* over S, we mean a pair (L, φ), where L is an $\mathscr{O}_S$-module locally free of finite type, and $\varphi : L \to L$ is a semi-linear endomorphism with respect to the absolute Frobenius $F_S : \mathscr{O}_S \to \mathscr{O}_S$. When there is no risk of confusions, we omit φ from the notation. We denote by $p\text{-}\mathfrak{Lie}_S$ the category of commutative p-Lie algebras over S.
Let (L, φ) be an object of $p\text{-}\mathfrak{Lie}_S$. We denote by

$$\mathscr{U}(L) = \mathrm{Sym}(L) = \oplus_{n\geq 0} \mathrm{Sym}^n(L),$$

the symmetric algebra of L over $\mathscr{O}_S$. Let $\mathscr{I}_p(L)$ be the ideal sheaf of $\mathscr{U}(L)$ defined, for an open subset $V \subset S$, by

$$\Gamma(V, \mathscr{I}_p(L)) = \{x^{\otimes p} - \varphi(x) \ ; \ x \in \Gamma(V, \mathscr{U}(L))\},$$

where $x^{\otimes p} = x \otimes x \otimes \cdots \otimes x \in \Gamma(V, \mathrm{Sym}^p(L))$. We put $\mathscr{U}_p(L) = \mathscr{U}(L)/\mathscr{I}_p(L)$, and call it the *p-enveloping algebra of* (L, φ). We endow $\mathscr{U}_p(L)$ with the structure of a Hopf-algebra with the comultiplication given by $\Delta(x) = 1 \otimes x + x \otimes 1$ and the coinverse given by $i(x) = -x$.
Let G be a commutative group scheme, locally free of finite type over S. We say that G is *of coheight one* if the Verschiebung $V_G : G^{(p)} \to G$ is the zero homomorphism. We denote by $\mathfrak{G}\mathrm{V}_S$ the category of such objects. For an object G of $\mathfrak{G}\mathrm{V}_S$, the Frobenius $F_{G^\vee}$ of $G^\vee$ is zero, so the Lie algebra $\mathrm{Lie}(G^\vee)$ is locally free of finite type over $\mathscr{O}_S$ ([DG] $\mathrm{VII_A}$ Théo. 7.4(iii)). The Hasse-Witt map of G (2.1.2) endows $\mathrm{Lie}(G^\vee)$ with a commutative p-Lie algebra structure over S.

PROPOSITION 2.3 ([DG] $\mathrm{VII_A}$, Théo. 7.2 et 7.4). *The functor $\mathfrak{G}\mathrm{V}_S \to p\text{-}\mathfrak{Lie}_S$ defined by $G \mapsto \mathrm{Lie}(G^\vee)$ is an anti-equivalence of categories; a quasi-inverse is given by $(L, \varphi) \mapsto \mathrm{Spec}(\mathscr{U}_p(L))$.*

2.4. Assume $S = \mathrm{Spec}(A)$ affine. Let (L, φ) be an object of $p\text{-}\mathfrak{Lie}_S$ such that L is free of rank n over $\mathscr{O}_S$, $(e_1, \cdots, e_n)$ be a basis of L over $\mathscr{O}_S$, $(h_{ij})_{1\leq i,j\leq n}$ be the matrix of φ under the basis $(e_1, \cdots, e_n)$, *i.e.* $\varphi(e_j) = \sum_{i=1}^n h_{ij} e_i$ for

$1 \leq j \leq n$. Then the group scheme attached to (L, φ) is explicitly given by

$$\mathrm{Spec}(\mathscr{U}_p(L)) = \mathrm{Spec}\left(A[X_1, \cdots, X_n]/(X_j^p - \sum_{i=1}^{n} h_{ij} X_i)_{1 \leq j \leq n}\right),$$

with the comultiplication $\Delta(X_j) = 1 \otimes X_j + X_j \otimes 1$. By the Jacobian criterion of étaleness [EGA, IV_0 22.6.7], the finite group scheme $\mathrm{Spec}(\mathscr{U}_p(L))$ is étale over S if and only if the matrix $(h_{ij})_{1 \leq i,j \leq n}$ is invertible. This condition is equivalent to that the linearization of φ is an isomorphism.

Corollary 2.5. *An object G of $\mathfrak{G}\mathrm{V}_S$ is étale over S, if and only if the linearization of its Hasse-Witt map* (2.1.2) *is an isomorphism.*

Proof. The problem being local over S, we may assume S affine and $L = \mathrm{Lie}(G^\vee)$ free over $\mathscr{O}_S$. By Theorem 2.3, G is isomorphic to $\mathrm{Spec}(\mathscr{U}_p(L))$, and we conclude by the last remark of 2.4. □

2.6. Let G be a BT-group over S of height $c + d$ and dimension d. The Lie algebra $\mathrm{Lie}(G^\vee)$ is an $\mathscr{O}_S$-module locally free of rank c, and canonically identified with $\mathrm{Lie}(G^\vee(1))$([BBM] 3.3.2). We define the *Hasse-Witt map* of G

$$\varphi_G : \mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G^\vee) \tag{2.6.1}$$

to be that of $G(1)$ (2.1.2).

2.7. Let k be a field of characteristic $p > 0$, G be a BT-group over k. Recall that we have a canonical exact sequence of BT-groups over k

$$0 \to G^\circ \to G \to G^{\text{ét}} \to 0 \tag{2.7.1}$$

with G° connected and $G^{\text{ét}}$ étale ([Dem] Chap.II, §7). This induces an exact sequence of Lie algebras

$$0 \to \mathrm{Lie}(G^{\text{ét}\vee}) \to \mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G^{\circ\vee}) \to 0, \tag{2.7.2}$$

compatible with Hasse-Witt maps.

Proposition 2.8. *Let k be a field of characteristic $p > 0$, G be a BT-group over k. Then* $\mathrm{Lie}(G^{\text{ét}\vee})$ *is the unique maximal k-subspace V of* $\mathrm{Lie}(G^\vee)$ *with the following properties:*
(a) *V is stable under φ_G;*
(b) *the restriction of φ_G to V is injective.*

Proof. It is clear that $\mathrm{Lie}(G^{\text{ét}\vee})$ satisfies property (a). We note that the Verschiebung of $G^{\text{ét}}(1)$ vanishes; so $G^{\text{ét}}(1)$ is in the category $\mathfrak{G}\mathrm{V}_{\mathrm{Spec}(k)}$. Since k is a field, 2.5 implies that the restriction of φ_G to $\mathrm{Lie}(G^{\text{ét}\vee})$, which coincides with $\varphi_{G^{\text{ét}}}$, is injective. This proves that $\mathrm{Lie}(G^{\text{ét}\vee})$ verifies (b). Conversely, let V be an arbitrary k-subspace of $\mathrm{Lie}(G^\vee)$ with properties (a) and (b). We have to show that $V \subset \mathrm{Lie}(G^{\text{ét}\vee})$. Let σ be the Frobenius endomorphism of k. If M is a k-vector space, for each integer $n \geq 1$, we put $M^{(p^n)} = k \otimes_{\sigma^n} M$, *i.e.* we have $1 \otimes ax = \sigma^n(a) \otimes x$ in $k \otimes_{\sigma^n} M$ for $a \in k, x \in M$. Since $\varphi_G|_V : V \to V$ is injective by assumption, the linearization $\widetilde{\varphi_G^n}|_{V^{(p^n)}} : V^{(p^n)} \to V$ of $\varphi_G^n|_V$

is injective (hence bijective) for any $n \geq 1$. We have $V = \widetilde{\varphi^n_G}(V^{(p^n)})$. Since G° is connected, there is an integer $n \geq 1$ such that the n-th iterated Frobenius $F^n_{G^\circ(1)} : G^\circ(1) \to G^\circ(1)^{(p^n)}$ vanishes. Hence by definition, the linearized n-iterated Hasse-Witt map $\widetilde{\varphi^n_{G^\circ}} : \mathrm{Lie}(G^{\circ\vee})^{(p^n)} \to \mathrm{Lie}(G^{\circ\vee})$ is zero. By the compatibility of Hasse-Witt maps, we have $\widetilde{\varphi^n_G}(\mathrm{Lie}(G^\vee)^{(p^n)}) \subset \mathrm{Lie}(G^{\text{ét}\vee})$; in particular, we have $V = \widetilde{\varphi^n_G}(V^{(p^n)}) \subset \mathrm{Lie}(G^{\text{ét}\vee})$. This completes the proof. $\square$

COROLLARY 2.9. *Let k be a field of characteristic $p > 0$, G be a BT-group over k. Then G is connected if and only if φ_G is nilpotent.*

Proof. In the proof of the proposition, we have seen that the Hasse-Witt map of the connected part of G is nilpotent. So the "only if" part is verified. Conversely, if φ_G is nilpotent, $\mathrm{Lie}(G^{\text{ét}\vee})$ is zero by the proposition. Therefore G is connected. $\square$

DEFINITION 2.10. Let S be a scheme of characteristic $p > 0$, G be a BT-group over S. We say that G is *ordinary* if there exists an exact sequence of BT-groups over S

$$0 \to G^{\mathrm{mult}} \to G \to G^{\text{ét}} \to 0, \tag{2.10.1}$$

such that G^{mult} is multiplicative and $G^{\text{ét}}$ is étale.

We note that when it exists, the exact sequence (2.10.1) is unique up to a unique isomorphism, because there is no non-trivial homomorphisms between a multiplicative BT-group and an étale one in characteristic $p > 0$. The property of being ordinary is clearly stable under arbitrary base change and Serre duality. If S is the spectrum of a field of characteristic $p > 0$, G is ordinary if and only if its connected part G° is of multiplicative type.

PROPOSITION 2.11. *Let G be a BT-group over S. The following conditions are equivalent:*
(a) *G is ordinary over S.*
(b) *For every $x \in S$, the fiber $G_x = G \otimes_S \kappa(x)$ is ordinary over $\kappa(x)$.*
(c) *The finite group scheme $\mathrm{Ker}\, V_G$ is étale over S.*
(c') *The finite group scheme $\mathrm{Ker}\, F_G$ is of multiplicative type over S.*
(d) *The linearization of the Hasse-Witt map φ_G is an isomorphism.*

First, we prove the following lemmas.

LEMMA 2.12. *Let T be a scheme, H be a commutative group scheme locally free of finite type over T. Then H is étale (*resp.* of multiplicative type) over T if and only if, for every $x \in T$, the fiber $H \otimes_T \kappa(x)$ is étale (*resp.* of multiplicative type) over $\kappa(x)$.*

Proof. We will consider only the étale case; the multiplicative case follows by duality. Since H is T-flat, it is étale over T if and only if it is unramified over T. By [EGA, IV 17.4.2], this condition is equivalent to that $H \otimes_T \kappa(x)$ is unramified over $\kappa(x)$ for every point $x \in T$. Hence the conclusion follows. $\square$

Lemma 2.13. *Let G be a BT-group over S. Then* $\operatorname{Ker} V_G$ *is an object of the category* $\mathfrak{G}\mathrm{V}_S$, i.e. *it is locally free of finite type over S, and its Verschiebung is zero. Moreover, we have a canonical isomorphism* $(\operatorname{Ker} V_G)^\vee \simeq \operatorname{Ker} F_{G^\vee}$, *which induces an isomorphism of Lie algebras* $\mathrm{Lie}\big((\operatorname{Ker} V_G)^\vee\big) \simeq \mathrm{Lie}(\operatorname{Ker} F_{G^\vee}) = \mathrm{Lie}(G^\vee)$, *and the Hasse-Witt map* (2.1.2) *of* $\operatorname{Ker} V_G$ *is identified with* φ_G (2.6.1).

Proof. The group scheme $\operatorname{Ker} V_G$ is locally free of finite type over S ([Ill] 1.3(b)), and we have a commutative diagram

$$\begin{tikzcd}
(\operatorname{Ker} V_G)^{(p)} \arrow[r, "V_{\operatorname{Ker} V_G}"] \arrow[d, hook] & \operatorname{Ker} V_G \arrow[d, hook] \\
(G^{(p)})^{(p)} \arrow[r, "V_{G^{(p)}}"] & G^{(p)}
\end{tikzcd}$$

By the functoriality of Verschiebung, we have $V_{G^{(p)}} = (V_G)^{(p)}$ and $\operatorname{Ker} V_{G^{(p)}} = (\operatorname{Ker} V_G)^{(p)}$. Hence the composition of the left vertical arrow with $V_{G^{(p)}}$ vanishes, and the Verschiebung of $\operatorname{Ker} V_G$ is zero.
By Cartier duality, we have $(\operatorname{Ker} V_G)^\vee = \operatorname{Coker}(F_{G^\vee(1)})$. Moreover, the exact sequence

$$\cdots \to G^\vee(1) \xrightarrow{F_{G^\vee(1)}} \big(G^\vee(1)\big)^{(p)} \xrightarrow{V_{G^\vee(1)}} G^\vee(1) \to \cdots,$$

induces a canonical isomorphism

$$\operatorname{Coker}(F_{G^\vee(1)}) \xrightarrow{\sim} \operatorname{Im}(V_{G^\vee(1)}) = \operatorname{Ker} F_{G^\vee(1)} = \operatorname{Ker} F_{G^\vee}. \tag{2.13.1}$$

Hence, we deduce that

$$(\operatorname{Ker} V_G)^\vee \simeq \operatorname{Coker}(F_{G^\vee(1)}) \xrightarrow{\sim} \operatorname{Ker} F_{G^\vee} \hookrightarrow G^\vee(1). \tag{2.13.2}$$

Since the natural injection $\operatorname{Ker} F_{G^\vee} \to G^\vee(1)$ induces an isomorphism of Lie algebras, we get

$$\mathrm{Lie}\big((\operatorname{Ker} V_G)^\vee\big) \simeq \mathrm{Lie}(\operatorname{Ker} F_{G^\vee}) = \mathrm{Lie}(G^\vee(1)) = \mathrm{Lie}(G^\vee). \tag{2.13.3}$$

It remains to prove the compatibility of the Hasse-Witt maps with (2.13.3). We note that the dual of the morphism (2.13.2) is the canonical map $F : G(1) \to \operatorname{Ker} V_G = \operatorname{Im}(F_{G(1)})$ induced by $F_{G(1)}$. Hence by (2.1.1), the isomorphism (2.13.3) is identified with the functorial map

$$\mathscr{H}om_{S_{\mathrm{fppf}}}(\operatorname{Ker} V_G, \mathbb{G}_a) \to \mathscr{H}om_{S_{\mathrm{fppf}}}(G(1), \mathbb{G}_a)$$

induced by F, and its compatibility with the Hasse-Witt maps follows easily from the definition (2.1.2). □

Proof of 2.11. (a)⇒(b). Indeed, the ordinarity of G is stable by base change.
(b)⇒(c). By Lemma 2.12, it suffices to verify that for every point $x \in S$, the fiber $(\operatorname{Ker} V_G) \otimes_S \kappa(x) \simeq \operatorname{Ker} V_{G_x}$ is étale over $\kappa(x)$. Since G_x is assumed to be ordinary, its connected part $(G_x)^\circ$ is multiplicative. Hence, the Verschiebung of

$(G_x)^\circ$ is an isomorphism, and $\operatorname{Ker} V_{G_x}$ is canonically isomorphic to $\operatorname{Ker} V_{G_x^{\text{ét}}} \subset (G_x^{\text{ét}})^{(p)} \simeq (G_x^{(p)})^{\text{ét}}$, so our assertion follows.
$(c) \Leftrightarrow (d)$. It follows immediately from Lemma 2.13 and Corollary 2.5.
(c)$\Leftrightarrow$(c'). By 2.12, we may assume that S is the spectrum of a field. So the category of commutative finite group schemes over S is abelian. We will just prove (c)$\Rightarrow$(c'); the converse can be proved by duality. We have a fundamental short exact sequence of finite group schemes

$$0 \to \operatorname{Ker} F_G \to G(1) \xrightarrow{F} \operatorname{Ker} V_G \to 0, \tag{2.13.4}$$

where F is induced by $F_{G(1)}$, That induces a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \big(\operatorname{Ker} F_G\big)^{(p)} \arrow[r] \arrow[d, "V'"] & \big(G(1)\big)^{(p)} \arrow[r, "F^{(p)}"] \arrow[d, "V_{G(1)}"] & \big(\operatorname{Ker} V_G\big)^{(p)} \arrow[r] \arrow[d, "V''"] & 0 \\
0 \arrow[r] & \operatorname{Ker} F_G \arrow[r] & G(1) \arrow[r, "F"] & \operatorname{Ker} V_G \arrow[r] & 0
\end{tikzcd}$$

where vertical arrows are the Verschiebung homomorphisms. We have seen that $V'' = 0$ (2.13). Therefore, by the snake lemma, we have a long exact sequence

(2.13.5)
$$\begin{aligned}0 \to \operatorname{Ker} V' \to \operatorname{Ker} V_{G(1)} \xrightarrow{\alpha} \big(\operatorname{Ker} V_G\big)^{(p)} \to \\ \to \operatorname{Coker} V' \to \operatorname{Coker} V_{G(1)} \xrightarrow{\beta} \operatorname{Ker} V_G \to 0,\end{aligned}$$

where the map α is the Frobenius of $\operatorname{Ker} V_G$ and β is the composed isomorphism

$$\operatorname{Coker}(V_{G(1)}) \simeq G(1)/\operatorname{Ker} F_{G(1)} \xrightarrow{\sim} \operatorname{Im}(F_{G(1)}) \simeq \operatorname{Ker} V_G.$$

Then condition (c) is equivalent to that α is an isomorphism; it implies that $\operatorname{Ker} V' = \operatorname{Coker} V' = 0$, *i.e.* the Verschiebung of $\operatorname{Ker} F_G$ is an isomorphism, and hence (c').
(c)$\Rightarrow$(a). For every integer $n > 0$, we denote by F_G^n the composed homomorphism

$$G \xrightarrow{F_G} G^{(p)} \xrightarrow{F_{G^{(p)}}} \cdots \xrightarrow{F_{G^{(p^{n-1})}}} G^{(p^n)},$$

and by V_G^n the composed homomorphism

$$G^{(p^n)} \xrightarrow{V_{G^{(p^{n-1})}}} G^{(p^{n-1})} \xrightarrow{V_{G^{(p^{n-2})}}} \cdots \xrightarrow{V_G} G;$$

F_G^n and V_G^n are isogenies of BT-groups. From the relation $V_G^n \circ F_G^n = p^n$, we deduce an exact sequence

$$0 \to \operatorname{Ker} F_G^n \to G(n) \xrightarrow{F^n} \operatorname{Ker} V_G^n \to 0, \tag{2.13.6}$$

where F^n is induced by F_G^n. For $1 \leq j < n$, we have a commutative diagram

$$\begin{tikzcd} G^{(p^n)} \arrow[rr, "V^{n-j}_{G^{(p^j)}}"] \arrow[dr, "V_G^n"'] & & G^{(p^j)} \arrow[dl, "V_G^j"] \\ & G. & \end{tikzcd} \tag{2.13.7}$$

One notices that $\operatorname{Ker} V^{n-j}_{G^{(p^j)}} = (\operatorname{Ker} V_G^{n-j})^{(p^j)}$ by the functoriality of Verschiebung . Since all maps in (2.13.7) are isogenies, we have an exact sequence

$$0 \to (\operatorname{Ker} V_G^{n-j})^{(p^j)} \xrightarrow{i'_{n-j,n}} \operatorname{Ker} V_G^n \xrightarrow{p_{n,j}} \operatorname{Ker} V_G^j \to 0. \tag{2.13.8}$$

Therefore, condition (c) implies by induction that $\operatorname{Ker} V_G^n$ is an étale group scheme over S. Hence the j-th iteration of the Frobenius $\operatorname{Ker} V_G^{n-j} \to (\operatorname{Ker} V_G^{n-j})^{(p^j)}$ is an isomorphism, and $\operatorname{Ker} V_G^{n-j}$ is identified with a closed subgroup scheme of $\operatorname{Ker} V_G^n$ by the composed map

$$i_{n-j,n} : \operatorname{Ker} V_G^{n-j} \xrightarrow{\sim} (\operatorname{Ker} V_G^{n-j})^{(p^j)} \xrightarrow{i'_{n-j,n}} \operatorname{Ker} V_G^n.$$

We claim that the kernel of the multiplication by p^{n-j} on $\operatorname{Ker} V_G^n$ is $\operatorname{Ker} V_G^{n-j}$. Indeed, from the relation $p^{n-j} \cdot \mathrm{Id}_{G^{(p^n)}} = F^{n-j}_{G^{(p^j)}} \circ V^{n-j}_{G^{(p^j)}}$, we deduce a commutative diagram (without dotted arrows)

$$\begin{tikzcd} \operatorname{Ker} V_G^n \arrow[rr] \arrow[dd, "p^{n-j}"'] \arrow[dr, dashed, "p_{n,j}"] & & G^{(p^n)} \arrow[dd, "p^{n-j}"'] \arrow[dr, "V^{n-j}_{G^{(p^j)}}"] & \\ & \operatorname{Ker} V_G^j \arrow[rr, dashed] \arrow[dl, dashed, "i_{j,n}"] & & G^{(p^j)} \arrow[dl, "F^{n-j}_{G^{(p^j)}}"] \\ \operatorname{Ker} V_G^n \arrow[rr] & & G^{(p^n)}. & \end{tikzcd} \tag{2.13.9}$$

It follows from (2.13.8) that the subgroup $\operatorname{Ker} V_G^n$ of $G^{(p^n)}$ is sent by $V^{n-j}_{G^{(p^j)}}$ onto $\operatorname{Ker} V_G^j$. Therefore diagram (2.13.9) remains commutative when completed by the dotted arrows, hence our claim. It follows from the claim that $(\operatorname{Ker} V_G^n)_{n \geq 1}$ constitutes an étale BT-group over S, denoted by $G^{\text{ét}}$. By duality, we have an exact sequence

$$0 \to \operatorname{Ker} F_G^j \to \operatorname{Ker} F_G^n \to (\operatorname{Ker} F_G^{n-j})^{(p^j)} \to 0. \tag{2.13.10}$$

Condition (c') implies by induction that $\operatorname{Ker} F_G^n$ is of multiplicative type. Hence the j-th iteration of Verschiebung $(\operatorname{Ker} F_G^{n-j})^{(p^j)} \to \operatorname{Ker} F_G^{n-j}$ is an isomorphism. We deduce from (2.13.10) that $(\operatorname{Ker} F_G^n)_{n \geq 1}$ form a multiplicative BT-group over S that we denote by G^{mult}. Then the exact sequences (2.13.6) give a decomposition of G of the form (2.10.1). $\square$

COROLLARY 2.14. *Let G be a BT-group over S, and S^{ord} be the locus in S of the points $x \in S$ such that $G_x = G \otimes_S \kappa(x)$ is ordinary over $\kappa(x)$. Then S^{ord} is open in S, and the canonical inclusion $S^{\mathrm{ord}} \to S$ is affine.*

The open subscheme S^{ord} of S is called the *ordinary locus* of G.

3. PRELIMINARIES ON DIEUDONNÉ THEORY AND DEFORMATION THEORY

3.1. We will use freely the conventions of 1.8. Let S be a scheme of characteristic $p > 0$, G be a Barsotti-Tate group over S, and $\mathbf{M}(G) = \mathbb{D}(G)_{(S,S)}$ be the coherent $\mathscr{O}_S$-module obtained by evaluating the (contravariant) Dieudonné crystal of G at the trivial divided power immersion $S \hookrightarrow S$ [BBM, 3.3.6]. Recall that $\mathbf{M}(G)$ is an $\mathscr{O}_S$-module locally free of finite type satisfying the following properties:
(i) Let $F_M : \mathbf{M}(G)^{(p)} \to \mathbf{M}(G)$ and $V_M : \mathbf{M}(G) \to \mathbf{M}(G)^{(p)}$ be the $\mathscr{O}_S$-linear maps induced respectively by the Frobenius and the Verschiebung of G. We have the following exact sequence:

$$\cdots \to \mathbf{M}(G)^{(p)} \xrightarrow{F_M} \mathbf{M}(G) \xrightarrow{V_M} \mathbf{M}(G)^{(p)} \to \cdots .$$

(ii) There is a connection $\nabla : \mathbf{M}(G) \to \mathbf{M}(G) \otimes_{\mathscr{O}_S} \Omega^1_{S/\mathbb{F}_p}$ for which F_M and V_M are horizontal morphisms.
(iii) We have two canonical filtrations on $\mathbf{M}(G)$ by $\mathscr{O}_S$-modules locally free of finite type:

$$0 \to \omega_G \to \mathbf{M}(G) \to \mathrm{Lie}(G^\vee) \to 0, \tag{3.1.1}$$

called the *Hodge filtration* on $\mathbf{M}(G)$ [BBM, 3.3.5], and the *conjugate filtration* on $\mathbf{M}(G)$

$$0 \to \mathrm{Lie}(G^\vee)^{(p)} \xrightarrow{\phi_G} \mathbf{M}(G) \to \omega_G^{(p)} \to 0, \tag{3.1.2}$$

which is obtained by applying the Dieudonné functor to the exact sequence of finite group schemes $0 \to \operatorname{Ker} F_G \to G(1) \to \operatorname{Ker} V_G \to 0$ [BBM, 4.3.1, 4.3.6, 4.3.11]. Moreover, we have the following commutative diagram (cf. [Ka1, 2.3.2

and 2.3.4])
(3.1.3)

$$\begin{tikzcd}
& 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
& \omega_G^{(p)} \arrow[d] & \omega_G \arrow[r, "\psi_G"] \arrow[d] & \omega_G^{(p)} \arrow[d] & \\
{} \arrow[r] & \mathbf{M}(G)^{(p)} \arrow[r, "F_M"] \arrow[d] & \mathbf{M}(G) \arrow[r, "V_M"] \arrow[d] \arrow[ur] & \mathbf{M}(G)^{(p)} \arrow[r] \arrow[d] & {,} \\
& \mathrm{Lie}(G^\vee)^{(p)} \arrow[r, "\widetilde{\varphi_G}"] \arrow[d] \arrow[ur, hook, "\phi_G"] & \mathrm{Lie}(G^\vee) \arrow[d] & \mathrm{Lie}(G^\vee)^{(p)} \arrow[d] & \\
& 0 & 0 & 0 &
\end{tikzcd}$$

where the columns are the Hodge filtrations and the anti-diagonal is the conjugate filtration. By functoriality, we see easily that $\widetilde{\varphi_G}$ above is nothing but the linearization of the Hasse-Witt map φ_G (2.6.1), and the morphism $\psi_G^* : \mathrm{Lie}(G)^{(p)} \to \mathrm{Lie}(G)$, which is obtained by applying the functor $\mathscr{H}om_{\mathscr{O}_S}(\_, \mathscr{O}_S)$ to ψ_G, is identified with the linearization $\widetilde{\varphi_{G^\vee}}$ of $\varphi_{G^\vee}$.
The formation of these structures on $\mathbf{M}(G)$ commutes with arbitrary base changes of S. In the sequel, we will use $(\mathbf{M}(G), F_M, \nabla)$ to emphasize these structures on $\mathbf{M}(G)$.

3.2. In the reminder of this section, k will denote an algebraically closed field of characteristic $p > 0$. Let S be a scheme formally smooth over k such that $\Omega^1_{S/\mathbb{F}_p} = \Omega^1_{S/k}$ is an $\mathscr{O}_S$-module locally free of finite type, *e.g.* $S = \mathrm{Spec}(A)$ with A a formally smooth k-algebra with a finite p-basis over k. Let G be a BT-group over S. We put KS to be the composed morphism

$$\text{(3.2.1)} \qquad \mathrm{KS} : \omega_G \to \mathbf{M}(G) \xrightarrow{\nabla} \mathbf{M}(G) \otimes_{\mathscr{O}_S} \Omega^1_{S/k} \xrightarrow{pr} \mathrm{Lie}(G^\vee) \otimes_{\mathscr{O}_S} \Omega^1_{S/k}$$

which is $\mathscr{O}_S$-linear. We put $\mathscr{T}_{S/k} = \mathscr{H}om_{\mathscr{O}_S}(\Omega^1_{S/k}, \mathscr{O}_S)$, and define the *Kodaira-Spencer map* of G

$$\text{(3.2.2)} \qquad \mathrm{Kod} : \mathscr{T}_{S/k} \to \mathscr{H}om_{\mathscr{O}_S}(\omega_G, \mathrm{Lie}(G^\vee))$$

to be the morphism induced by KS. We say that G is *versal* if Kod is surjective.

3.3. Let r be an integer ≥ 1, $R = k[[t_1, \cdots, t_r]]$, $\mathfrak{m}$ be the maximal ideal of R. We put $\mathscr{S} = \mathrm{Spf}(R)$, $S = \mathrm{Spec}(R)$, and for each integer $n \geq 0$, $S_n = \mathrm{Spec}(R/\mathfrak{m}^{n+1})$. By a BT-group $\mathscr{G}$ over the formal scheme $\mathscr{S}$, we mean a sequence of BT-groups $(G_n)_{n\geq 0}$ over $(S_n)_{n\geq 0}$ equipped with isomorphisms $G_{n+1} \times_{S_{n+1}} S_n \simeq G_n$.

According to [deJ, 2.4.4], *the functor* $G \mapsto (G\times_S S_n)_{n\geq 0}$ *induces an equivalence of categories between the category of BT-groups over* S *and the category of BT-groups over* $\mathscr{S}$. For a BT-group $\mathscr{G}$ over $\mathscr{S}$, the corresponding BT-group G over S is called the *algebraization* of $\mathscr{G}$. We say that $\mathscr{G}$ is *versal* over $\mathscr{S}$, if its algebraization G is versal over S. Since S is local, by Nakayama's Lemma, $\mathscr{G}$ or G is versal if and only if the reduction of Kod modulo the maximal ideal

$$\mathrm{Kod}_0 : \mathscr{T}_{S/k}\otimes_{\mathscr{O}_S} k \longrightarrow \mathrm{Hom}_k(\omega_{G_0}, \mathrm{Lie}(G_0^\vee)) \tag{3.3.1}$$

is surjective.

3.4. We recall briefly the deformation theory of a BT-group. Let $\mathfrak{AL}_k$ be the category of local artinian k-algebras with residue field k. We notice that all morphisms of $\mathfrak{AL}_k$ are local. A morphism $A'\to A$ in $\mathfrak{AL}_k$ is called a *small extension*, if it is surjective and its kernel I satisfies $I\cdot\mathfrak{m}_{A'}=0$, where $\mathfrak{m}_{A'}$ is the maximal ideal of A'.
Let G_0 be a BT-group over k, and A an object of $\mathfrak{AL}_k$. A deformation of G_0 over A is a pair (G,ϕ), where G is a BT-group over $\mathrm{Spec}(A)$ and ϕ is an isomorphism $\phi: G\otimes_A k \xrightarrow{\sim} G_0$. When there is no risk of confusions, we will denote a deformation (G,ϕ) simply by G. Two deformations (G,ϕ) and (G',ϕ') over A are isomorphic if there exists an isomorphism of BT-groups $\psi: G\xrightarrow{\sim} G'$ over A such that $\phi=\phi'\circ(\psi\otimes_A k)$. Let's denote by $\mathcal{D}$ the functor which associates with each object A of $\mathfrak{AL}_k$ the set of isomorphsm classes of deformations of G_0 over A. If $f: A\to B$ is a morphism of $\mathfrak{AL}_k$, then the map $\mathcal{D}(f):\mathcal{D}(A)\to\mathcal{D}(B)$ is given by extension of scalars. We call $\mathcal{D}$ the *deformation functor* of G_0 over $\mathfrak{AL}_k$.

PROPOSITION 3.5 ([Ill], 4.8). *Let* G_0 *be a BT-group over* k *of dimension* d *and height* $c+d$, $\mathcal{D}$ *be the deformation functor of* G_0 *over* $\mathfrak{AL}_k$.
(i) *Let* $A'\to A$ *be a small extension in* $\mathfrak{AL}_k$ *with ideal* I, $x=(G,\phi)$ *be an element in* $\mathcal{D}(A)$, $\mathcal{D}_x(A')$ *be the subset of* $\mathcal{D}(A')$ *with image* x *in* $\mathcal{D}(A)$. *Then the set* $\mathcal{D}_x(A')$ *is a nonempty homogenous space under the group* $\mathrm{Hom}_k(\omega_{G_0},\mathrm{Lie}(G_0^\vee))\otimes_k I$.
(ii) *The functor* $\mathcal{D}$ *is pro-representable by a formally smooth formal scheme* $\mathscr{S}$ *over* k *of relative dimension* cd, i.e. $\mathscr{S}=\mathrm{Spf}(R)$ *with* $R\simeq k[[(t_{ij})_{1\leq i\leq c,1\leq j\leq d}]]$, *and there exists a unique deformation* $(\mathscr{G},\psi)$ *of* G_0 *over* $\mathscr{S}$ *such that, for any object* A *of* $\mathfrak{AL}_k$ *and any deformation* (G,ϕ) *of* G_0 *over* A, *there is a unique homomorphism of local* k*-algebras* $\varphi: R\to A$ *with* $(G,\phi)=\mathcal{D}(\varphi)(\mathscr{G},\psi)$.
(iii) *Let* $\mathscr{T}_{\mathscr{S}/k}(0)=\mathscr{T}_{\mathscr{S}/k}\otimes_{\mathscr{O}_{\mathscr{S}}} k$ *be the tangent space of* $\mathscr{S}$ *at its unique closed point,*

$$\mathrm{Kod}_0:\mathscr{T}_{\mathscr{S}/k}(0)\longrightarrow \mathrm{Hom}_k(\omega_{G_0},\mathrm{Lie}(G_0^\vee))$$

be the Kodaira-Spencer map of $\mathscr{G}$ *evaluated at the closed point of* $\mathscr{S}$. *Then* Kod_0 *is bijective, and it can be described as follows. For an element* $f\in\mathscr{T}_{\mathscr{S}/k}(0)$, i.e. *a homomorphism of local* k*-algebras* $f: R\to k[\epsilon]/\epsilon^2$, $\mathrm{Kod}_0(f)$ *is the difference of deformations*

$$[\mathscr{G}\otimes_R (k[\epsilon]/\epsilon^2)] - [G_0\otimes_k (k[\epsilon]/\epsilon^2)],$$

which is a well-defined element in $\mathrm{Hom}_k(\omega_{G_0},\mathrm{Lie}(G_0^\vee))$ *by* (i).

Remark 3.6. Let $(e_j)_{1\leq j\leq d}$ be a basis of ω_{G_0}, $(f_i)_{1\leq i\leq c}$ be a basis of $\mathrm{Lie}(G_0^\vee)$. In view of 3.5(iii), we can choose a system of parameters $(t_{ij})_{1\leq i\leq c, 1\leq j\leq d}$ of $\mathscr{S}$ such that

$$\mathrm{Kod}_0(\frac{\partial}{\partial t_{ij}}) = e_j^* \otimes f_i,$$

where $(e_j^*)_{1\leq j\leq d}$ is the dual basis of $(e_j)_{1\leq j\leq d}$. Moreover, if $\mathfrak{m}$ is the maximal ideal of R, the parameters t_{ij} are determined uniquely modulo $\mathfrak{m}^2$.

Corollary 3.7 (Algebraization of the universal deformation). *The assumptions being those of* (3.5)*, we put moreover* $\mathbf{S} = \mathrm{Spec}(R)$ *and* $\mathbf{G}$ *the algebraization of the universal formal deformation* $\mathscr{G}$*. Then the BT-group* $\mathbf{G}$ *is versal over* $\mathbf{S}$*, and satisfies the following universal property: Let A be a noetherian complete local k-algebra with residue field k, G be a BT-group over A endowed with an isomorphism $G\otimes_A k \simeq G_0$. Then there exists a unique continuous homomorphism of local k-algebras $\varphi : R \to A$ such that* $G \simeq \mathbf{G}\otimes_R A$.

Proof. By the last remark of 3.3, $\mathbf{G}$ is clearly versal. It remains to prove that it satisfies the universal property in the corollary. Let G be a deformation of G_0 over a noetherian complete local k-algebra A with residue field k. We denote by $\mathfrak{m}_A$ the maximal ideal of A, and put $A_n = A/\mathfrak{m}_A^{n+1}$ for each integer $n \geq 0$. Then by 3.5(b), there exists a unique local homomorphism $\varphi_n : R \to A_n$ such that $G\otimes A_n \simeq \mathbf{G}\otimes_R A_n$. The φ_n's form a projective system $(\varphi_n)_{n\geq 0}$, whose projective limit $\varphi : R \to A$ answers the question. □

Definition 3.8. The notations are those of (3.7). We call $\mathbf{S}$ the *local moduli in characteristic* p of G_0, and $\mathbf{G}$ the *universal deformation of* G_0 *in characteristic* p.

If there is no confusions, we will omit "in characteristic p" for short.

3.9. Let G be a BT-group over k, G° be its connected part, and $G^{\text{ét}}$ be its étale part. Let r be the height of $G^{\text{ét}}$. Then we have $G^{\text{ét}} \simeq (\mathbb{Q}_p/\mathbb{Z}_p)^r$, since k is algebraically closed. Let $\mathcal{D}_G$ (*resp.* $\mathcal{D}_{G^\circ}$) be the deformation functor of G (*resp.* G°) over $\mathfrak{AL}_k$. If A is an object in $\mathfrak{AL}_k$ and $\mathscr{G}$ is a deformation of G (*resp.* G°) over A, we denote by $[\mathscr{G}]$ its isomorphism class in $\mathcal{D}_G(A)$ (*resp.* in $\mathcal{D}_{G^\circ(A)}$).

Proposition 3.10. *The assumptions are as above, let $\Theta : \mathcal{D}_G \to \mathcal{D}_{G^\circ}$ be the morphism of functors that maps a deformation of G to its connected component.*
(i) *The morphism Θ is formally smooth of relative dimension r.*
(ii) *Let A be an object of $\mathfrak{AL}_k$, and $\mathscr{G}^\circ$ be a deformation of G° over A. Then the subset $\Theta_A^{-1}([\mathscr{G}^\circ])$ of $\mathcal{D}_G(A)$ is canonically identified with $\mathrm{Ext}^1_A(\mathbb{Q}_p/\mathbb{Z}_p, \mathscr{G}^\circ)^r$, where Ext^1_A means the group of extensions in the category of abelian* fppf*-sheaves on* $\mathrm{Spec}(A)$.

Proof. (i) Since $\mathcal{D}_G$ and $\mathcal{D}_{G^\circ}$ are both pro-representable by a noetherian local complete k-algebra and formally smooth over k (3.5), by a formal completion version of [EGA, IV 17.11.1(d)], we only need to check that the tangent map

$$\Theta_{k[\epsilon]/\epsilon^2} : \mathcal{D}_G(k[\epsilon]/\epsilon^2) \to \mathcal{D}_{G^\circ}(k[\epsilon]/\epsilon^2)$$

is surjective with kernel of dimension r over k. By 3.5(iii), $\mathcal{D}_G(k[\epsilon]/\epsilon^2)$ (*resp.* $\mathcal{D}_{G^\circ}(k[\epsilon]/\epsilon^2)$) is isomorphic to $\mathrm{Hom}_k(\omega_G, \mathrm{Lie}(G^\vee))$ (*resp.* $\mathrm{Hom}_k(\omega_{G^\circ}, \mathrm{Lie}(G^{\circ\vee}))$) by the Kodaira-Spencer morphism. In view of the canonical isomorphism $\omega_G \simeq \omega_{G^\circ}$, $\Theta_{k[\epsilon]/\epsilon^2}$ corresponds to the map

$$\Theta'_{k[\epsilon]/\epsilon^2} : \mathrm{Hom}_k(\omega_G, \mathrm{Lie}(G^\vee)) \to \mathrm{Hom}_k(\omega_G, \mathrm{Lie}(G^{\circ\vee}))$$

induced by the canonical surjection $\mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G^{\circ\vee})$. It is clear that $\Theta'_{k[\epsilon]/\epsilon^2}$ is surjective of kernel $\mathrm{Hom}_k(\omega_G, \mathrm{Lie}(G^{\text{ét}\vee}))$, which has dimension r over k.

(ii) Since $G^{\text{ét}}$ is isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^r$, every element in $\mathrm{Ext}^1_A(\mathbb{Q}_p/\mathbb{Z}_p, \mathscr{G}^\circ)^r$ defines clearly an element of $\mathcal{D}_G(A)$ with image $[\mathscr{G}^\circ]$ in $\mathcal{D}_{G^\circ}(A)$. Conversely, for any $\mathscr{G} \in \mathcal{D}_G(A)$ with connected component isomorphic to $\mathscr{G}^\circ$, the isomorphism $G^{\text{ét}} \simeq (\mathbb{Q}_p/\mathbb{Z}_p)^r$ lifts uniquely to an isomorphism $\mathscr{G}^{\text{ét}} \simeq (\mathbb{Q}_p/\mathbb{Z}_p)^r$ because A is henselian. The canonical exact sequence $0 \to \mathscr{G}^\circ \to \mathscr{G} \to \mathscr{G}^{\text{ét}} \to 0$ shows that $\mathscr{G}$ comes from an element of $\mathrm{Ext}^1_A(\mathbb{Q}_p/\mathbb{Z}_p, \mathscr{G}^\circ)^r$.

□

4. HW-cyclic Barsotti-Tate Groups

DEFINITION 4.1. Let S be a scheme of characteristic $p > 0$, G be a BT-group over S such that $c = \dim(G^\vee)$ is constant. We say that G is *HW-cyclic*, if $c \geq 1$ and there exists an element $v \in \Gamma(S, \mathrm{Lie}(G^\vee))$ such that

$$v, \varphi_G(v), \cdots, \varphi_G^{c-1}(v)$$

generate $\mathrm{Lie}(G^\vee)$ as an $\mathscr{O}_S$-module, where φ_G is the Hasse-Witt map (2.6.1) of G.

REMARK 4.2. It is clear that a BT-group G over S is HW-cyclic, if and only if $\mathrm{Lie}(G^\vee)$ is free over $\mathscr{O}_S$ and there exists a basis of $\mathrm{Lie}(G^\vee)$ over $\mathscr{O}_S$ under which φ_G is expressed by a matrix of the form

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_c \end{pmatrix}, \tag{4.2.1}$$

where $a_i \in \Gamma(S, \mathscr{O}_S)$ for $1 \leq i \leq c$.

LEMMA 4.3. *Let R be a local ring of characteristic $p > 0$, k be its residue field.*
(i) *A BT-group G over R is HW-cyclic if and only if so is $G \otimes k$.*
(ii) *Let $0 \to G' \to G \to G'' \to 0$ be an exact sequence of BT-groups over R. If G is HW-cyclic, then so is G'. In particular, if R is henselian, the connected part of a HW-cyclic BT-group over R is HW-cyclic.*

Proof. (i) The property of being HW-cyclic is clearly stable under arbitrary base changes, so the "only if" part is clear. Assume that $G_0 = G \otimes k$ is HW-cyclic. Let $\overline{v}$ be an element of $\mathrm{Lie}(G_0^\vee) = \mathrm{Lie}(G^\vee) \otimes k$ such that

$(\overline{v}, \varphi_{G_0}(\overline{v}), \cdots, \varphi_{G_0}^{c-1}(\overline{v}))$ is a basis of $\mathrm{Lie}(G_0^\vee)$. Let v be any lift of $\overline{v}$ in $\mathrm{Lie}(G^\vee)$. Then by Nakayama's lemma, $(v, \varphi_G(v), \cdots, \varphi_G^{c-1}(v))$ is a basis of $\mathrm{Lie}(G^\vee)$.
(ii) By statement (i), we may assume $R = k$. The exact sequence of BT-groups induces an exact sequence of Lie algebras

$$0 \to \mathrm{Lie}(G''^\vee) \to \mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G'^\vee) \to 0, \tag{4.3.1}$$

and the Hasse-Witt map $\varphi_{G'}$ is induced by φ_G by functoriality. Assume that G is HW-cyclic and $G^\vee$ has dimension c. Let u be an element of $\mathrm{Lie}(G^\vee)$ such that

$$u, \varphi_G(u), \cdots, \varphi_G^{c-1}(u)$$

form a basis of $\mathrm{Lie}(G^\vee)$ over k. We denote by u' the image of u in $\mathrm{Lie}(G'^\vee)$. Let $r \leq c$ be the maximal integer such that the vectors

$$u', \ \varphi_{G'}(u'), \ \cdots, \ \varphi_{G'}^{r-1}(u')$$

are linearly independent over k. It is easy to see that they form a basis of the k-vector space $\mathrm{Lie}(G'^\vee)$. Hence G' is HW-cyclic. □

Lemma 4.4. *Let $S = \mathrm{Spec}(R)$ be an affine scheme of characteristic $p > 0$, G be a HW-cyclic BT-group over R with $c = \dim(G^\vee)$ constant, and*

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_c \end{pmatrix} \in \mathrm{M}_{c\times c}(R),$$

be a matrix of φ_G. Put $a_{c+1} = 1$, and $P(X) = \sum_{i=0}^{c} a_{i+1} X^{p^i} \in R[X]$.
(i) *Let $V_G : G^{(p)} \to G$ be the Verschiebung homomorphism of G. Then $\mathrm{Ker}\, V_G$ is isomorphic to the group scheme $\mathrm{Spec}(R[X]/P(X))$ with comultiplication given by $X \mapsto 1 \otimes X + X \otimes 1$.*
(ii) *Let $x \in S$, and G_x be the fibre of G at x. Put*

$$i_0(x) = \min_{0 \leq i \leq c} \{i; a_{i+1}(x) \neq 0\}, \tag{4.4.1}$$

where $a_i(x)$ denotes the image of a_i in the residue field of x. Then the étale part of G_x has height $c - i_0(x)$, and the connected part of G_x has height $d + i_0(x)$. In particular, G_x is connected if and only if $a_i(x) = 0$ for $1 \leq i \leq c$.

Proof. (i) By 2.3 and 2.13, $\mathrm{Ker}\, V_G$ is isomorphic to the group scheme

$$\mathrm{Spec}\Big(R[X_1, \ldots, X_c]/(X_1^p - X_2, \cdots, X_{c-1}^p - X_c, X_c^p + a_1 X_1 + \cdots + a_c X_c)\Big)$$

with comultiplication $\Delta(X_i) = 1 \otimes X_i + X_i \otimes 1$ for $1 \leq i \leq c$. By sending $(X_1, X_2, \cdots, X_c) \mapsto (X, X^p, \cdots, X^{p^{c-1}})$, we see that the above group scheme is isomorphic to $\mathrm{Spec}(R[X]/P(X))$ with comultiplication $\Delta(X) = 1 \otimes X + X \otimes 1$.

(ii) By base change, we may assume that $S = x = \mathrm{Spec}(k)$ and hence $G = G_x$. Let $G(1)$ be the kernel of the multiplication by p on G. Then we have an exact sequence

$$0 \to \operatorname{Ker} F_G \to G(1) \to \operatorname{Ker} V_G \to 0.$$

Since $\operatorname{Ker} F_G$ is an infinitesimal group scheme over k, we have $G(1)(\overline{k}) = (\operatorname{Ker} V_G)(\overline{k})$, where $\overline{k}$ is an algebraic closure of k. By the definition of $i_0(x)$, we have $P(X) = Q(X^{p^{i_0(x)}})$, where $Q(X)$ is an additive sepearable polynomial in $k[X]$ with $\deg(Q) = p^{c-i_0(x)}$. Hence the roots of $P(X)$ in $\overline{k}$ form an $\mathbb{F}_p$-vector space of dimension $c - i_0(x)$. By (i), $(\operatorname{Ker} V_G)(\overline{k})$ can be identified with the additive group consisting of the roots of $P(X)$ in $\overline{k}$. Therefore, the étale part of G has height $c - i_0(x)$, and the connected part of G has height $d + i_0(x)$. □

4.5. Let k be a perfect field of characteristic $p > 0$, and $\alpha_p = \mathrm{Spec}(k[X]/X^p)$ be the finite group scheme over k with comultiplication map $\Delta(X) = 1 \otimes X + X \otimes 1$. Let G be a BT-group over k. Following Oort, we call

$$a(G) = \dim_k \mathrm{Hom}_{k_{\mathrm{fppf}}}(\alpha_p, G)$$

the a-number of G, where $\mathrm{Hom}_{k_{\mathrm{fppf}}}$ means the homomorphisms in the category of abelian fppf-sheaves over k. Since the Frobenius of α_p vanishes, any morphism of α_p in G factorize through $\mathrm{Ker}(F_G)$. Therefore we have

$$\begin{aligned}\mathrm{Hom}_{k_{\mathrm{fppf}}}(\alpha_p, G) &= \mathrm{Hom}_{k-gr}(\alpha_p, \mathrm{Ker}(F_G)) \\ &= \mathrm{Hom}_{k-gr}(\mathrm{Ker}(F_G)^\vee, \alpha_p) \\ &= \mathrm{Hom}_{p\text{-}\mathfrak{Lie}_k}(\mathrm{Lie}(\alpha_p), \mathrm{Lie}(\mathrm{Ker}(F_G))),\end{aligned}$$

where Hom_{k-gr} denotes the homomorphisms in the category of commutative group schemes over k, and the last equality uses Proposition 2.3. Since we have a canonical isomorphism $\mathrm{Lie}(\mathrm{Ker}(F_G)) \simeq \mathrm{Lie}(G)$ and $\mathrm{Lie}(\alpha_p)$ has dimension one over k with $\varphi_{\alpha_p} = 0$, we get

$$a(G) = \dim_k\{x \in \mathrm{Lie}(G) | \varphi_{G^\vee}(x) = 0\} = \dim_k \mathrm{Ker}(\varphi_{G^\vee}). \tag{4.5.1}$$

Due to the perfectness of k, we have also $a(G) = \dim_k \mathrm{Ker}(\widetilde{\varphi_{G^\vee}})$, where $\widetilde{\varphi_{G^\vee}}$ is the linearization of $\varphi_{G^\vee}$. By Proposition 2.11, we see that $a(G) = 0$ if and only if G is ordinary.

LEMMA 4.6. *Let G be a BT-group over k, and $G^\vee$ its Serre dual. Then we have $a(G) = a(G^\vee)$.*

Proof. Let $\psi_G : \omega_G \to \omega_G^{(p)}$ be the k-linear map induced by the Verschiebung of G. Then ψ_G^*, the morphism obtained by applying the functor $\mathrm{Hom}_k(\_, k)$ to ψ_G, is identified with $\widetilde{\varphi_{G^\vee}}$. By (4.5.1) and the exactitude of the functor $\mathrm{Hom}_k(\_, k)$, we have $a(G) = \dim_k \mathrm{Ker}(\psi_G^*) = \dim_k \mathrm{Coker}(\psi_G)$. Using the additivity of $\dim_k$, we get finally $a(G) = \dim_k \mathrm{Ker}(\psi_G)$. By considering the commutative diagram (3.1.3), we have

$$a(G) = \dim_k\Big(\omega_G \cap \phi_G(\mathrm{Lie}(G^\vee)^{(p)})\Big).$$

On the other hand, it follows also from (3.1.3) that

$$a(G^\vee) = \dim_k \operatorname{Ker}(\widetilde{\varphi_G}) = \dim_k \Big(\phi_G(\operatorname{Lie}(G^\vee)^{(p)}) \cap \omega_G\Big).$$

The lemma now follows immediately.

□

Proposition 4.7. *Let k be a perfect field of characteristic $p > 0$, G a BT-group over k. Consider the following conditions:*
(i) *G is HW-cyclic and non-ordinary;*
(ii) *the connected part G° of G is HW-cyclic and not of multiplicative type;*
(iii) *$a(G^\vee) = a(G) = 1$.*
We have (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii). *If k is algebraically closed, we have moreover* (ii) $\Rightarrow$ (i).

Remark 4.8. In [Oo1, Lemma 2.2], Oort proved the following assertion, which is a generalization of (iii) $\Rightarrow$ (ii): Let k be an algebraically closed field of characteristic $p > 0$, and G be a connected BT-group with $a(G) = 1$. Then there exists a basis of the Dieudonné module M of G over $W(k)$, such that the action of Frobenius on M is given by a display-matrix of "normal form" in the sense of [Oo1, 2.1].

Proof. (i) $\Rightarrow$ (ii) follows from 4.3(ii).
(ii) $\Rightarrow$ (iii). First, we note that $a(G) = a(G^\circ)$, so we may assume G connected. Since G is not of multiplicative type, we have $c = \dim(G^\vee) \geq 1$. By Lemma 4.4(ii), there exists a basis of $\operatorname{Lie}(G^\vee)$ over k under which φ_G is expressed by

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \in \mathrm{M}_{c\times c}(k).$$

According to (4.5.1), $a(G^\vee)$ equals to $\dim_k \operatorname{Ker}(\varphi_G)$, *i.e.* the k-dimension of the solutions of the equation system in $(x_1, \cdots, x_c)$

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1^p \\ x_2^p \\ \vdots \\ x_c^p \end{pmatrix} = 0$$

The solutions $(x_1, \cdots, x_c)$ form clearly a vector space over k of dimension 1, *i.e.* we have $a(G^\vee) = 1$.
(iii) $\Rightarrow$ (ii). Let $G^{\text{ét}}$ be the étale part of G. Since k is perfect, the exact sequence (2.7.1) splits [Dem, Chap. II §7]; so we have $G \simeq G^\circ \times G^{\text{ét}}$. We put $M = \operatorname{Lie}(G^\vee)$, $M_1 = \operatorname{Lie}(G^{\circ\vee})$ and $M_2 = \operatorname{Lie}(G^{\text{ét}\vee})$ for short. By 2.8 and 2.9, we have a decomposition $M = M_1 \oplus M_2$, such that M_1, M_2 are stable under φ_G, and the action of φ_G is nilpotent on M_1 and bijective on M_2. We note

that $a(G^{\circ\vee}) = a(G^\circ) = a(G) = 1$. By the last remark of 4.5, G° is not of multiplicative type, hence $\dim_k M_1 = \dim(G^{\circ\vee}) \geq 1$. It remains to prove that G° is HW-cyclic. Let n be the minimal integer such that $\varphi_G^n(M_1) = 0$. We have a strictly increasing filtration

$$0 \subsetneq \mathrm{Ker}(\varphi_G) \subsetneq \cdots \subsetneq \mathrm{Ker}(\varphi_G^n) = M_1.$$

If $n = 1$, then M_1 is one-dimensional, hence G° is clearly HW-cyclic. Assume $n \geq 2$. For $2 \leq m \leq n$, φ_G^{m-1} induces an injective map

$$\overline{\varphi_G^{m-1}} : \mathrm{Ker}(\varphi_G^m)/\mathrm{Ker}(\varphi_G^{m-1}) \longrightarrow \mathrm{Ker}(\varphi_G).$$

Since $\dim_k \mathrm{Ker}(\varphi_G) = a(G^{\circ\vee}) = 1$, $\overline{\varphi_G^{m-1}}$ is necessarily bijective. So we have $\dim_k \mathrm{Ker}(\varphi_G^m) = m$ for $1 \leq m \leq n$. Let v be an element of M_1 but not in $\mathrm{Ker}(\varphi_G^{n-1})$. Then $v, \varphi_G(v), \cdots, \varphi_G^{n-1}(v)$ are linearly independant, hence they form a basis of M_1 over k. This proves that G° is HW-cyclic.
Assume k algebraically closed. We prove that (ii) $\Rightarrow$ (i). Noting that G is ordinary if and only if G° is of multiplicative type, we only need to check that G is HW-cyclic. We conserve the notations above. Since φ_G is bijective on M_2 and k algebraically closed, there exists a basis $(e_1, \cdots, e_m)$ of M_2 such that $\varphi_G(e_i) = e_i$ for $1 \leq i \leq m$. Let $v \in M_1$ but not in $\mathrm{Ker}(\varphi_G^{n-1})$ as above, and put $u = v + \lambda_1 e_1 + \cdots \lambda_m e_m$, where $\lambda_i (1 \leq i \leq m)$ are some elements in k to be determined later. Then we have

$$\begin{pmatrix} \varphi_G^n(u) \\ \vdots \\ \varphi_G^{n+m-1}(u) \end{pmatrix} = \begin{pmatrix} \lambda_1^{p^n} & \cdots & \lambda_m^{p^n} \\ \vdots & \ddots & \vdots \\ \lambda_1^{p^{n+m-1}} & \cdots & \lambda_m^{p^{n+m-1}} \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_m \end{pmatrix}.$$

Let $L(\lambda_1, \cdots, \lambda_m) \in k[\lambda_1, \cdots, \lambda_m]$ be the determinant polynomial of the matrix on the right side. An elementary computation shows that the polynomial $L(\lambda_1, \cdots, \lambda_m)$ is not null. We can choose $\lambda_1, \cdots, \lambda_m \in k$ such that $L(\lambda_1, \cdots, \lambda_m) \neq 0$ because k is algebraically closed. So $\varphi_G^n(u), \cdots, \varphi_G^{n+m-1}(u)$ form a basis of M_2 over k. Since

$$\varphi_G^i(u) \equiv \varphi_G^i(v) \mod M_2 \quad \text{for} \quad 0 \leq i \leq n,$$

by the choice of u, we see that $\{u, \varphi_G(u), \cdots, \varphi_G^{n+m-1}(u)\}$ form a basis of $M = \mathrm{Lie}(G^\vee)$ over k. □

By combining 4.6 and 4.7, we obtain the following

COROLLARY 4.9. *Let k be an algebraically closed field of characteristic $p > 0$. Then a BT-group over k is HW-cyclic if and only if so is its Serre dual.*

4.10. EXAMPLES. Let k be a perfect field, $W(k)$ be the ring of Witt vectors with coefficients in k, and σ be the Frobenius automorphism of $W(k)$. Let s, r be relatively prime integers such that $0 \leq s \leq r$ and $r \neq 0$; put $\lambda = \frac{s}{r}$. We consider the Dieudonné module $M^\lambda \simeq W(k)[F, V]/(F^{r-s} - V^s)$, where $W(k)[F, V]$ is the non-commutative ring with relations $FV = VF = p$, $Fa = \sigma(a)F$ and $V\sigma(a) = aV$ for all $a \in W(k)$. We note that M^λ is free of rank

r over $W(k)$ and $M^\lambda/VM^\lambda \simeq k[F]/F^{r-s}$. By the contravariant Dieudonné theory, M^λ corresponds to a BT-group G^λ over k of height r with $\mathrm{Lie}(G^{\lambda\vee}) = M^\lambda/VM^\lambda$. We see easily that G^λ is HW-cyclic, and we call it the *elementary BT-group of slope* λ. We note that $G^0 \simeq \mathbb{Q}_p/\mathbb{Z}_p$, $G^1 \simeq \mu_{p^\infty}$, and $(G^\lambda)^\vee \simeq G^{1-\lambda}$ for $0 \leq \lambda \leq 1$.
Assume k algebraically closed. Then by the Dieudonné-Manin's classification of isocrystals [Dem, Chap.IV §4], any BT-group over k is isogenous to a finite product of G^λ's; moreover, any connected one-dimensional BT-group over k of height r is necessarily isomorphic to $G^{1/r}$ [Dem, Chap.IV §8], hence in particular HW-cyclic.

Proposition 4.11. *Let k be an algebraically closed field of characteristic $p > 0$, R be a noetherian complete regular local k-algebra with residue field k, and $S = \mathrm{Spec}(R)$. Let G be a connected HW-cyclic BT-group over R of dimension $d \geq 1$ and height $c + d$,*

$$\mathfrak{h} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_c \end{pmatrix} \in \mathrm{M}_{c\times c}(R)$$

be a matrix of φ_G.
(i) *If G is versal over S, then $\{a_1, \cdots, a_c\}$ is a subset of a regular system of parameters of R.*
(ii) *Assume that $d = 1$. The converse of* (i) *is also true,* i.e. *if $\{a_1, \cdots, a_c\}$ is a subset of a regular system of parameters of R then G is versal over S. Furthermore, G is the universal deformation of its special fiber if and only if $\{a_1, \cdots, a_c\}$ is a system of regular parameters of R.*

Proof. Let $(\mathbf{M}(G), F_M, \nabla)$ be the finite free $\mathscr{O}_S$-module equipped with a semi-linear endomorphism F_M and a connection $\nabla : \mathbf{M}(G) \to \mathbf{M}(G) \otimes_{\mathscr{O}_S} \Omega^1_{S/k}$, obtained by evaluating the Dieudonné crystal of G at the trivial immersion $S \hookrightarrow S$ (cf. 3.1). Recall that we have a commutative diagram

(4.11.1)
$$\begin{tikzcd} \mathbf{M}(G)^{(p)} \arrow[r, "F_M"] \arrow[d, "pr"'] & \mathbf{M}(G) \arrow[d, "pr"] \\ \mathrm{Lie}(G^\vee)^{(p)} \arrow[r, "\widetilde{\varphi_G}"] \arrow[ur, hook, "\phi_G"] & \mathrm{Lie}(G^\vee), \end{tikzcd}$$

where ϕ_G is universally injective (3.1.3). Let $\{v_1, \cdots, v_c\}$ be a basis of $\mathrm{Lie}(G^\vee)$ over $\mathscr{O}_S$ under which φ_G is expressed by $\mathfrak{h}$, *i.e.* we have $\varphi_G^{i-1}(v_1) = v_i$ for $1 \leq i \leq c$ and $\varphi_G^c(v_1) = \varphi_G(v_c) = -\sum_{i=1}^c a_i v_i$. Let f_1 be a lift of v_1 to $\Gamma(S, \mathbf{M}(G))$, and put $f_{i+1} = \phi_G(v_i^{(p)})$ for $1 \leq i \leq c-1$, where $v_i^{(p)} = 1 \otimes v_i \in \Gamma(S, \mathrm{Lie}(G^\vee)^{(p)})$. The image of f_i in $\Gamma(S, \mathrm{Lie}(G^\vee))$ is thus v_i for $1 \leq i \leq c$ by

(4.11.1). We put

$$e_1 = \phi_G(v_c^{(p)}) + a_1 f_1 + \cdots + a_c f_c \in \Gamma(S, \mathbf{M}(G)). \tag{4.11.2}$$

The image of e_1 in $\Gamma(S, \mathrm{Lie}(G^\vee))$ is $\varphi_G(v_c) + \sum_{i=1}^c a_i v_i = 0$; so we have $e_1 \in \Gamma(S, \omega_G)$. By 4.4(ii), we notice that $a_1, \cdots, a_c$ belong to the maximal ideal $\mathfrak{m}_R$ of R, as G is connected. Hence, we have $\overline{e_1} = \overline{\phi_G(v_c^{(p)})}$, where for a R-module M and $x \in M$, we denote by $\overline{x}$ the canonical image of x in $M \otimes k$. Since ϕ_G commutes with base change and is universally injective, we get $\overline{e_1} = \overline{\phi_G(v_c^{(p)})} = \phi_{G\otimes k}(\overline{v_c^{(p)}}) \neq 0$. Therefore, we can choose $e_2, \cdots, e_d \in \Gamma(S, \omega_G)$ such that $(e_1, \cdots, e_d)$ becomes a basis of ω_G over $\mathscr{O}_S$, so $(e_1, \cdots, e_d, f_1, \cdots, f_c)$ is a basis of $\mathbf{M}(G)$. Since F_M is horizontal for the connection ∇ (cf. 3.1(ii)), we have

$$\nabla(\phi_G(v_c^{(p)})) = \nabla(F_M(f_c^{(p)})) = 0.$$

In view of (4.11.2), we get

$$\begin{aligned} \nabla(e_1) &= \sum_{i=1}^c f_i \otimes da_i + \sum_{i=1}^c a_i \nabla(f_i) \\ &\equiv \sum_{i=1}^c f_i \otimes da_i \pmod{\mathfrak{m}_R}. \end{aligned} \tag{4.11.3}$$

Let KS_0 and Kod_0 be respectively the reductions modulo $\mathfrak{m}_R$ of (3.2.1) and (3.2.2). Since $(\overline{v_i})_{1\leq i\leq c}$ is a base of $\mathrm{Lie}(G^\vee) \otimes k$, we can write

$$\mathrm{KS}_0(e_j) = \sum_{i=1}^c \overline{v_i} \otimes \theta_{i,j} \qquad \text{for } 1 \leq j \leq d,$$

where $\theta_{i,j} \in \Omega_{S/k} \otimes k$. From (4.11.3), we deduce that $\theta_{i,1} = da_i$. By the definition of Kod_0, we have

$$\mathrm{Kod}_0(\partial) = \sum_{j=1}^d \sum_{i=1}^c < \partial, \theta_{i,j} > \overline{e_j}^* \otimes \overline{v_i} \tag{4.11.4}$$

where $\partial \in \mathscr{T}_{S/k} \otimes k$, $< \bullet, \bullet >$ is the canonical pairing between $\mathscr{T}_{S/k} \otimes k$ and $\Omega^1_{S/k} \otimes k$, and $(\overline{e_i}^*)_{1\leq i\leq d}$ denotes the dual basis of $(\overline{e_i})_{1\leq i\leq d}$. Now assume that G is versal over S, *i.e.* Kod_0 is surjective by definition (3.2). In particular, there are $\partial_1, \cdots, \partial_c \in \mathscr{T}_{S/k} \otimes k$ such that $\mathrm{Kod}_0(\partial_i) = \overline{e_1}^* \otimes v_i$ for $1 \leq i \leq c$, *i.e.* we have

$$< \partial_i, da_j > = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \text{for } 1 \leq i, j \leq c, \tag{4.11.5}$$

and

$$< \partial_i, \theta_{j,\ell} > = 0 \qquad \text{for } 1 \leq i, j \leq c, 2 \leq \ell \leq d.$$

From (4.11.5), we see easily that $da_1, \cdots, da_c$ are linearly independent in $\Omega_{S/k} \otimes k \simeq \mathfrak{m}_R/\mathfrak{m}_R^2$; therefore, $(a_1, \cdots, a_c)$ is a part of a regular system of parameters of R. Statement (i) is proved.

For statement (ii), we assume $d = 1$ and that $(a_1, \cdots, a_c)$ is a part of a regular system of parameters of R. Then the formula (4.11.4) is simplified as

$$\mathrm{Kod}_0(\partial) = \sum_{i=1}^{c} < \partial, da_i > \overline{e_1}^* \otimes \overline{v_i}.$$

Since $da_1, \cdots, da_c$ are linearly independent in $\Omega^1_{S/k} \otimes k$, there exist $\partial_1, \cdots, \partial_c \in \mathscr{T}_{S/k} \otimes k$ such that (4.11.5) holds, *i.e.* $(\overline{e_1}^* \otimes \overline{v_i})_{1 \leq i \leq c}$ are in the image of Kod_0. But the elements $(\overline{e_1}^* \otimes \overline{v_i})_{1 \leq i \leq c}$ form already a basis of $\mathscr{H}om_{\mathscr{O}_S}(\omega_G, \mathrm{Lie}(G^\vee)) \otimes k$. So Kod_0 is surjective, and hence G is versal over S by Nakayama's lemma. Let G_0 be the special fiber of G. It remains to prove that when $d = 1$, G is the universal deformation of G_0 if and only if $\dim(S) = c$ and G is versal over S. Let $\mathbf{S}$ be the local moduli in characteristic p of G_0. By the universal property of $\mathbf{G}$ (3.7), there exists a unique morphism $f : S \to \mathbf{S}$ such that $G \simeq \mathbf{G} \times_{\mathbf{S}} S$. Since S and $\mathbf{S}$ are local complete regular schemes over k with residue field k of the same dimension, f is an isomorphism if and only if the tangent map of f at the closed point of S, denoted by T_f, is an isomorphism. By the functoriality of Kodaira-Spencer maps (3.2.2), we have a commutative diagram

$$\begin{tikzcd}
\mathscr{T}_{S/k} \otimes_{\mathscr{O}_S} k \arrow[r, "\mathrm{Kod}_0^S"] \arrow[d, "T_f"'] & \mathrm{Hom}_k(\omega_{G_0}, \mathrm{Lie}(G_0^\vee)) \arrow[d, equal] \\
\mathscr{T}_{\mathbf{S}/k} \otimes_{\mathscr{O}_{\mathbf{S}}} k \arrow[r, "\mathrm{Kod}_0^{\mathbf{S}}"] & \mathrm{Hom}_k(\omega_{G_0}, \mathrm{Lie}(G_0^\vee))
\end{tikzcd},$$

where horizontal arrows are the Kodaira-Spencer maps evaluated at the closed points (3.3.1). Since Kod_0^S and $\mathrm{Kod}_0^{\mathbf{S}}$ are isomorphisms according to the first part of this propostion, we deduce that so is T_f. This completes the proof. $\square$

5. Monodromy of a HW-cyclic BT-group over a Complete Trait of Characteristic $p > 0$

5.1. Let k be an algebraically closed field of characteristic $p > 0$, A be a complete discrete valuation ring of characteristic p, with residue field k and fraction field K. We put $S = \mathrm{Spec}(A)$, and denote by s its closed point, by η its generic point. Let $\overline{K}$ be an algebraic closure of K, K^{sep} be the maximal separable extension of K contained in $\overline{K}$, K^{t} be the maximal tamely ramified extension of K contained in K^{sep}. We put $I = \mathrm{Gal}(K^{\mathrm{sep}}/K)$, $I_p = \mathrm{Gal}(K^{\mathrm{sep}}/K^{\mathrm{t}})$ and $I_t = I/I_p = \mathrm{Gal}(K^{\mathrm{t}}/K)$.
Let π be a uniformizer of A; so we have $A \simeq k[[\pi]]$. Let $\mathtt{v}$ be the valuation on K normalized by $\mathtt{v}(\pi) = 1$; we denote also by $\mathtt{v}$ the unique extension of $\mathtt{v}$ to $\overline{K}$. For every $\alpha \in \mathbb{Q}$, we denote by $\mathfrak{m}_\alpha$ (*resp.* by $\mathfrak{m}_\alpha^+$) the set of elements $x \in K^{\mathrm{sep}}$ such that $\mathtt{v}(x) \geq \alpha$ (*resp.* $\mathtt{v}(x) > \alpha$). We put

$$V_\alpha = \mathfrak{m}_\alpha / \mathfrak{m}_\alpha^+, \tag{5.1.1}$$

which is a k-vector space of dimension 1 equipped with a continuous action of the Galois group I.

5.2. First, we recall some properties of the inertia groups I_p and I_t [Se1, Chap. IV]. The subgroup I_p, called the *wild inertia subgroup*, is the unique maximal pro-p-group contained in I and hence normal in I. The quotient $I_t = I/I_p$ is a commutative profinite group, called the *tame inertia group*. We have a canonical isomorphism

$$\theta : I_t \xrightarrow{\sim} \varprojlim_{(d,p)=1} \mu_d, \tag{5.2.1}$$

where the projective system is taken over positive integers prime to p, μ_d is the group of d-th roots of unity in k, and the transition maps $\mu_m \to \mu_d$ are given by $\zeta \mapsto \zeta^{m/d}$, whenever d divides m. We denote by $\theta_d : I_t \to \mu_d$ the projection induced by (5.2.1). Let q be a power of p, $\mathbb{F}_q$ be the finite subfield of k with q elements. Then $\mu_{q-1} = \mathbb{F}_q^\times$, and we can write $\theta_{q-1} : I_t \to \mathbb{F}_q^\times$. The character θ_d is characterized by the following property.

PROPOSITION 5.3 ([Se3] Prop.7). *Let a, d be relatively prime positive integers with d prime to p. Then the natural action of I_p on the k-vector space $V_{a/d}$ (5.1.1) is trivial, and the induced action of I_t on $V_{a/d}$ is given by the character $(\theta_d)^a : I_t \to \mu_d$. In particular, if q is a power of p, the action of I_t on $V_{1/(q-1)}$ is given by the character $\theta_{q-1} : I_t \to \mathbb{F}_q^\times$ and any I-equivariant $\mathbb{F}_p$-subspace of $V_{1/(q-1)}$ is an $\mathbb{F}_q$-vector space.*

5.4. Let G be a BT-group over S. We define $h(G)$ to be the valuation of the determinant of a matrix of φ_G if $\dim(G^\vee) \geq 1$, and $h(G) = 0$ if $\dim(G^\vee) = 0$. We call $h(G)$ the *Hasse invariant* of G.
(a) $h(G)$ does not depend on the choice of the matrix representing φ_G. Indeed, let c be the rank of $\mathrm{Lie}(G^\vee)$ over A, $\mathfrak{h} \in \mathrm{M}_{c\times c}(A)$ be a matrix of φ_G. Any other matrix representing φ_G can be written in the form $U^{-1} \cdot \mathfrak{h} \cdot U^{(p)}$, where $U \in \mathrm{GL}_c(A)$, U^{-1} is the inverse of U, and $U^{(p)}$ is the matrix obtained by applying the Frobenius map of A to the coefficients of U.
(b) By 2.11, the generic fiber G_η is ordinary if and only if $h(G) < \infty$; G is ordinary over T if and only $h(G) = 0$.
(c) Let $0 \to G' \to G \to G'' \to 0$ be a short exact sequence of BT-groups over T, then we have $h(G) = h(G') + h(G'')$. Indeed, the exact sequence of BT-groups induces a short exact sequence of Lie algebras (cf. [BBM] 3.3.2)

$$0 \to \mathrm{Lie}(G''^\vee) \to \mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G'^\vee) \to 0,$$

from which our assertion follows easily.

PROPOSITION 5.5. *Let G be a BT-group over S. Then we have $h(G) = h(G^\vee)$.*

Proof. The proof is very similar to that of Lemma 4.6. First, we have

$$h(G) = \mathrm{leng}\big(\mathrm{Lie}(G^\vee)/\widetilde{\varphi_G}(\mathrm{Lie}(G^\vee)^{(p)})\big),$$

where $\widetilde{\varphi_G}$ is the linearization of φ_G, and "leng" means the length of a finite A-module (note that this formulae holds even if $\dim(G^\vee) = 0$). By the commutative diagram (3.1.3), we have

$$h(G) = \mathrm{leng}\,\mathbf{M}(G)/(\phi_G(\mathrm{Lie}(G^\vee)^{(p)}) + \omega_G).$$

On the other hand, by applying the functor $\mathrm{Hom}_A(\_, A)$ to the A-linear map $\widetilde{\varphi_{G^\vee}} : \mathrm{Lie}(G)^{(p)} \to \mathrm{Lie}(G)$, we obtain a map $\psi_G : \omega_G \to \omega_G^{(p)}$. If U is a matrix of $\widetilde{\varphi_{G^\vee}}$, then the transpose of U, denoted by U^t, is a matrix of ψ_G. So we have

$$h(G^\vee) = \mathtt{v}(\det(U)) = \mathtt{v}(\det(U^t)) = \mathrm{leng}\big(\omega_G^{(p)}/\psi_G(\omega_G)\big).$$

By diagram 3.1.3, we get

$$h(G^\vee) = \mathrm{leng}\,\mathbf{M}(G)/(\phi_G(\mathrm{Lie}(G^\vee)^{(p)}) + \omega_G) = h(G).$$

□

5.6. Let G be a BT-group over S, $c = \dim(G^\vee)$. We put

$$\mathrm{T}_p(G) = \varprojlim_n G(n)(\overline{K}) \tag{5.6.1}$$

the Tate module of G, where $G(n)$ is the kernel of $p^n : G \to G$. It is a free $\mathbb{Z}_p$-module of rank $\leq c$, and the equality holds if and only if the generic fiber G_η is ordinary. The Galois group I acts continuously on $\mathrm{T}_p(G)$. We are interested in the image of the monodromy representation

$$\rho : I = \mathrm{Gal}(K^{\mathrm{sep}}/K) \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(G)). \tag{5.6.2}$$

We denote by

$$\overline{\rho} : I = \mathrm{Gal}(K^{\mathrm{sep}}/K) \to \mathrm{Aut}_{\mathbb{F}_p}\big(G(1)(\overline{K})\big) \tag{5.6.3}$$

its reduction mod p.

Theorem 5.7 (Reformulation of Igusa's theorem). *Let G be a connected BT-group over S of height 2 and dimension 1. Then G is versal (3.2) if and only if $h(G) = 1$; moreover, if this condition is satisfied, the monodromy representation $\rho : I \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(G)) \simeq \mathbb{Z}_p^\times$ is surjective.*

Proof. Since $\mathrm{Lie}(G^\vee)$ is an $\mathscr{O}_S$-module free of rank 1, the condition that $h(G) = 1$ is equivalent to that any matrix of φ_G is represented by a uniformizer of A. Hence the first part of this theorem follows from Proposition 4.11(ii).
We follow [Ka2, Thm 4.3] to prove the surjectivity of ρ under the assumption that $h(G) = 1$. For each integer $n \geq 1$, let

$$\rho_n : I \to \mathrm{Aut}_{\mathbb{Z}/p^n\mathbb{Z}}(G(n)(\overline{K})) \simeq (\mathbb{Z}/p^n\mathbb{Z})^\times$$

be the reduction mod p^n of ρ, K_n be the subfield of K^{sep} fixed by the kernel of ρ_n. Then ρ_n induces an injective homomorphism $\mathrm{Gal}(K_n/K) \to (\mathbb{Z}/p^n\mathbb{Z})^\times$. By taking projective limits, we are reduced to proving the surjectivity of ρ_n for every $n \geq 1$. It suffices to verify that

$$|\,\mathrm{Im}(\rho_n)| = [K_n : K] \geq p^{n-1}(p-1)$$

(then the equality holds automatically).

We regard G as a formal group over S. Then by [Ka2, 3.6], there exists a parameter X of the formal group G normalized by the condition that $[\xi](X) = \xi(X)$ for all $(p-1)$-th root of unity $\xi \in \mathbb{Z}_p$. For such a parameter, we have

$$[p](X) = a_1X^p + \alpha X^{p^2} + \sum_{m\geq 2} c_m X^{p(1+m(p-1))} \in A[[X]],$$

where we have $\mathtt{v}(a_1) = h(G) = 1$ by [Ka2, 3.6.1 and 3.6.5], and $\mathtt{v}(\alpha) = 0$, as G is of height 2. For each integer $i \geq 0$, we put

$$V^{(p^i)}(X) = a_1^{p^i} X + \alpha^{p^i} X^p + \sum_{m\geq 2} c_m^{p^i} X^{1+m(p-1)} \in A[[X]];$$

then we have $[p^n](X) = V^{(p^{n-1})} \circ V^{(p^{n-2})} \circ \cdots \circ V(X^{p^n})$. Hence each point of $G(n)(\overline{K})$ is given by a sequence $y_1, \cdots, y_n \in K^{\mathrm{sep}}$ (or simply an element $y_n \in K^{\mathrm{sep}}$) satisfying the equations

$$\begin{cases} V(y_1) = a_1y_1 + \alpha y_1^p + \cdots = 0; \\ V^{(p)}(y_2) = a_1^p y_2 + \alpha^p y_2^p + \cdots = y_1; \\ \vdots \\ V^{(p^{n-1})}(y_n) = a_1^{p^{n-1}} y_n + \alpha^{p^{n-1}} y_n^p + \cdots = y_{n-1}. \end{cases}$$

Let $y_n \in K^{\mathrm{sep}}$ be such that $y_1 \neq 0$. By considering the Newton polygons of the equations above, we verify that

$$\mathtt{v}(y_i) = \frac{1}{p^{i-1}(p-1)} \qquad \text{for } 1 \leq i \leq n.$$

In particular, the ramification index $e(K_n/K)$ is at least $p^{n-1}(p-1)$. By the definition of K_n, the Galois group $\mathrm{Gal}(K^{\mathrm{sep}}/K_n)$ must fix $y_n \in K^{\mathrm{sep}}$, *i.e.* K_n is an extension of $K(y_n)$. Therefore, we have $[K_n : K] \geq [K(y_n) : K] \geq e(K(y_n)/K) \geq p^{n-1}(p-1)$. $\square$

PROPOSITION 5.8. *Let G be a HW-cyclic BT-group over S of height $c+d$ and dimension d such that $G \otimes K$ is ordinary,*

$$\mathfrak{h} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_c \end{pmatrix}$$

be a matrix of φ_G. Put $q = p^c$, $a_{c+1} = 1$, and $P(X) = \sum_{i=0}^{c} a_{i+1} X^{p^i} \in A[X]$.
(i) *Assume that G is connected and the Hasse invariant $h(G) = 1$. Then the representation $\overline{\rho}$ (5.6.3) is tame, $G(1)(\overline{K})$ is endowed with the structure of an $\mathbb{F}_q$-vector space of dimension 1, and the induced action of I_t is given by the character $\theta_{q-1} : I_t \to \mathbb{F}_q^{\times}$.*
(ii) *Assume that $c > 1$, $\mathtt{v}(a_i) \geq 2$ for $1 \leq i \leq c-1$ and $\mathtt{v}(a_c) = 1$. Then the order of $\mathrm{Im}(\overline{\rho})$ is divisible by $p^{c-1}(p-1)$.*

(iii) *Put* $i_0 = \min_{0\leq i\leq c}\{i; \mathtt{v}(a_{i+1}) = 0\}$. *Assume that there exists* $\alpha \in k$ *such that* $\mathtt{v}(P(\alpha)) = 1$. *Then we have* $i_0 \leq c-1$ *and the order of* $\mathrm{Im}(\overline{\rho})$ *is divisible by* p^{i_0}.

Proof. Since G is generically ordinary, we have $a_1 \neq 0$ by 2.11(d). Hence $P(X) \in K[X]$ is a separable polynomial. By 4.4, $G(1)(\overline{K}) \simeq (\mathrm{Ker}\, V_G)(K^{\mathrm{sep}})$ is identified with the additive group consisting of the roots of $P(X)$ in K^{sep}.
(i) By definition of the Hasse invariant, we have $\mathtt{v}(a_1) = h(G) = 1$. By 4.4(ii), the assumption that G is connected is equivalent to saying $\mathtt{v}(a_i) \geq 1$ for $1 \leq i \leq c$. From the Newton polygon of $P(X)$, we deduce that all the non-zero roots of $P(X)$ in K^{sep} have the same valuation $1/(q-1)$. We denote by

$$\psi : G(1)(\overline{K}) \to V_{1/(q-1)}$$

the map which sends each root $x \in K^{\mathrm{sep}}$ of $P(X)$ to the class of x in $V_{1/(q-1)} = \mathfrak{m}_{1/(q-1)}/\mathfrak{m}^+_{1/(q-1)}$ (5.1.1). We remark that $G(1)(\overline{K})$ is an $\mathbb{F}_p$-vector space of dimension c. Hence $G(1)(\overline{K})$ is automatically of dimension 1 over $\mathbb{F}_q$ once we know it is an $\mathbb{F}_q$-vector space. By 5.3, it suffices to show that ψ is an injective I-equivariant homomorphism of groups. By 4.4(i), ψ is obviously an I-equivariant homomorphism of groups. Let x_0 be a root of $P(X)$, and put $Q(y) = P(x_0 y)$. Then the polynomial $Q(y)$ has the form $Q(y) = x_0^q Q_1(y)$, where

$$Q_1(y) = y^q + b_c y^{p^{c-1}} + \cdots + b_2 y^p + b_1 y$$

with $b_i = a_i/x_0^{(q-p^{i-1})} \in K^{\mathrm{sep}}$. We have $\mathtt{v}(b_i) > 0$ for $2 \leq i \leq c$ and $\mathtt{v}(b_1) = 0$. Let $\overline{b}_1$ be the class of b_1 in the residue field $k = \mathfrak{m}_0/\mathfrak{m}_0^+$. Then the images of the roots of $P(X)$ in $V_{1/(q-1)}$ are $x_0 \overline{b}_1^{1/(q-1)}\zeta$, where ζ runs over the finite field $\mathbb{F}_q$. Therefore, ψ is injective.
(ii) By computing the slopes of the Newton polygon of $P(X)$, we see that $P(X)$ has $p^{c-1}(p-1)$ roots of valuation $1/(p^c - p^{c-1})$. Let L be the sub-extension of K^{sep} obtained by adding to K all the roots of $P(x)$. Then the ramification index $e(L/K)$ is divisible by $p^{c-1}(p-1)$. Let $\widetilde{L}$ be the sub-extension of K^{sep} fixed by the kernel of $\overline{\rho}$ (5.6.3). The Galois group $\mathrm{Gal}(K^{\mathrm{sep}}/\widetilde{L})$ fixes the roots of $P(x)$ by definition. Hence we have $L \subset \widetilde{L}$, and $|\,\mathrm{Im}(\overline{\rho})| = [\widetilde{L} : K]$ is divisible by $[L : K]$; in particular, it is divisible by $p^{c-1}(p-1)$.
(iii) Note that the relation $i_0 \leq c-1$ is equivalent to saying that G is not connected by 4.4(ii). Assume conversely $i_0 = c$, *i.e.* G is connected. Then we would have

$$P(X) \equiv X^q \mod (\pi A[X]).$$

But $\mathtt{v}(P(\alpha)) = 1$ implies that $\alpha^{p^c} \in \pi A$, *i.e.* $\alpha = 0$; hence we would have $P(\alpha) = 0$, which contradicts the condition $\mathtt{v}(P(\alpha)) = 1$.
We put $Q(X) = P(X+\alpha) = P(X) + P(\alpha)$. As $\mathtt{v}(P(\alpha)) = 1$, then $(0,1)$ and $(p^{i_0}, 0)$ are the first two break points of the Newton polygon of $Q(X)$. Hence there exists p^{i_0} roots of $Q(X)$ of valuation $1/p^{i_0}$. Let L be the subextension of K in K^{sep} generated by the roots of $P(X)$. The ramification index $e(L/K)$ is divisible by p^{i_0}. As in the proof of (ii), if $\widetilde{L}$ is the subextension of K^{sep}

fixed by the kernel of $\overline{\rho}$, then it is an extension of L. Therefore, we have $|\operatorname{Im}(\overline{\rho})| = [\widetilde{L} : K]$ is divisible by $[L : K]$, and in particular, divisible by p^{i_0}. $\square$

5.9. Let G be a BT-group over S with connected part G°, and étale part $G^{\text{ét}}$ of height r. We have a canonical exact sequence of I-modules

$$0 \to G^\circ(1)(\overline{K}) \to G(1)(\overline{K}) \to G^{\text{ét}}(1)(\overline{K}) \to 0 \tag{5.9.1}$$

giving rise to a class $\overline{C} \in \operatorname{Ext}^1_{\mathbb{F}_p[I]}(G^{\text{ét}}(1)(\overline{K}), G^\circ(1)(\overline{K}))$, which vanishes if and only if (5.9.1) splits. Since I acts trivially on $G^{\text{ét}}(1)(\overline{K})$, we have an isomorphism of I-modules $G^{\text{ét}}(1)(\overline{K}) \simeq \mathbb{F}_p^r$. Recall that for any $\mathbb{F}_p[I]$-module M, we have a canonical isomorphism ([Se1] Chap.VII, §2)

$$\operatorname{Ext}^1_{\mathbb{F}_p[I]}(\mathbb{F}_p, M) \simeq H^1(I, M).$$

Hence we deduce that

$$\overline{C} \in \operatorname{Ext}^1_{\mathbb{F}_p[I]}(G^{\text{ét}}(1)(\overline{K}), G^\circ(1)(\overline{K})) \simeq H^1(I, G^\circ(1)(\overline{K}))^r. \tag{5.9.2}$$

PROPOSITION 5.10. *Let G be a HW-cyclic BT-group over S such that $h(G) = 1$, $\overline{\rho}$ (5.6.3) be the representation of I on $G(1)(\overline{K})$. Then the cohomology class $\overline{C}$ does not vanish if and only if the order of the group $\operatorname{Im}(\overline{\rho})$ is divisible by p.*

First, we prove the following result on cohomology of groups.

LEMMA 5.11. *Let F be a field, Γ be a commutative group, and $\chi : \Gamma \to F^\times$ be a non-trivial character of Γ. We denote by $F(\chi)$ an F-vector space of dimension 1 endowed with an action of Γ given by χ. Then we have $H^1(\Gamma, F(\chi)) = 0$.*

Proof. Let C be a 1-cocycle of Γ with values in $F(\chi)$. We prove that C is a 1-coboundary. For any $g, h \in \Gamma$, we have

$$\begin{aligned} C(gh) &= C(g) + \chi(g)C(h), \\ C(hg) &= C(h) + \chi(h)C(g). \end{aligned}$$

Since Γ is commutative, it follows from the relation $C(gh) = C(hg)$ that

$$(\chi(g) - 1)C(h) = (\chi(h) - 1)C(g). \tag{5.11.1}$$

If $\chi(g) \neq 1$ and $\chi(h) \neq 1$, then

$$\frac{1}{\chi(g) - 1} C(g) = \frac{1}{\chi(h) - 1} C(h).$$

Therefore, there exists $x \in F(\overline{\chi})$ such that $C(g) = (\chi(g) - 1)x$ for all $g \in \Gamma$ with $\chi(g) \neq 1$. If $\chi(g) = 1$, we have also $C(g) = 0 = (\chi(g) - 1)x$ by (5.11.1). This shows that C is a 1-coboundary. $\square$

Proof of 5.10. By 4.3(ii) and 5.4(c), the connected part G° of G is HW-cyclic with $h(G^\circ) = h(G) = 1$. Assume that $\mathrm{T}_p(G^\circ)$ has rank ℓ over $\mathbb{Z}_p$, and $\mathrm{T}_p(G^{\text{ét}})$ has rank r. Then by 5.8(a), $G^\circ(1)(\overline{K})$ is an $\mathbb{F}_q$-vector space of dimension 1 with $q = p^\ell$, and the action of I on $G^\circ(1)(\overline{K})$ factors through the character $\overline{\chi} : I \to I_t \xrightarrow{\theta_{q-1}} \mathbb{F}_q^\times$. We write $G^\circ(1)(\overline{K}) = \mathbb{F}_q(\overline{\chi})$ for short. If the cohomology class $\overline{C}$ is zero, then the exact sequence (5.9.1) splits, *i.e.* we have an isomorphism

of Galois modules $G(1)(\overline{K}) \simeq \mathbb{F}_q(\chi) \oplus \mathbb{F}_p^r$. It is clear that the group $\mathrm{Im}(\overline{\rho})$ has order $q-1$.
Conversely, if the cohomology class $\overline{C}$ is not zero, we will show that there exists an element in $\mathrm{Im}(\overline{\rho})$ of order p. We choose a basis adapted to the exact sequence (5.9.1) such that the action of $g \in I$ is given by

$$\overline{\rho}(g) = \begin{pmatrix} \overline{\chi}(g) & \overline{C}(g) \\ 0 & \mathbf{1}_r \end{pmatrix}, \tag{5.11.2}$$

where $\mathbf{1}_r$ is the unit matrix of type (r,r) with coefficients in $\mathbb{F}_p$, and the map $g \mapsto \overline{C}(g)$ gives rise to a 1-cocycle representing the cohomology class $\overline{C}$. Let I_1 be the kernel of $\overline{\chi} : I \to \mathbb{F}_q^\times$, Γ be the quotient I/I_1, so $\overline{\chi}$ induces an isomorphism $\overline{\chi} : \Gamma \xrightarrow{\sim} \mathbb{F}_q^\times$. We have an exact sequence

$$0 \to H^1(\Gamma, \mathbb{F}_q(\overline{\chi}))^r \xrightarrow{\mathrm{Inf}} H^1(I, \mathbb{F}_q(\overline{\chi}))^r \xrightarrow{\mathrm{Res}} H^1(I_1, \mathbb{F}_q(\overline{\chi}))^r,$$

where "Inf" and "Res" are respectively the inflation and restriction homomorphisms in group cohomology. Since $H^1(\Gamma, \mathbb{F}_q(\overline{\chi}))^r = 0$ by 5.11, the restriction of the cohomology class $\overline{C}$ to $H^1(I_1, \mathbb{F}_q(\overline{\chi}))^r$ is non-zero. Hence there exists $h \in I_1$ such that $\overline{C}(h) \neq 0$. As we have $\overline{\chi}(h) = 1$, then

$$\overline{\rho}(h)^p = \begin{pmatrix} \mathbf{1}_\ell & p\overline{C}(h) \\ 0 & \mathbf{1}_r \end{pmatrix} = \mathbf{1}_{\ell+r}.$$

Thus the order of $\overline{\rho}(h)$ is p. □

Corollary 5.12. *Let G be a HW-cyclic BT-group over S,*

$$\mathfrak{h} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_c \end{pmatrix}$$

be a matrix of φ_G, $P(X) = X^{p^c} + a_c X^{p^{c-1}} + \cdots + a_1 X \in A[X]$. If $h(G) = 1$ and if there exists $\alpha \in k \subset A$ such that $\mathrm{v}(P(\alpha)) = 1$, then the cohomology class (5.9.2) *is not zero,* i.e. *the extension of I-modules* (5.9.1) *does not split.*

Proof. Since $\mathrm{v}(a_1) = h(G) = 1$, the integer i_0 defined in 5.8(iii) is at least 1. Then the corollary follows from 5.8(iii) and 5.10. □

6. Lemmas in Group Theory

In this section, we fix a prime number $p \geq 2$ and an integer $n \geq 1$.

6.1. Recall that the general linear group $\mathrm{GL}_n(\mathbb{Z}_p)$ admits a natural exhaustive decreasing filtration by normal subgroups

$$\mathrm{GL}_n(\mathbb{Z}_p) \supset 1 + p\mathrm{M}_n(\mathbb{Z}_p) \supset \cdots \supset 1 + p^m\mathrm{M}_n(\mathbb{Z}_p) \supset \cdots,$$

where $\mathrm{M}_n(\mathbb{Z}_p)$ denotes the ring of matrix of type (n,n) with coefficients in $\mathbb{Z}_p$. We endow $\mathrm{GL}_n(\mathbb{Z}_p)$ with the topology for which $(1 + p^m\mathrm{M}_n(\mathbb{Z}_p))_{m\geq 1}$ form a fundamental system of neighborhoods of 1. Then $\mathrm{GL}_n(\mathbb{Z}_p)$ is a complete and separated topological group.

6.2. Let $\mathfrak{G}$ be a profinite group, $\rho : \mathfrak{G} \to \mathrm{GL}_n(\mathbb{Z}_p)$ be a continuous homomorphism of topological groups. By taking inverse images, we obtain a decreasing filtration $(F^m\mathfrak{G}, m \in \mathbb{Z}_{\geq 0})$ on $\mathfrak{G}$ by open normal subgroups:

$$F^0\mathfrak{G} = \mathfrak{G}, \quad \text{and} \quad F^m\mathfrak{G} = \rho^{-1}(1 + p^m\mathrm{M}_n(\mathbb{Z}_p)) \text{ for } m \geq 1.$$

Furthermore, the homomorphism ρ induces a sequence of injective homomorphisms of finite groups

$$\rho_0 \colon F^0\mathfrak{G}/F^1\mathfrak{G} \longrightarrow \mathrm{GL}_n(\mathbb{F}_p) \tag{6.2.1}$$

$$\rho_m \colon F^m\mathfrak{G}/F^{m+1}\mathfrak{G} \to \mathrm{M}_n(\mathbb{F}_p), \quad \text{for } m \geq 1. \tag{6.2.2}$$

LEMMA 6.3. *The homomorphism ρ is surjective if and only if the following conditions are satisfied:*
(i) *The homomorphism ρ_0 is surjective.*
(ii) *For every integer $m \geq 1$, the subgroup $\mathrm{Im}(\rho_m)$ of $\mathrm{M}_n(\mathbb{F}_p)$ contains an element of the form*

$$\begin{pmatrix} x & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

with $x \neq 0$; or equivalently, there exists, for every $m \geq 1$, an element $g_m \in \mathfrak{G}$ such that $\rho(g_m)$ is of the form

$$\begin{pmatrix} 1 + p^m a_{1,1} & p^{m+1}a_{1,2} & \cdots & p^{m+1}a_{1,n} \\ p^{m+1}a_{2,1} & 1 + p^{m+1}a_{2,2} & \cdots & p^{m+1}a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p^{m+1}a_{n,1} & p^{m+1}a_{n,2} & \cdots & 1 + p^{m+1}a_{n,n} \end{pmatrix},$$

where $a_{i,j} \in \mathbb{Z}_p$ for $1 \leq i, j \leq n$ and $a_{1,1}$ is not divisible by p.

Proof. We notice first that ρ is surjective if and only if ρ_m is surjective for every $m \geq 0$, because $\mathfrak{G}$ is complete and $\mathrm{GL}_n(\mathbb{Z}_p)$ is separated [Bou, Chap. III §2 n°8 Cor.2 au Théo. 1]. The surjectivity of ρ_0 is condition (i). Condition (ii) is clearly necessary. We prove that it implies the surjectivity of ρ_m for all $m \geq 1$, under the assumption of (i). First, we remark that under condition (i), if A lies in $\mathrm{Im}(\rho_m)$, then for any $U \in \mathrm{GL}_n(\mathbb{F}_p)$ the conjuagate matrix $U \cdot A \cdot U^{-1}$ lies also in $\mathrm{Im}(\rho_m)$. In fact, let $\widetilde{A}$ be a lift of A in $\mathrm{M}_n(\mathbb{Z}_p)$ and $\widetilde{U} \in \mathrm{GL}_n(\mathbb{Z}_p)$ a lift of U. By assumption, there exist $g, h \in \mathfrak{G}$ such that

$$\rho(g) \equiv 1 + p^m\widetilde{A} \mod (1 + p^{m+1}M_n(\mathbb{Z}_p)) \text{ and } \rho(h) \equiv \widetilde{U} \mod (1 + p\mathrm{M}_n(\mathbb{Z}_p)).$$

Therefore, we have $\rho(hgh^{-1}) \equiv (1 + p^m\widetilde{U} \cdot \widetilde{A} \cdot \widetilde{U}^{-1}) \mod (1 + p^{m+1}\mathrm{M}_n(\mathbb{Z}_p))$. Hence $hgh^{-1} \in F^m\mathfrak{G}$ and $\rho_m(hgh^{-1}) = U \cdot A \cdot U^{-1}$.
For $1 \leq i, j \leq n$, let $E_{i,j} \in \mathrm{M}_n(\mathbb{F}_p)$ be the matrix whose (i, j)-th entry is 0 and the other entries are 0. The matrices $E_{i,j} (1 \leq i, j \leq n)$ form clearly

a basis of $\mathrm{M}_n(\mathbb{F}_p)$ over $\mathbb{F}_p$. To prove the surjectivity of ρ_m, we only need to verify that $E_{i,j} \in \mathrm{Im}(\rho_m)$ for $1 \leq i,j \leq n$, because $\mathrm{Im}(\rho_m)$ is an $\mathbb{F}_p$-subspace of $\mathrm{M}_n(\mathbb{F}_p)$. By assumption, we have $E_{1,1} \in \mathrm{Im}(\rho_m)$. For $2 \leq i \leq n$, we put $U_i = E_{1,i} - E_{i,1} + \sum_{j \neq 1,i} E_{j,j}$. Then we have $U_i \in \mathrm{GL}_n(\mathbb{Z}_p)$ and $U_i \cdot E_{1,1} \cdot U_i^{-1} = E_{i,i} \in \mathrm{Im}(\rho_m)$. For $1 \leq i < j \leq n$, we put $U_{i,j} = I + E_{i,j}$ where I is the unit matrix. Then we have $U_{i,j} \cdot E_{i,i} \cdot U_{i,j}^{-1} = E_{i,i} + E_{i,j} \in \mathrm{Im}(\rho_m)$, and hence $E_{i,j} \in \mathrm{Im}(\rho_m)$. This completes the proof. □

Remark 6.4. By using the arguments in [Se2, Chap. IV 3.4 Lemma 3], we can prove the following stronger form of Lemma 6.3: *If $p = 2$, condition* (i) *and* (ii) *for $m = 1, 2$ are sufficient to guarantee the surjectivity of ρ; if $p \geq 3$, then* (i) *and* (ii) *just for $m = 1$ suffice already.*

A subgroup C of $\mathrm{GL}_n(\mathbb{F}_p)$ is called a *non-split Cartan subgroup*, if the subset $C \cup \{0\}$ of the matrix algebra $\mathrm{M}_n(\mathbb{F}_p)$ is a field isomorphic to $\mathbb{F}_{p^n}$; such a group is cyclic of order $p^n - 1$.

Lemma 6.5. *Assume that $n \geq 2$. We denote by H the subgroup of $\mathrm{GL}_n(\mathbb{F}_p)$ consisting of all the elements of the form $\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}$, where $A \in \mathrm{GL}_{n-1}(\mathbb{F}_p)$ and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_{n-1} \end{pmatrix}$ with $b_i \in \mathbb{F}_p (1 \leq i \leq n-1)$. Let G be a subgroup of $\mathrm{GL}_n(\mathbb{F}_p)$. Then $G = \mathrm{GL}_n(\mathbb{F}_p)$ if and only if G contains H and a non-split Cartan subgroup of $\mathrm{GL}_n(\mathbb{F}_p)$.*

Proof. The "only if" part is clear. For the "if" part, let C be a non-split Cartan subgroup contained in G. For a finite group Λ, we denote by $|\Lambda|$ its order. An easy computation shows that $|\mathrm{GL}_n(\mathbb{F}_p)| = |H| \cdot |C|$. So we just need to prove that $U \cap C = \{1\}$; since then we will have $|\mathrm{GL}_n(\mathbb{F}_p)| = |G|$, hence $G = \mathrm{GL}_n(\mathbb{F}_p)$. Let $g \in H \cap C$, and $P(T) \in \mathbb{F}_p[T]$ be its characteristic polynomial. We fix an isomorphism $C \simeq \mathbb{F}_{p^n}^\times$, and let $\zeta \in \mathbb{F}_{p^n}^\times$ be the element corresponding to g. We have $P(T) = \prod_{\sigma \in \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)} (T - \sigma(\zeta))$ in $\mathbb{F}_{p^n}[T]$. On the other hand, the fact that $g \in H$ implies that $(T-1)$ divises $P(T)$. Therefore, we get $\zeta = 1$, *i.e.* $g = 1$. □

Remark 6.6. E. Lau point out the following strengthened version of 6.5: *When $n \geq 3$, a subgroup $G \subset \mathrm{GL}_n(\mathbb{F}_p)$ coincides with $\mathrm{GL}_n(\mathbb{F}_p)$ if and only if G contains a non-split Cartan subgroup and the subgroup $\begin{pmatrix} \mathrm{GL}_{n-1}(\mathbb{F}_p) & 0 \\ 0 & 1 \end{pmatrix}$.* This can be used to simplify the induction process in the proof of Theorem 7.3 when $n \geq 3$.

7. Proof of Theorem 1.3 in the One-dimensional Case

7.1. We start with a general remark on the monodromy of BT-groups. Let X be a scheme, G be an ordinary BT-group over a scheme X, $G^{\text{ét}}$ be its étale part (2.10.1). If $\overline{\eta}$ is a geometric point of X, we denote by

$$\mathrm{T}_p(G,\overline{\eta}) = \varprojlim_n G(n)(\overline{\eta}) = \varprojlim_n G^{\text{ét}}(n)(\overline{\eta})$$

the Tate module of G at $\overline{\eta}$, and by $\rho(G)$ the monodromy representation of $\pi_1(X,\overline{\eta})$ on $\mathrm{T}_p(G,\overline{\eta})$. Let $f: Y \to X$ be a morphism of schemes, $\overline{\xi}$ be a geometric point of Y, $G_Y = G \times_X Y$. Then by the functoriality, we have a commutative diagram

$$\begin{tikzcd} \pi_1(Y,\xi) \arrow[r, "\pi_1(f)"] \arrow[d, "\rho(G_Y)"'] & \pi_1(X, f(\overline{\xi})) \arrow[d, "\rho(G)"] \\ \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(G_Y,\overline{\xi})) \arrow[r, equal] & \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(G, f(\overline{\xi}))) \end{tikzcd} \tag{7.1.1}$$

In particular, the monodromy of G_Y is a subgroup of the monodromy of G. In the sequel, diagram (7.1.1) will be refereed as the *functoriality of monodromy* for the BT-group G and the morphism f.

7.2. Let k be an algebraically closed field of characteristic $p > 0$, G be the unique connected BT-group over k of dimension 1 and height $n+1 \geq 2$ (4.10). We denote by $\mathbf{S}$ the algebraic local moduli of G in characteristic p, by $\mathbf{G}$ the universal deformation of G over $\mathbf{S}$, and by $\mathbf{U}$ the ordinary locus of $\mathbf{G}$ over $\mathbf{S}$ (3.8). Recall that $\mathbf{S}$ is affine of ring $R \simeq k[[t_1,\cdots,t_n]]$ (3.7), and that G and $\mathbf{G}$ are HW-cyclic (cf. 4.3(i) and 4.10). Let $\overline{\eta}$ be a geometric point of $\mathbf{U}$ over its generic point. We put

$$\mathrm{T}_p(\mathbf{G},\overline{\eta}) = \varprojlim_{m\in\mathbb{Z}_{\geq 1}} \mathbf{G}(m)(\overline{\eta})$$

to be the Tate module of $\mathbf{G}$ at the point $\overline{\eta}$. This is a free $\mathbb{Z}_p$-module of rank n. We have the monodromy representation

$$\rho_n : \pi_1(\mathbf{U},\overline{\eta}) \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(\mathbf{G},\overline{\eta})) \simeq \mathrm{GL}_n(\mathbb{Z}_p).$$

The following is the one-dimensional case of Theorem 1.3.

Theorem 7.3. *Under the above assumptions, the homomorphism ρ_n is surjective for $n \geq 1$.*

7.4. First, we assume $n \geq 2$. By Proposition 4.11(ii), we may assume that

$$\mathfrak{h} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -t_1 \\ 1 & 0 & \cdots & 0 & -t_2 \\ 0 & 1 & \cdots & 0 & -t_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -t_n \end{pmatrix} \tag{7.4.1}$$

is a matrix of the Hasse-Witt map $\varphi_{\mathbf{G}}$. Let $\mathfrak{p}$ be the prime ideal of R generated by $t_1, \cdots, t_{n-1}$. Then the closed subscheme of $\mathbf{S}$ defined by $\mathfrak{p}$ is just the locus where the p-rank of $\mathbf{G}$ is ≤ 1 by 4.4(ii). Let $K_0 \simeq k((t_n))$ be the fraction field of $R/\mathfrak{p}$, $R' = \widehat{R}_{\mathfrak{p}}$ be the completion of the localization of R at $\mathfrak{p}$, and $\mathscr{G}_{R'} = \mathbf{G} \otimes_R R'$. Since the natural map $R \to R'$ is injective, for any $a \in R$, we will denote also by a its image in R'. Since the Hasse-Witt map commutes with base change, the image of $\mathfrak{h}$ in $\mathrm{M}_{n\times n}(R')$, denoted also by $\mathfrak{h}$, is a matrix of $\varphi_{\mathscr{G}_{R'}}$. We see easily that the étale part of $\mathscr{G}_{R'}$ has height 1 and its connected part $\mathscr{G}^\circ_{R'}$ has height n. We have an exact sequence of BT-groups over R'

$$0 \to \mathscr{G}^\circ_{R'} \to \mathscr{G}_{R'} \to \mathscr{G}^{\text{ét}}_{R'} \to 0. \tag{7.4.2}$$

We fix an imbedding $i : K_0 \to \overline{K}_0$ of K_0 into an algebraically closed field. Put $\mathscr{G}^*_{\overline{K}_0} = \mathscr{G}^*_{R'} \otimes \overline{K}_0$ for $* = \emptyset, \text{ét}, \circ$. We have $\mathscr{G}^{\text{ét}}_{\overline{K}_0} \simeq \mathbb{Q}_p/\mathbb{Z}_p$, and $\mathscr{G}^\circ_{\overline{K}_0}$ is the unique connected one-dimensional BT-group over $\overline{K}_0$ of height n (cf. 4.10). We put $\widetilde{R'} = \overline{K}_0[[x_1, \cdots, x_{n-1}]]$, and

$$\Sigma = \{\text{ring homomorphisms } \sigma : R' \to \widetilde{R'} \text{ lifting } R' \to K_0 \xrightarrow{i} \overline{K}_0\} \tag{7.4.3}$$

Let $\sigma \in \Sigma$. We deduce from (7.4.2) by base change an exact sequence of BT-groups over $\widetilde{R'}$

$$0 \to \mathscr{G}^\circ_{\widetilde{R'},\sigma} \to \mathscr{G}_{\widetilde{R'},\sigma} \to \mathscr{G}^{\text{ét}}_{\widetilde{R'},\sigma} \to 0, \tag{7.4.4}$$

where we have put $\mathscr{G}^*_{\widetilde{R'},\sigma} = \mathscr{G}^*_{R'} \otimes_\sigma \widetilde{R'}$ for $* = \circ, \emptyset, \text{ét}$. Due to the henselian property of $\widetilde{R'}$, the isomorphism $\mathscr{G}^{\text{ét}}_{\overline{K}_0} \simeq \mathbb{Q}_p/\mathbb{Z}_p$ lifts uniquely to an isomorphism $\mathscr{G}^{\text{ét}}_{\widetilde{R'},\sigma} \simeq \mathbb{Q}_p/\mathbb{Z}_p$. Assume that $\mathscr{G}^\circ_{\widetilde{R'},\sigma}$ is generically ordinary over $\widetilde{S}' = \mathrm{Spec}(\widetilde{R'})$. Let $\widetilde{U}'_\sigma \subset \widetilde{S}'$ be its ordinary locus, and $\overline{x}$ be a geometric point over the generic point of $\widetilde{U}'_\sigma$. The exact sequence (7.4.4) induces an exact sequence of Tate modules

$$0 \to \mathrm{T}_p(\mathscr{G}^\circ_{\widetilde{R'},\sigma}, \overline{x}) \to \mathrm{T}_p(\mathscr{G}_{\widetilde{R'},\sigma}, \overline{x}) \to \mathrm{T}_p(\mathscr{G}^{\text{ét}}_{\widetilde{R'},\sigma}, \overline{x}) \to 0 \tag{7.4.5}$$

compatible with the actions of $\pi_1(\widetilde{U}'_\sigma, \overline{x})$. Since we have $\mathrm{T}_p(\mathscr{G}^{\text{ét}}_{\widetilde{R'},\sigma}, \overline{x}) \simeq \mathrm{T}_p(\mathbb{Q}_p/\mathbb{Z}_p, \overline{x}) = \mathbb{Z}_p$, this determines a cohomology class

$$C_\sigma \in \mathrm{Ext}^1_{\mathbb{Z}_p[\pi_1(\widetilde{U}'_\sigma, \overline{x})]}(\mathbb{Z}_p, \mathrm{T}_p(\mathscr{G}^\circ_{\widetilde{R'},\sigma}, \overline{x})) \simeq H^1(\pi_1(\widetilde{U}'_\sigma, \overline{x}), \mathrm{T}_p(\mathscr{G}^\circ_{\widetilde{R'},\sigma}, \overline{x})). \tag{7.4.6}$$

We consider also the "mod-p version" of (7.4.5)

$$0 \to \mathscr{G}^\circ_{\widetilde{R'},\sigma}(1)(\overline{x}) \to \mathscr{G}_{\widetilde{R'},\sigma}(1)(\overline{x}) \to \mathbb{F}_p \to 0,$$

which determines a cohomology class

$$\overline{C}_\sigma \in \mathrm{Ext}^1_{\mathbb{F}_p[\pi_1(\widetilde{U}'_\sigma, \overline{x})]}(\mathbb{F}_p, \mathscr{G}^\circ_{\widetilde{R'},\sigma}(1)(\overline{x})) \simeq H^1(\pi_1(\widetilde{U}'_\sigma, \overline{x}), \mathscr{G}^\circ_{\widetilde{R'},\sigma}(1)(\overline{x})). \tag{7.4.7}$$

It is clear that $\overline{C}_\sigma$ is the image of C_σ by the canonical reduction map

$$H^1(\pi_1(\widetilde{U}'_\sigma, \overline{x}), \mathrm{T}_p(\mathscr{G}^\circ_{\widetilde{R'},\sigma}, \overline{x})) \to H^1(\pi_1(\widetilde{U}'_\sigma, \overline{x}), \mathscr{G}^\circ_{\widetilde{R'},\sigma}(1)(\overline{x})).$$

LEMMA 7.5. *Under the above assumptions, there exist $\sigma_1, \sigma_2 \in \Sigma$ satisfying the following properties:*
(i) *We have $\mathscr{G}^{\circ}_{\widetilde{R'},\sigma_1} = \mathscr{G}^{\circ}_{\widetilde{R'},\sigma_2}$, and it is the universal deformation of $\mathscr{G}^{\circ}_{\overline{K}_0}$.*
(ii) *We have $C_{\sigma_1} = 0$ and $\overline{C}_{\sigma_2} \neq 0$.*

Before proving this lemma, we prove first Theorem 7.3.

PROOF OF 7.3. First, we notice that the monodromy of a BT-group is independent of the base point. So we can change $\overline{\eta}$ to any geometric point of $\mathbf{U}$ when discussing the monodromy of $\mathbf{G}$. We make an induction on the codimension $n = \dim(G^{\vee})$. The case of $n = 1$ is proved in Theorem 5.7. Assume that $n \geq 2$ and the theorem is proved for $n-1$. We denote by

$$\overline{\rho}_n : \pi_1(\mathbf{U}, \overline{\eta}) \to \mathrm{Aut}_{\mathbb{F}_p}(\mathbf{G}(1)(\overline{\eta})) \simeq \mathrm{GL}_n(\mathbb{F}_p)$$

the reduction of ρ_n modulo by p. By Lemma 6.3 and 6.5, to prove the surjectivity of ρ_n, we only need to verify the following conditions:
(a) $\mathrm{Im}(\overline{\rho}_n)$ contains a non-split Cartan subgroup of $\mathrm{GL}_n(\mathbb{F}_p)$;
(b) $\mathrm{Im}(\rho_n)$ contains the subgroup $H \subset \mathrm{GL}_n(\mathbb{Z}_p)$ consisting of all the elements of the form $\begin{pmatrix} B & b \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_n(\mathbb{Z}_p)$, with $B \in \mathrm{GL}_{n-1}(\mathbb{Z}_p)$ and $b \in \mathrm{M}_{(n-1)\times 1}(\mathbb{Z}_p)$;
For condition (a), let $A = k[[\pi]]$, $T = \mathrm{Spec}(A)$, ξ be its generic point, $\overline{\xi}$ be a geometric point over ξ, and $I = \mathrm{Gal}(\overline{\xi}/\xi)$ be the absolute Galois group over ξ. We keep the notations of 7.4. Let $f^* : R \to A$ be the homomorphism of k-algebras such that $f^*(t_1) = \pi$ and $f^*(t_i) = 0$ for $2 \leq i \leq n$. We denote by $f : T \to \mathbf{S}$ the corresponding morphism of schemes, and put $G_T = \mathbf{G} \times_{\mathbf{S}} T$. By the functoriality of Hasse-Witt maps,

$$\mathfrak{h}_T = \begin{pmatrix} 0 & 0 & \cdots & 0 & -\pi \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

is a matrix of φ_{G_T}. By definition 5.4, the Hasse invariant of G_T is $h(G) = 1$. Hence G_T is generically ordinary; so $f(\xi) \in \mathbf{U}$. Let

$$\overline{\rho}_T : I = \mathrm{Gal}(\overline{\xi}/\xi) \to \mathrm{Aut}_{\mathbb{F}_p}(G_T(1)(\overline{\xi}))$$

be the mod-p monodromy representation attached to G_T. Proposition 5.8(i) implies that $\mathrm{Im}(\overline{\rho}_T)$ is a non-split Cartan subgroup of $\mathrm{GL}_n(\mathbb{F}_p)$. On the other hand, by the functoriality of monodromy, we get $\mathrm{Im}(\overline{\rho}_T) \subset \mathrm{Im}(\overline{\rho}_n)$. This verifies condition (a).
To check condition (b), we consider the constructions in 7.4. Let $S' = \mathrm{Spec}(R')$, $f : S' \to \mathbf{S}$ be the morphism of schemes corresponding to the natural ring homomorphism $R \to R'$, U' be the ordinary locus of $\mathscr{G}_{R'}$, and $\overline{\xi}$ be a geometric point of U'. From (7.4.2), we deduce an exact sequence of Tate modules

$$0 \to \mathrm{T}_p(\mathscr{G}^{\circ}_{R'}, \overline{\xi}) \to \mathrm{T}_p(\mathscr{G}_{R'}, \overline{\xi}) \to \mathrm{T}_p(\mathscr{G}^{\text{ét}}_{R'}, \overline{\xi}) \to 0. \tag{7.5.1}$$

Let $\rho_{\mathscr{G}'} : \pi_1(U', \overline{\xi}) \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(\mathscr{G}_{R'}, \overline{\xi})) \simeq \mathrm{GL}_n(\mathbb{Z}_p)$ be the monodromy represention of $\mathscr{G}_{R'}$. Under any basis of $\mathrm{T}_p(\mathscr{G}_{R'}, \overline{\xi})$ adapted to (7.5.1), the action of $\pi_1(U', \overline{\xi})$ on $\mathrm{T}_p(\mathscr{G}_{R'}, \overline{\xi})$ is given by

$$\rho_{\mathscr{G}_{R'}} : g \in \pi_1(U', \overline{\xi}) \mapsto \begin{pmatrix} \rho_{\mathscr{G}^{\circ}_{R'}}(g) & * \\ 0 & \rho_{\mathscr{G}^{\text{ét}}_{R'}}(g), \end{pmatrix}$$

where $g \mapsto \rho_{\mathscr{G}^{\circ}_{R'}}(g) \in \mathrm{GL}_{n-1}(\mathbb{Z}_p)$ (*resp.* $g \mapsto \rho_{\mathscr{G}^{\text{ét}}_{R'}}(g) \in \mathbb{Z}_p^{\times}$) gives the action of $\pi_1(U', \overline{\xi})$ on $\mathrm{T}_p(\mathscr{G}^{\circ}_{R'}, \overline{\xi})$ (*resp.* on $\mathrm{T}_p(\mathscr{G}^{\text{ét}}_{R'}, \overline{\xi})$). Note that $f(U') \subset \mathbf{U}$. So by the functoriality of monodromy, we get $\mathrm{Im}(\rho_{\mathscr{G}'}) \subset \mathrm{Im}(\rho_n)$. To complete the proof of Theorem 7.3, it suffices to check condition (b) with ρ_n replaced by $\rho_{\mathscr{G}_{R'}}$ under the induction hypothesis that 7.3 is valide for $n-1$. Let $\sigma_1, \sigma_2 : R' \to \widetilde{R'}$ be the homomorphisms given by 7.5. For $i = 1, 2$, we denote by $f_i : \widetilde{S}' = \mathrm{Spec}(\widetilde{R'}) \to S' = \mathrm{Spec}(R')$ the morphism of schemes corresponding to σ_i, and put $\mathscr{G}_i = \mathscr{G}_{\widetilde{R'}, \sigma_i} = \mathscr{G}_{R'} \otimes_{\sigma_i} \widetilde{R'}$ to simply the notations. By condition 7.5(i), we can denote by $\mathscr{G}^{\circ}$ the common connected component of $\mathscr{G}_1$ and $\mathscr{G}_2$. Let $\widetilde{U'} \subset \widetilde{S}'$ be the ordinary locus of $\mathscr{G}^{\circ}$. Then we have $f_i(\widetilde{U'}) \subset U'$ for $i = 1, 2$. Let $\overline{x}$ be a geometric point over the generic point of $\widetilde{U'}$. We have an exact sequence of Tate modules

$$0 \to \mathrm{T}_p(\mathscr{G}^{\circ}, \overline{x}) \to \mathrm{T}_p(\mathscr{G}_i, \overline{x}) \to \mathrm{T}_p(\mathbb{Q}_p/\mathbb{Z}_p, \overline{x}) \to 0 \tag{7.5.2}$$

compatible with the actions of $\pi_1(\widetilde{U'}, \overline{x})$. We denote by

$$\rho_{\mathscr{G}_i} : \pi_1(\widetilde{U'}, \overline{x}) \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(\mathscr{G}_i, \overline{x})) \simeq \mathrm{GL}_n(\mathbb{Z}_p)$$

the monodromy representation of $\mathscr{G}_i$. In a basis adapted to (7.5.2), the action of $\pi_1(\widetilde{U'}, \overline{x})$ on $\mathrm{T}_p(\mathscr{G}_i, \overline{x})$ is given by

$$\rho_{\mathscr{G}_i} : g \mapsto \begin{pmatrix} \rho_{\mathscr{G}^{\circ}}(g) & C_{\sigma_i}(g) \\ 0 & 1 \end{pmatrix},$$

where $\rho_{\mathscr{G}^{\circ}} : \pi_1(\widetilde{U'}, \overline{x}) \to \mathrm{GL}_{n-1}(\mathbb{Z}_p)$ is the monodromy representation of $\mathscr{G}^{\circ}$, and the cohomology class in $H^1(\pi_1(\widetilde{U'}, \overline{x}), \mathrm{T}_p(\mathscr{G}^{\circ}))$ given by $g \mapsto C_{\sigma_i}(g)$ is nothing but the class defined in (7.4.6). By 7.5(i) and the induction hypothesis, $\rho_{\mathscr{G}^{\circ}}$ is surjective. Since the cohomology class $C_{\sigma_1} = 0$ by 7.5(ii), we may assume $C_{\sigma_1}(g) = 0$ for all $g \in \pi_1(U', \overline{x})$. Therefore $\mathrm{Im}(\rho_{\mathscr{G}_1})$ contains all the matrix of the form $\begin{pmatrix} B & 0 \\ 0 & 1 \end{pmatrix}$ with $B \in \mathrm{GL}_{n-1}(\mathbb{Z}_p)$. By the functoriality of monodromy, $\mathrm{Im}(\rho_{\mathscr{G}_{R'}})$ contains $\mathrm{Im}(\rho_{\mathscr{G}_1})$. Hence we have

$$\begin{pmatrix} \mathrm{GL}_{n-1}(\mathbb{Z}_p) & 0 \\ 0 & 1 \end{pmatrix} \subset \mathrm{Im}(\rho_{\mathscr{G}_1}) \subset \mathrm{Im}(\rho_{\mathscr{G}_{R'}}). \tag{7.5.3}$$

On the other hand, since the cohomology class $\overline{C}_{\sigma_2} \neq 0$, there exists a $g \in \pi_1(\widetilde{U'}, \overline{x})$ such that $b_2 = \overline{C}_{\sigma_2}(g) \neq 0$. Hence the matrix $\rho_{\mathscr{G}_2}(g)$ has the form $\begin{pmatrix} B_2 & b_2 \\ 0 & 1 \end{pmatrix}$ such that $B_2 \in \mathrm{GL}_{n-1}(\mathbb{Z}_p)$ and the image of $b_2 \in \mathrm{M}_{1\times n-1}(\mathbb{Z}_p)$

in $\mathrm{M}_{1\times n-1}(\mathbb{F}_p)$ is non-zero. By the functoriality of monodromy, we have $\mathrm{Im}(\rho_{\mathscr{G}_2}) \subset \mathrm{Im}(\rho_{\mathscr{G}_{R'}})$; in particular, we have $\begin{pmatrix} B_2 & b_2 \\ 0 & 1 \end{pmatrix} \in \mathrm{Im}(\rho_{\mathscr{G}_{R'}})$. In view of (7.5.3), we get

$$\begin{pmatrix} \mathrm{GL}_{n-1}(\mathbb{Z}_p) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} B_2 & b_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{GL}_{n-1}(\mathbb{Z}_p) & 0 \\ 0 & 1 \end{pmatrix} \subset \mathrm{Im}(\rho_{\mathscr{G}_{R'}}). \tag{7.5.4}$$

But the subset of $\mathrm{GL}_n(\mathbb{Z}_p)$ on the left hand side is just the subgroup H described in condition (b). Therefore, condition (b) is verified for $\rho_{\mathscr{G}_{R'}}$, and the proof of 7.3 is complete.

The rest of this section is dedicated to the proof of Lemma 7.5.

LEMMA 7.6. *Let k be an algebraically closed field of characteristic $p > 0$, A be a noetherian henselian local k-algebra with residue field k, G be a BT-group over A, and $G^{\text{ét}}$ be its étale part. Put*

$$\mathrm{Lie}(G^\vee)^{\varphi=1} = \{x \in \mathrm{Lie}(G^\vee) \textit{ such that } \varphi_G(x) = x\}.$$

Then $\mathrm{Lie}(G^\vee)^{\varphi=1}$ is an $\mathbb{F}_p$-vector space of dimension equal to the rank of $\mathrm{Lie}(G^{\text{ét}\vee})$, and the A-submodule $\mathrm{Lie}(G^{\text{ét}\vee})$ of $\mathrm{Lie}(G^\vee)$ is generated by $\mathrm{Lie}(G^\vee)^{\varphi=1}$.

Proof. Let r be the rank of $\mathrm{Lie}(G^{\text{ét}\vee})$, G° be the connected part of G, and s be the height of $\mathrm{Lie}(G^{\circ\vee})$. We have an exact sequence of A-modules

$$0 \to \mathrm{Lie}(G^{\text{ét}\vee}) \to \mathrm{Lie}(G^\vee) \to \mathrm{Lie}(G^{\circ\vee}) \to 0,$$

compatible with Hasse-Witt maps. We choose a basis of $\mathrm{Lie}(G^\vee)$ adapted to this exact sequence, so that φ_G is expressed by a matrix of the form $\begin{pmatrix} U & W \\ 0 & V \end{pmatrix}$ with $U \in \mathrm{M}_{r\times r}(A)$, $V \in \mathrm{M}_{s\times s}(A)$, and $W \in \mathrm{M}_{r\times s}(A)$. An element of $\mathrm{Lie}(G^\vee)^{\varphi=1}$ is given by a vector $\begin{pmatrix} x \\ y \end{pmatrix}$, where $x = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ \vdots \\ y_s \end{pmatrix}$ with $x_i, y_j \in A$, satisfying

$$\begin{pmatrix} U & W \\ 0 & V \end{pmatrix} \cdot \begin{pmatrix} x^{(p)} \\ y^{(p)} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} U \cdot x^{(p)} + W \cdot y^{(p)} = x \\ V \cdot y^{(p)} = y. \end{cases} \tag{7.6.1}$$

where $x^{(p)}$ (*resp.* $y^{(p)}$) is the vector obtained by applying $a \mapsto a^p$ to each $x_i (1 \leq i \leq r)$ (*resp.* $y_j (1 \leq j \leq s)$). By 2.9, the Hasse-Witt map of the special fiber of G° is nilpotent. So there exists an integer $N \geq 1$ such that $\varphi^N_{G^\circ}(\mathrm{Lie}(G^{\circ\vee})) \subset \mathfrak{m}_A \cdot \mathrm{Lie}(G^{\circ\vee})$, *i.e.* we have $V \cdot V^{(p)} \cdots V^{(p^{N-1})} \equiv 0 \pmod{\mathfrak{m}_A}$. From the equation $V \cdot y^{(p)} = y$, we deduce that

$$y = V \cdot V^{(p)} \cdots V^{(p^{N-1})} \cdot y^{(p^N)} \equiv 0 \pmod{\mathfrak{m}_A}.$$

But this implies that $y^{(p^N)} \equiv 0 \pmod{\mathfrak{m}_A^{p^N}}$. Hence we get $y = V \cdot y^{(p)} \equiv 0 \pmod{\mathfrak{m}_A^{p^N+1}}$. Repeting this argument, we get finally $y \equiv 0 \pmod{\mathfrak{m}_A^{\ell}}$ for all integers $\ell \geq 1$, so $y = 0$. This implies that $\mathrm{Lie}(G^\vee)^{\varphi=1} \subset \mathrm{Lie}(G^{\text{ét}\vee})$, and the equation (7.6.1) is simplified as $U \cdot x^{(p)} = x$. Since the linearization of $\varphi_{G^{\text{ét}}}$ is bijective by 2.11, we have $U \in \mathrm{GL}_r(A)$. Let $\overline{U}$ be the image of U in $\mathrm{GL}_r(k)$, and Sol be the solutions of the equation $\overline{U} \cdot x^{(p)} = x$. As k is algebraically closed, Sol is an $\mathbb{F}_p$-space of dimension r, and $\mathrm{Lie}(G^{\text{ét}\vee}) \otimes k$ is generated by Sol (cf. [Ka2, Prop. 4.1]). By the henselian property of A, every elements in Sol lifts uniquely to a solution of $U \cdot x^{(p)} = x$, *i.e.* the reduction map $\mathrm{Lie}(G^\vee)^{\varphi=1} \xrightarrow{\sim} \mathrm{Sol}$ is bijective. By Nakayama's lemma, $\mathrm{Lie}(G^\vee)^{\varphi=1}$ generates the A-module $\mathrm{Lie}(G^{\text{ét}\vee})$. □

7.7. We keep the notations of 7.4. Let $\mathbf{Comp}_{\overline{K}_0}$ be the category of noetherian complete local $\overline{K}_0$-algebras with residue field $\overline{K}_0$, $\mathcal{D}_{\mathscr{G}_{\overline{K}_0}}$ (*resp.* $\mathcal{D}_{\mathscr{G}^\circ_{\overline{K}_0}}$) be the functor which associates to every object A of $\mathbf{Comp}_{\overline{K}_0}$ the set of isomorphsm classes of deformations of $\mathscr{G}_{\overline{K}_0}$ (*resp.* $\mathscr{G}^\circ_{\overline{K}_0}$) . If A is an object in $\mathbf{Comp}_{\overline{K}_0}$ and G is a deformation of $\mathscr{G}_{\overline{K}_0}$ (*resp.* $\mathscr{G}^\circ_{\overline{K}_0}$) over A, we denote by $[G]$ its isomorphic class in $\mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(A)$ (*resp.* in $\mathcal{D}_{\mathscr{G}^\circ_{\overline{K}_0}}$).

LEMMA 7.8. *Let Σ be the set defined in* (7.4.3).
(i) *The morphism of sets $\Phi : \Sigma \to \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'})$ given by $\sigma \mapsto [\mathscr{G}_{\widetilde{R'},\sigma}]$ is bijective.*
(ii) *Let $\sigma \in \Sigma$. Then there exists a basis of $\mathrm{Lie}(\mathscr{G}^{\circ\vee}_{\widetilde{R'},\sigma})$ such that $\varphi_{\mathscr{G}^\circ_{\widetilde{R'},\sigma}}$ is represented by a matrix of the form*

$$\mathfrak{h}^\circ_\sigma = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_1 \\ 1 & 0 & \cdots & 0 & a_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & a_{n-1} \end{pmatrix} \tag{7.8.1}$$

with $a_i \equiv \alpha \cdot \sigma(t_i) \pmod{\mathfrak{m}^2_{\widetilde{R'}}}$ for $1 \leq i \leq n-1$, where $\alpha \in \widetilde{R'}^{\times}$ and $\mathfrak{m}_{\widetilde{R'}}$ is the maximal ideal of $\widetilde{R'}$. In particular, $\mathscr{G}^\circ_{\widetilde{R'},\sigma}$ is the universal deformation of $\mathscr{G}^\circ_{\overline{K}_0}$ if and only if $\{\sigma(t_1), \cdots, \sigma(t_{n-1})\}$ is a system of regular parameters of $\widetilde{R'}$.

Proof. (i) We begin with a remark on the Kodaira-Spencer map of $\mathscr{G}_{R'}$. Let $\mathscr{T}_{\mathbf{S}/k} = \mathscr{H}om_{\mathscr{O}_\mathbf{S}}(\Omega^1_{\mathbf{S}/k}, \mathscr{O}_\mathbf{S})$ be the tangent sheaf of $\mathbf{S}$. Since $\mathbf{G}$ is universal, the Kodaira-Spencer map (3.2.2)

$$\mathrm{Kod} : \mathscr{T}_{\mathbf{S}/k} \xrightarrow{\sim} \mathscr{H}om_{\mathscr{O}_\mathbf{S}}(\omega_\mathbf{G}, \mathrm{Lie}(\mathbf{G}^\vee))$$

is an isomorphism. By functoriality, this induces an isomorphism of R'-modules

$$\mathrm{Kod}_{R'} : T_{R'/k} \xrightarrow{\sim} \mathrm{Hom}_{R'}(\omega_{\mathscr{G}_{R'}}, \mathrm{Lie}(\mathscr{G}^\vee_{R'})), \tag{7.8.2}$$

where $T_{R'/k} = \mathrm{Hom}_{R'}(\Omega^1_{R'/k}, R') = \Gamma(\mathbf{S}, \mathscr{T}_{\mathbf{S}/k}) \otimes_R R'$.

For each integer $\nu \geq 0$, we put $\widetilde{R'}_\nu = \widetilde{R'}/\mathfrak{m}^{\nu+1}_{\widetilde{R'}}$, Σ_ν to be the set of liftings of $R \to K_0 \to \overline{K}_0$ to $R \to \widetilde{R'}_\nu$, and $\Phi_\nu : \Sigma_\nu \to \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'}_\nu)$ to be the morphism of sets $\sigma_\nu \mapsto [\mathscr{G}_{R'} \otimes_{\sigma_\nu} \widetilde{R'}_\nu]$. We prove by induction on ν that Φ_ν is bijective for all $\nu \geq 0$. This will complete the proof of (i). For $\nu = 0$, the claim holds trivially. Assume that it holds for $\nu - 1$ with $\nu \geq 1$. We have a commutative diagram

$$\begin{tikzcd} \Sigma_\nu \arrow[r, "\Phi_\nu"] \arrow[d] & \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'}_\nu) \arrow[d] \\ \Sigma_{\nu-1} \arrow[r, "\Phi_{\nu-1}"] & \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'}_{\nu-1}), \end{tikzcd}$$

where the vertical arrows are the canonical reductions, and the lower arrow is an isomorphism by induction hypothesis. Let τ be an arbitrary element of $\Sigma_{\nu-1}$. We denote by $\Sigma_{\nu,\tau} \subset \Sigma_\nu$ the preimage of τ, and by $\mathcal{D}_{\Phi_{\nu-1}(\tau)}(\widetilde{R'}_\nu) \subset \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'}_\nu)$ the preimage of $\Phi_{\nu-1}(\tau)$. It suffices to prove that Φ_ν induces a bijection between $\Sigma_{\nu,\tau}$ and $\mathcal{D}_{\Phi_{\nu-1}(\tau)}(\widetilde{R'}_\nu)$. Let $I_\nu = \mathfrak{m}^\nu_{\widetilde{R'}}/\mathfrak{m}^{\nu+1}_{\widetilde{R'}}$ be the ideal of the reduction map $\widetilde{R'}_\nu \to \widetilde{R'}_{\nu-1}$. By [EGA, 0_{IV} 21.2.5 and 21.9.4], we have $\Omega^1_{R'/k} \simeq \widehat{\Omega}^1_{R'/k}$, and they are free over A of rank n. By [EGA, 0_{IV} 20.1.3], $\Sigma_{\nu,\tau}$ is a (nonempty) homogenous space under the group

$$\mathrm{Hom}_{K_0}(\Omega^1_{R'/k} \otimes_{R'} K_0, I_\nu) = T_{R'/k} \otimes_{R'} I_\nu.$$

On the other hand, according to 3.5(i), $\mathcal{D}_{\Phi_{\nu-1}(\tau)}(\widetilde{R'}_\nu)$ is a homogenous space under the group

$$\mathrm{Hom}_{\overline{K}_0}(\omega_{\mathscr{G}_{\overline{K}_0}}, \mathrm{Lie}(\mathscr{G}^\vee_{\overline{K}_0})) \otimes_{\overline{K}_0} I_\nu = \mathrm{Hom}_{R'}(\omega_{\mathscr{G}_{R'}}, \mathrm{Lie}(\mathscr{G}^\vee_{R'})) \otimes_{R'} I_\nu.$$

Moreover, it is easy to check that the morphism of sets $\Phi_\nu : \Sigma_{\nu,\tau} \to \mathcal{D}_{\Phi_{\nu-1}(\tau)}(\widetilde{R'}_\nu)$ is compatible with the homomorphism of groups

$$\mathrm{Kod}_{R'} \otimes_{R'} \mathrm{Id} : T_{R'/k} \otimes_{R'} I_\nu \to \mathrm{Hom}_{R'}(\omega_{\mathscr{G}_{R'}}, \mathrm{Lie}(\mathscr{G}^\vee_{R'})) \otimes_{R'} I_\nu,$$

where $\mathrm{Kod}_{R'}$ is the Kodaira-Spencer map (7.8.2) associated to $\mathscr{G}_{R'}$. The bijectivity of Φ_ν now follows from the fact that $\mathrm{Kod}_{R'}$ is an isomorphism.
(ii) The second part of the statement follows immediately from 4.11. It remains to compute the Hasse-Witt map of $\mathscr{G}^\circ_{\widetilde{R'},\sigma}$. We determine first the submodule $\mathrm{Lie}(\mathscr{G}^{\text{ét}\vee}_{\widetilde{R'},\sigma})$ of $\mathrm{Lie}(\mathscr{G}^\vee_{\widetilde{R'},\sigma})$. We choose a basis of $\mathrm{Lie}(\mathbf{G}^\vee)$ over $\mathscr{O}_{\mathbf{S}}$ such that $\varphi_{\mathbf{G}}$ is expressed by the matrix $\mathfrak{h}$ (7.4.1). As $\mathscr{G}_{\widetilde{R'},\sigma}$ derives from $\mathbf{G}$ by base change $R \to R' \xrightarrow{\sigma} \widetilde{R'}$, there exists a basis $(e_1, \cdots, e_n)$ of $\mathrm{Lie}(\mathscr{G}^\vee_{\widetilde{R'},\sigma})$ such that $\varphi_{\mathscr{G}_{\widetilde{R'},\sigma}}$ is expressed by

$$\mathfrak{h}^\sigma = \begin{pmatrix} 0 & 0 & \cdots & 0 & -\sigma(t_1) \\ 1 & 0 & \cdots & 0 & -\sigma(t_2) \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -\sigma(t_n) \end{pmatrix}.$$

By Lemma 7.6, $\mathrm{Lie}(\mathscr{G}^{\text{ét}\vee}_{\widetilde{R'},\sigma})$ is generated by $\mathrm{Lie}(\mathscr{G}^\vee_{\widetilde{R'},\sigma})^{\varphi=1}$. If $\sum_{i=1}^n x_n e_n \in \mathrm{Lie}(\mathscr{G}^\vee_{\widetilde{R'},\sigma})^{\varphi=1}$ with $x_i \in \widetilde{R'}$ for $1 \leq i \leq n$, then $(x_i)_{1\leq i\leq n}$ must satisfy the equation $\mathfrak{h}^\sigma \cdot \begin{pmatrix} x_1^p \\ \vdots \\ x_n^p \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$; or equivalently,

$$\begin{cases} x_1 = -\sigma(t_1)x_n^p \\ x_2 = -\sigma(t_2)x_n^p - \sigma(t_1)^p x_n^{p^2} \\ \cdots \\ x_{n-1} = -\sigma(t_{n-1})x_n^p - \cdots - \sigma(t_1)^{p^{n-2}}x_n^{p^{n-1}} \\ \sigma(t_1)^{p^{n-1}}x_n^{p^n} + \sigma(t_2)^{p^{n-2}}x_n^{p^{n-1}} + \cdots + \sigma(t_n)x_n^p + x_n = 0. \end{cases} \tag{7.8.3}$$

We note that $\sigma(t_i) \in \mathfrak{m}_{\widetilde{R'}}$ for $1 \leq i \leq n-1$ and $\sigma(t_n) \in \widetilde{R'}^\times$ with image $i(t_n) \in \overline{K}_0$, where $i : K_0 \to \overline{K}_0$ is the fixed immbedding. By Hensel's lemma, every solution in $\overline{K}_0$ of the equation $i(t_n)x_n^p + x_n = 0$ lifts uniquely to a solution of (7.8.3). As $\mathrm{Lie}(\mathscr{G}^{\text{ét}\vee}_{\widetilde{R'},\sigma})$ has rank 1, by Lemma 7.6, these are all the solutions of (7.8.3). Let $(\lambda_1, \cdots, \lambda_n)$ be a non-zero solution of (7.8.3). We have

$$\lambda_n \in \widetilde{R'}^\times \quad \text{and} \quad \lambda_i \equiv -\lambda_n^p\sigma(t_i) \pmod{\mathfrak{m}^2_{\widetilde{R'}}}. \tag{7.8.4}$$

We put $v = \lambda_1 e_1 + \cdots + \lambda_n e_n$; so v is a basis of $\mathrm{Lie}(\mathscr{G}^{\text{ét}\vee}_{\widetilde{R'},\sigma})$ by 7.6. For $1 \leq i \leq n$, let f_i be the image of e_i in $\mathrm{Lie}(\mathscr{G}^{\circ\vee}_{\widetilde{R'},\sigma})$. Then $f_1, \cdots, f_n$ clearly generate $\mathrm{Lie}(\mathscr{G}^{\circ\vee}_{\widetilde{R'},\sigma})$. By the explicit description above of $\mathrm{Lie}(\mathscr{G}^{\text{ét}\vee}_{\widetilde{R'},\sigma})$, we have $f_n = -\lambda_n^{-1}(\lambda_1 f_1 \cdots + \lambda_{n-1}f_{n-1})$. Hence $f_1, \cdots, f_{n-1}$ form a basis of $\mathrm{Lie}(\mathscr{G}^{\circ\vee}_{\widetilde{R'},\sigma})$. By the functoriality of Hasse-Witt maps, we have $\varphi_{\mathscr{G}^\circ_{\widetilde{R'}}}(f_i) = f_{i+1}$ for $1 \leq i \leq n-1$, or equivalently,

$$\varphi_{\mathscr{G}^\circ_{\widetilde{R'},\sigma}}(f_1, \cdots, f_{n-1}) = (f_1, \cdots, f_{n-1}) \cdot \begin{pmatrix} 0 & 0 & \cdots & 0 & -\lambda_n^{-1}\lambda_1 \\ 1 & 0 & \cdots & 0 & -\lambda_n^{-1}\lambda_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -\lambda_n^{-1}\lambda_{n-1} \end{pmatrix}.$$

In view of (7.8.4), we see that the above matrix has the form of (7.8.1) by setting $\alpha = \lambda_n^{p-1} \in \widetilde{R'}^\times$. The second part of statement (ii) follows immediately from Proposition 4.11(ii) and the description above of $\varphi_{\mathscr{G}^\circ_{\widetilde{R'},\sigma}}$. □

Now we can turn to the proof of 7.5.

7.9. Proof of Lemma 7.5. First, suppose that we have found a $\sigma_2 \in \Sigma$ such that $\overline{C}_{\sigma_2} \neq 0$ and $\mathscr{G}^\circ_{\widetilde{R'},\sigma_2}$ is the universal deformation of $\mathscr{G}^\circ_{\overline{K}_0}$. Since $\Phi : \Sigma \xrightarrow{\sim} \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'})$ is bijective by 7.8(i), there exists a $\sigma_1 \in \Sigma$ corresponding to the deformation $[\mathscr{G}^\circ_{\widetilde{R'},\sigma_2} \oplus \mathbb{Q}_p/\mathbb{Z}_p] \in \mathcal{D}_{\mathscr{G}_{\overline{K}_0}}(\widetilde{R'})$. It is clear that $\mathscr{G}^\circ_{\widetilde{R'},\sigma_1} \simeq \mathscr{G}^\circ_{\widetilde{R'},\sigma_2}$. Besides, the exact sequence (7.4.5) for σ_1 splits; so we have $C_{\sigma_1} = 0$. It remains to prove the existence of σ_2. We note first that $\overline{K}_0$ can be canonically imbedded into $\widetilde{R'}$, since it is perfect. Since R' is formally smooth over k and

$(t_1, \cdots, t_n)$ is a p-basis of R' over k, by [EGA, 0_{IV} 21.2.7], there is a $\sigma \in \Sigma$ such that $\sigma(t_i)$ $(1 \leq i \leq n-1)$ form a system of regular parameters of $\widetilde{R'}$ and $\sigma(t_n) \in \overline{K}_0 \subset \widetilde{R'}$. We claim that $\sigma_2 = \sigma$ answers the question. In fact, Lemma 7.8(ii) implies that $\mathscr{G}^{\circ}_{\widetilde{R'},\sigma}$ is the universal deformation of $\mathscr{G}^{\circ}_{\overline{K}_0}$. It remains to verify that $\overline{C}_\sigma \neq 0$.

Let $A = \overline{K}_0[[\pi]]$ be a complete discrete valuation ring of characteristic p with residue field $\overline{K}_0$, $T = \mathrm{Spec}(A)$, ξ be the generic point of T, $\overline{\xi}$ be a geometric over ξ, and $I = \mathrm{Gal}(\overline{\xi}/\xi)$ the Galois group. We define a homomorphism of $\overline{K}_0$-algebras $f^* : \widetilde{R'} \to A$ by putting $f^*(\sigma(t_1)) = \pi$ and $f^*(\sigma(t_i)) = 0$ for $2 \leq i \leq n-1$. This is possible, since $(\sigma(t_1), \cdots, \sigma(t_{n-1}))$ is a system of regular parameters of $\widetilde{R'}$. Let $f : T \to \widetilde{S'}$ be the homomorphism of schemes corresponding to f^*, and $\mathscr{G}_T = \mathscr{G}_{\widetilde{R'},\sigma} \times_{\widetilde{S'}} T$. By the functoriality of Hasse-Witt maps,

$$\mathfrak{h}_T = \begin{pmatrix} 0 & 0 & \cdots & 0 & -\pi \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -f^*(\sigma(t_n)) \end{pmatrix} \in \mathrm{M}_{n\times n}(\widetilde{R'})$$

is a matrix of $\varphi_{\mathscr{G}_T}$. By definition (5.4), the Hasse invariant of $\mathscr{G}_T$ is $h(\mathscr{G}_T) = 1$. In particular, $\mathscr{G}_T$ is generically ordinary. Let $\widetilde{U}'_\sigma \subset \widetilde{S}'$ be the ordinary locus of $\mathscr{G}_{\widetilde{R'},\sigma}$. We have $f(\xi) \in \widetilde{U}'_\sigma$. By the functoriality of fundamental groups, f induces a homomorphism of groups

$$\pi_1(f) : I = \mathrm{Gal}(\overline{\xi}/\xi) \to \pi_1(\widetilde{U}'_\sigma, f(\overline{\xi})) \simeq \pi_1(\widetilde{U}'_\sigma, \overline{x}).$$

Let $\mathscr{G}^{\circ}_T$ be the connected part of $\mathscr{G}_T$, and $\mathscr{G}^{\text{ét}}_T$ be the étale part of $\mathscr{G}_T$. Then $\mathscr{G}^{\text{ét}}_T \simeq \mathbb{Q}_p/\mathbb{Z}_p$. We have an exact sequence of $\mathbb{F}_p[I]$-modules

$$0 \to \mathscr{G}^{\circ}_T(1)(\overline{\xi}) \to \mathscr{G}_T(1)(\overline{\xi}) \to \mathscr{G}^{\text{ét}}_T(1)(\overline{\xi}) \to 0,$$

which determines a cohomology class $\overline{C}_T \in H^1(I, \mathscr{G}^{\circ}_T(1)(\overline{\xi}))$. We notice that $\mathscr{G}_T(1)(\overline{\xi})$ is isomorphic to $\mathscr{G}_{\widetilde{R'},\sigma}(1)(\overline{x})$ as an abelian group, and the action of I on $\mathscr{G}_T(1)(\overline{\xi})$ is induced by the action of $\pi_1(\widetilde{U}'_\sigma, \overline{x})$ on $\mathscr{G}_{\widetilde{R'},\sigma}(1)(\overline{x})$. Therefore, $\overline{C}_T$ is the image of $\overline{C}_\sigma$ by the functorial map

$$H^1\big(\pi_1(\widetilde{U}'_\sigma, \overline{x}), \mathscr{G}^{\circ}_{\widetilde{R'},\sigma}(1)(\overline{x})\big) \to H^1\big(I, \mathscr{G}^{\circ}_T(1)(\overline{\xi})\big).$$

To verify that $\overline{C}_\sigma \neq 0$, it suffices to check that $\overline{C}_T \neq 0$. We consider the polynomial $P(X) = X^{p^n} + f^*(\sigma(t_n))X^{p^{n-1}} + \pi X \in A[X]$. According to 5.12, it suffices to find a $\alpha \in \overline{K}_0 \subset A$ such that $P(\alpha)$ is a uniformizer of A. But by the choice of σ, we have $\sigma(t_n) \in \overline{K}_0$ and $\sigma(t_n) \neq 0$; so $f^*(\sigma(t_n)) \neq 0$ lies in $\overline{K}_0$. Let α be a $p^{n-1}(p-1)$-th root of $-f^*(\sigma(t_n))$ in $\overline{K}_0$. Then we have $\alpha \in \overline{K}_0^{\times}$, and $P(\alpha) = \alpha\pi$ is a uniformizer of A. This completes the proof of 7.5.

8. End of the Proof of Theorem 1.3

In this section, k denotes an algebraically closed field of characteristic $p > 0$.

8.1. First, we recall some preliminaries on Newton stratification due to F. Oort. Let G be an arbitrary BT-group over k, $\mathbf{S}$ be the local moduli of G in characteristic p, and $\mathbf{G}$ be the universal deformation of G over $\mathbf{S}$ (3.8). Put $d = \dim(G)$ and $c = \dim(G^\vee)$. We denote by $\mathcal{N}(G)$ the Newton polygon of G which has endpoints $(0,0)$ and $(c+d,d)$. Here we use the normalization of Newton polygons such that slope 0 corresponds to étale BT- groups and slope 1 corresponds to groups of multiplicative type.
Let $\mathcal{NP}(c+d,d)$ be the set of Newton polygons with endpoints $(0,0)$ and $(c+d,d)$ and slopes in $(0,1)$. For $\alpha,\beta \in \mathcal{NP}(c+d,d)$, we say that $\alpha \preceq \beta$ if no point of α lies below β; then "$\preceq$" is a partial order on $\mathcal{NP}(c+d,d)$. For each $\beta \in \mathcal{NP}(c+d,d)$, we denote by V_β the subset of $\mathbf{S}$ consisting of points x with $\mathcal{N}(\mathbf{G}_x) \preceq \beta$, and by V_β° the subset of $\mathbf{S}$ consisting of points x with $\mathcal{N}(\mathbf{G}_x) = \beta$. By Grothendieck-Katz's specialization theorem of Newton polygons, V_β is closed in $\mathbf{S}$, and V_β° is open (maybe empty) in V_β. We put

$\diamondsuit(\beta) =$
$\{(x,y) \in \mathbb{Z}\times\mathbb{Z} \mid 0 \leq y < d, y < x < c+d, (x,y) \text{lies on or above the polygon } \beta\}$,
and $\dim(\beta) = \#(\diamondsuit(\beta))$.

Theorem 8.2 ([Oo2] Theorem 2.11). *Under the above assumptions, for each $\beta \in \mathcal{NP}(c+d,d)$, the subset V_β° is non-empty if and only if $\mathcal{N}(G) \preceq \beta$. In that case, V_β is the closure of V_β° and all irreducible components of V_β have dimension* $\dim(\beta)$.

8.3. Let G be a connected and HW-cyclic BT-group over k of dimension $d = \dim(G) \geq 2$. Let $\beta \in \mathcal{NP}(c+d,d)$ be the Newton polygon given by the following slope sequence:

$$\beta = (\underbrace{1/(c+1),\cdots,1/(c+1)}_{c+1},\underbrace{1,\cdots,1}_{d-1}).$$

We have $\mathcal{N}(G) \preceq \beta$ since G is supposed to be connected. By Oort's Theorem 8.2, V_β is a equal dimensional closed subset of the local moduli $\mathbf{S}$ of dimension $c(d-1)$. We endow V_β with the structure of a reduced closed subscheme of $\mathbf{S}$.

Lemma 8.4. *Under the above assumptions, let R be the ring of $\mathbf{S}$, and*

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_c \end{pmatrix} \in \mathrm{M}_{c\times c}(R)$$

be a matrix of the Hasse-Witt map φ_G. Then the closed reduced subscheme V_β of $\mathbf{S}$ is defined by the prime ideal $(a_1,\cdots,a_c)$. In particular, V_β is irreducible.

Proof. Note first that $\{a_1, \cdots, a_c\}$ is a subset of a system of regular parameters of R by 4.11(i). Let I be the ideal of R defining V_β. Let x be an arbitrary point of V_β, we denote by $\mathfrak{p}_x$ the prime ideal of R corresponding to x. Since the Newton polygon of the fibre $\mathbf{G}_x$ lies above β, $\mathbf{G}_x$ is connected. By Lemma 4.4, we have $a_i \in \mathfrak{p}_x$ for $1 \leq i \leq c$. Since V_β is reduced, we have $a_i \in I$. Let $\mathfrak{P} = (a_1, \cdots, a_c)$, and $V(\mathfrak{P})$ the closed subscheme of $\mathbf{S}$ defined by $\mathfrak{P}$. Then $V(\mathfrak{P})$ is an integral scheme of dimension $c(d-1)$ and $V_\beta \subset V(\mathfrak{P})$. Since Theorem 8.2 implies that $\dim V_\beta = c(d-1)$, we have necessarily $V_\beta = V(\mathfrak{P})$. $\square$

We keep the assumptions above. Let $(t_{i,j})_{1\leq i\leq c, 1\leq j\leq d}$ be a regular system of parameters of R such that $t_{i,d} = a_i$ for all $1 \leq i \leq c$. Let x be the generic point of the Newton strata V_β, $k' = \kappa(x)$, and $R' = \widehat{\mathscr{O}}_{\mathbf{S},x}$. Since R is noetherian and integral, the canonical ring homomorphism $R \to \mathscr{O}_{\mathbf{S},x} \to R'$ is injective. The image in R' of an element $a \in R$ will be denoted also by a. By choosing a k-section $k' \to R'$ of the canonical projection $R' \to k'$, we get a (non-canonical) isomorphism of k-algebras $R' \simeq k'[[t_{1,d}, \cdots, t_{c,d}]]$. Let k'' be an algebraic closure of k', and $R'' = k''[[t_{1,d}, \cdots, t_{c,d}]]$. Then we have a natural injective homomorphism of k-algebras $R' \to R''$ mapping $t_{i,d}$ to $t_{i,d}$ for $1 \leq i \leq c$.
Let $S'' = \mathrm{Spec}(R'')$, $\overline{x}$ be its closed point. By the construction of S'', we have a morphism of k-schemes

$$f : S'' \to \mathbf{S} \tag{8.4.1}$$

sending $\overline{x}$ to x. We put $\mathscr{G} = \mathbf{G} \times_{\mathbf{S}} S''$. By the choice of the Newton polygon β, the closed fibre $\mathscr{G}_{\overline{x}}$ has a BT-subgroup $\mathscr{H}_{\overline{x}}$ of multiplicative type of height $d-1$. Since S'' is henselian, $\mathscr{H}_{\overline{x}}$ lifts uniquely to a BT-subgroup $\mathscr{H}$ of $\mathscr{G}$. We put $\mathscr{G}'' = \mathscr{G}/\mathscr{H}$. It is a connected BT-group over S'' of dimension 1 and height $c+1$.

Lemma 8.5. *Under the above assumptions, $\mathscr{G}''$ is the universal deformation in equal characteristic of its special fiber.*

This lemma is a particular case of [Lau, Lemma 3.1]. Here, we use 4.11(ii) to give a simpler proof.

Proof. We have an exact sequence of BT-groups over S''

$$0 \to \mathscr{H} \to \mathscr{G} \to \mathscr{G}'' \to 0,$$

which induces an exact sequence of Lie algebras $0 \to \mathrm{Lie}(\mathscr{G}''^\vee) \to \mathrm{Lie}(\mathscr{G}^\vee) \to \mathrm{Lie}(\mathscr{H}^\vee) \to 0$ compatible with Hasse-Witt maps. Since $\mathscr{H}$ is of multiplicative type, we get $\mathrm{Lie}(\mathscr{H}^\vee) = 0$ and an isomorphism of Lie algebras $\mathrm{Lie}(\mathscr{G}''^\vee) \simeq \mathrm{Lie}(\mathscr{G}^\vee)$. By the choice of the regular system $(t_{i,j})_{1\leq i\leq c, 1\leq j\leq d}$, there is a basis $(v_1, \cdots, v_c)$ of $\mathrm{Lie}(\mathscr{G}''^\vee)$ over $\mathscr{O}_{S''}$ such that $\varphi_{\mathscr{G}''}$ is given by the matrix

$$\mathfrak{h} = \begin{pmatrix} 0 & 0 & \cdots & 0 & -t_{1,d} \\ 1 & 0 & \cdots & 0 & -t_{2,d} \\ 0 & 1 & \cdots & 0 & -t_{3,d} \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -t_{c,d} \end{pmatrix}.$$

Now the lemma results from Proposition 4.11(ii). $\square$

8.6. Proof of Theorem 1.3. The one-dimensional case is treated in 7.3. If $\dim(G) \geq 2$, we apply the preceding discussion to obtain the morphism $f\colon S'' \to \mathbf{S}$ and the BT-groups $\mathscr{G} = \mathbf{G} \times_{\mathbf{S}} S''$ and $\mathscr{G}''$, which is the quotient of $\mathscr{G}$ by the maximal subgroup of $\mathscr{G}$ of multiplicative type. Let U'' be the common ordinary locus of $\mathscr{G}$ and $\mathscr{G}''$ over S'', and $\overline{\xi}$ be a geometric point of U''. Then f maps U'' into the ordinary locus $\mathbf{U}$ of $\mathbf{G}$. We denote by

$$\rho_{\mathscr{G}} : \pi_1(U'', \overline{\xi}) \to \mathrm{Aut}_{\mathbb{Z}_p}(\mathrm{T}_p(\mathscr{G}, \overline{\xi}))$$

the monodromy representation associated to $\mathscr{G}$, and the same notation for $\rho_{\mathscr{G}''}$. By the functoriality of monodromy, we have $\mathrm{Im}(\rho_{\mathscr{G}}) \subset \mathrm{Im}(\rho_{\mathbf{G}})$. On the other hand, the canonical map $\mathscr{G} \to \mathscr{G}''$ induces an isomorphism of Tate modules $\mathrm{T}_p(\mathscr{G}, \overline{\eta}) \xrightarrow{\sim} \mathrm{T}_p(\mathscr{G}'', \overline{\eta})$ compatible with the action of $\pi_1(U'', \overline{\eta})$. Therefore, the group $\mathrm{Im}(\rho_{\mathscr{G}})$ is identified with $\mathrm{Im}(\rho_{\mathscr{G}''})$. Since $\mathscr{G}''$ is one-dimensional, we conclude the proof by Lemma 8.5 and Theorem 7.3.

References

[AN] J. Achter and P. Norman, Local monodromy of p-divisible groups, to appear in *Transactions of A.M.S.*, (2006).

[BBM] P. Berthelot, L. Breen and W. Messing, *Théorie de Dieudonné Cristalline II*, Lect. notes in Math. 930, Springer-Verlag, (1982).

[Bou] N. Bourbaki, *Algèbre Commutative*, Masson, Paris (1985).

[Ch1] L. Chai, Local monodromy for deformations of one dimensional formal groups, *J. reine angew. Math.* 524, (2000), 227-238.

[Ch2] L. Chai, Methods for p-adic monodromy, to appear in *Jussieu J. Math.*, (2006).

[CO] L. Chai and F. Oort, Monodromy and irreducibility of leaves, available at the webpage of F. Oort, (2008).

[DR] P. Deligne and K. Ribet, Values of abelian L-functions at negative integers over totally real fields. *Inven. Math.* 59, (1980), 227-286.

[Dem] M. Demazure, *Lectures on p-Divisible Groups*, Lect. notes in Math. 302, Springer-Verlag, (1972).

[DG] M. Demazure and A. Grothendieck, *Schéma en Groupes* I (SGA 3_{I}) , Lect. notes in Math. 151, Springer-Verlag, (1970).

[Eke] T. Ekedahl, The action of monodromy on torsion points of Jacobians, *Arithmetic Algebraic Geometry*, G. van der Geer, F. Oort and J. Steenbrink, ed. Progress in Math. 89, Birkhäuser, (1991), 41-49.

[FC] G. Faltings and L. Chai, *Degeneration of Abelian Varieties*, Ergebnisse Bd 22, Springer-Verlag,(1990).

[Gro] B. Gross, Ramification in p-adic Lie extensions, *Journée de Géométrie Algébrique de Rennes III*, *Astérisque* 65, (1979), 81-102.

[EGA] A. Grothendieck, Éléments de Géométrie algébrique IV, Étude locale des schémas et des morphismes de schémas, *Publ. Math. Inst. Hautes Étud. Sci.* 20, 24, 28, 32 (1964-1967).

[Gr] A. GROTHENDIECK, *Groupes de Barsotti-Tate et Cristaux de Dieudonné*, les Presses de l'Université de Montréal, (1974).

[Hid] H. HIDA p-adic automorphic forms on reductive groups, *Astérisque* 296 (2005), 147-254.

[Ill] L. ILLUSIE, Déformations de groupes de Barsotti-Tate (d'après A. Grothendieck), *Astérisque* 127 (1985), 151-198.

[Igu] J. IGUSA, On the algebraic theory of elliptic modular functions. *J. Math. Soc. Japan* 20 (1968), 96-106.

[deJ] A. J. DE JONG, Crystalline Dieudonné module theory via formal and rigide geometry, *Publ. Math. Inst. Hautes Étud. Sci.* 82 (1995), 5-96.

[Ka1] N. KATZ, Algebraic solutions of differential equations (p-curvature and the Hodge filtration), *Inven. Math.* 18 (1972), 1-118.

[Ka2] N. KATZ, p-adic properties of modular schemes and modular forms, in *Modular Functions of One Variable III*, Lect. notes in Math. 350, Springer-Verlag, (1973).

[Lau] E. LAU, Tate modules and universal p-divisible groups, arXiv:0803.1390v1, (2008).

[Oo1] F. OORT, Newton polygons and formal groups: conjectures by Manin and Grothendieck, *Ann. of Math.* 152 (2000), 183-206.

[Oo2] F. OORT, Newton polygon strata in the moduli space of abelian varieties, in *Moduli space of Abelian Varieties* (Progress in Mathematics 195), Birkhäuser-Verlag, (2001), 417-440.

[Se1] J. P. SERRE, *Corps Locaux*, Hermann, Paris, (1968).

[Se2] J. P. SERRE, *Abelian ℓ-adic representations and elliptic curves*, A K Peters, Wellesley, MA, (1998). Originally published in 1968 by W. A. Benjamin.

[Se3] J. P. SERRE, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, *Inven. Math.* 15 (1972), 259-331.

[Str] M. STRAUCH, Galois actions on torsion points of universal one-dimensional formal modules, arXiv: 0709.3542, (2007).

[Ti1] Y. TIAN, Thesis at University Paris 13, defensed in November 2007.

[Ti2] Y. TIAN, p-adic monodromy of the universal deformation of an elementary Barsotti-Tate group, arXiv: 0708.2022, (2007).

Yichao Tian
Department of Mathematics
Princeton University
Princeton
New Jersey
08544
USA
yichaot@princeton.edu

Galois Representations and Lubin-Tate Groups

Mark Kisin and Wei Ren

Received: May 3, 2009
Revised: June 6, 2009

Communicated by Takeshi Saito

Abstract. Using Lubin-Tate groups, we develop a variant of Fontaine's theory of (φ, Γ)-modules, and we use it to give a description of the Galois stable lattices inside certain crystalline representations.

Introduction

In his Grothendieck Festschrift paper [Fo 1], Fontaine introduced a new way to classify local Galois representation, using the theory of so called (φ, Γ)-modules. To recall this, let k be a perfect field of characteristic p, $K_0 = \operatorname{Fr} W(k)$ and K/K_0 a finite, totally ramified extension. Fix an algebraic closure $\bar{K}$ of K. Fontaine's theory starts with an infinite extension K_∞/K which is required to have certain ramification properties. Miraculously, these properties ensure that $G_{K_\infty} = \mathrm{Gal}(\bar{K}/K_\infty)$ can be identified with the absolute Galois group of a local field of *equal* characteristic p, $X(K)$. It is well known that representations of such a Galois group on finite dimensional $\mathbb{F}_p$-vector spaces can be classified rather concretely in terms of finite dimensional vector spaces over $X(K)$ equipped with an étale Frobenius. If K_∞/K is Galois, then $\Gamma = \mathrm{Gal}(K_\infty/K)$ acts naturally on $X(K)$, and one obtains a classification of G_K-representations on finite dimensional $\mathbb{F}_p$-vector spaces by adding a semi-linear action of Γ to the étale φ-modules over $X(K)$.
To obtain a classification of G_K-representations on finite $\mathbb{Z}_p$-modules, one needs to lift the action of φ and Γ on $X(K)$ to commuting operators on a Cohen ring for $X(K)$. This is probably not always possible, but can be done when K_∞ is the p-cyclotomic extension of K. Much of the work on Fontaine's theory by Berger, Colmez, Wach and others has focused on this case. In this paper we focus on the case when K_∞ is generated by the p-power torsion points of a Lubin-Tate group for a finite extension $L/\mathbb{Q}_p$ contained in K. As an application we obtain a description of the G_K-stable lattices in a certain class of crystalline G_K-representations. This is possible using the p-cyclotomic theory only when K is an unramified extension of some $\mathbb{Q}_p(\mu_{p^n})$.

More precisely, let $\mathcal{G}$ be a Lubin-Tate group over $\mathcal{O}_L$, write k_L for the residue field of L, fix a uniformizer π_L of L, and write $R = \varprojlim \mathcal{O}_{\bar{K}}/p$, where the transition maps in the inverse limit are given by Frobenius. The action of G_K on the Tate module $T\mathcal{G}$ of $\mathcal{G}$ gives rise to a character $\chi : \Gamma \to \mathcal{O}_L^\times$. It turns out that, using the periods of $T\mathcal{G}$ one can construct a subring $\mathcal{O}_\mathcal{E} \subset W(\mathrm{Fr}\, R) \otimes_{W(k_L)} L$ which is naturally a Cohen ring for $X(K)$. The action of $\mathcal{O}_L^\times \subset \mathcal{O}_L$ on $\mathcal{G}$ gives rise to a natural lifting of the action of Γ to $\mathcal{O}_\mathcal{E}$ (via χ), while the action of π_L on $\mathcal{G}$ allows one to lift the q-Frobenius $\varphi_q = \varphi^r$ to $\mathcal{O}_\mathcal{E}$, where $q = |k_L|$.
This allows one to classify G_K-representations on finite $\mathcal{O}_L$-modules in terms of étale (φ_q, Γ)-modules (see Theorem 1.6 below), and is explained in §1 of the paper. At least some part of this construction was certainly known to experts. The construction of the periods involved is in Colmez's paper [Col 1], and some of the ideas go back to Coleman [Co]. This material is also closely related to the subject of Fourquaux's thesis [Fou, §1.4].
In §2,3 we use this classification to give a classification of Galois stable lattices in certain crystalline G_K-representations, assuming that $K \subset K_0 \cdot L_\infty$ where L_∞/L is the field generated by the torsion points of $\mathcal{G}$. To explain the classification, assume for simplicity that $K = K_0 \cdot L$, and let $\mathfrak{S}_L = \mathcal{O}_K[\![u]\!]$. Fix a co-ordinate X on $\mathcal{G}$, and for $a \in \mathcal{O}_L$ denote by $[a] \in \mathcal{O}_L[\![X]\!]$ the power series giving the action of a on $\mathcal{G}$. Then $\gamma \in \Gamma$ acts on $\mathfrak{S}_L$ by $u \mapsto [\chi(\gamma)](u)$, while φ_q acts on $\mathfrak{S}_L$ by $u \mapsto [\pi_L](u)$. Let $Q = [\pi_L](u)/u$. We denote by $\mathrm{Mod}_{/\mathfrak{S}_L}^{\varphi_q,\Gamma}$ the category of finite free $\mathfrak{S}_L$-modules equipped with a continuous semi-linear action of Γ which induces the trivial action of $\mathfrak{M}/u\mathfrak{M}$, and an isomorphism $\varphi_q^*\mathfrak{M}[1/Q] \xrightarrow{\sim} \mathfrak{M}[1/Q]$ such that the map $1 \otimes \varphi_q : \mathfrak{M} \to \mathfrak{M}[1/Q]$ commutes with the action of Γ. Inside this category is a subcategory $\mathrm{Mod}_{/\mathfrak{S}_L}^{\varphi_q,\Gamma,\mathrm{an}}$ consisting of objects $\mathfrak{M}$ on which the Γ-action is $\mathcal{O}_L$-*analytic*. this means that there is $\mathcal{O}_L$-linear map of Lie algebras $d\Gamma : \mathrm{Lie}\,\Gamma \to \mathrm{End}_K(\mathfrak{M} \otimes_{\mathfrak{S}_L} K[\![u]\!])$, such that the action of an open subgroup of Γ is obtained by exponentiating $d\Gamma$.
To describe the crystalline representations we allow, consider any crystalline G_K-representation on an L-vector space V. Then

$$D_{\mathrm{dR}}(V) = (B_{\mathrm{dR}} \otimes_{\mathbb{Q}_p} V)^{G_K} = \oplus_\mathfrak{m} D_{\mathrm{dR}}(V)_\mathfrak{m}$$

where $\mathfrak{m}$ runs over the maximal ideals of $K \otimes_{\mathbb{Q}_p} L$. We say that V is L-crystalline if the filtration on $D_{\mathrm{dR}}(V)_\mathfrak{m}$ is trivial, unless $\mathfrak{m}$ is the kernel of the natural map $K \otimes_{\mathbb{Q}_p} L \hookrightarrow K$ corresponding to the inclusion $L \to K$. One of our main results is then the following

Theorem (0.1). *There is an exact equivalence of $\otimes$-categories between $\mathrm{Mod}_{/\mathfrak{S}_L}^{\varphi_q,\Gamma,\mathrm{an}}$ and the category of G_K-stable $\mathcal{O}_L$-lattices in L-crystalline G_K-representations.*

The theorem is a generalization of the classification of G_K-stable lattices in crystalline representations in terms of Wach lattices due to Wach [Wa], Colmez [Col 2] and Berger [Be 3], when K_∞ is the p-cyclotomic extension and K is unramified.

It is also analogous to the classification G_{K_∞}-stable lattices, obtained in [Ki] in the case when K_∞ is obtained from K by adjoining the p-power roots of a uniformizer. The advantage of Theorem (0.1) is that it applies without restriction on the ramification of K, and gives a precise description of G_K-stable lattices. Unfortunately, it applies only to a rather special kind of crystalline G_K-representation. It seems likely that in order to obtain a classification valid for any crystalline G_K-representation one needs to consider higher dimensional subrings of $W(\mathrm{Fr}\, R)$, constructed using the periods of all the conjugates of $\mathcal{G}$.

Acknowledgment: The results presented here were to be the subject of the Ph. D thesis of Wei Ren. They were written up by the first author after Ren's premature and tragic death.
We would like to thank Barry Mazur and Jean-Marc Fontaine for useful conversations on some of the material presented here. Finally we thank the referee for a careful reading of the paper and some useful remarks.

§1 Étale (φ_q, Γ)-modules

(1.1) Throughout the paper we fix a perfect field k, of characteristic $p > 0$. Let $W = W(k)$, $K_0 = W[1/p]$ and K/K_0 a finite totally ramified extension with ring of integers $\mathcal{O}_K$, and uniformizer π. We also fix an algebraic closure $\bar{K}$ of K with ring of integers $\mathcal{O}_{\bar{K}}$, and set $G_K = \mathrm{Gal}(\bar{K}/K)$.
Let $L/\mathbb{Q}_p$ be a finite extension of $\mathbb{Q}_p$ contained in K. Let $\mathcal{O}_L$ denote the ring of integers of L, and $k_L \subset k$ its residue field. Write $\mathcal{O}_{L_0} = W(k_L)$, $L_0 = \mathcal{O}_{L_0}[1/p]$, and $q = p^r = |k_L|$. For an $\mathcal{O}_{L_0}$-algebra A, it will be convenient to write $A_L = A \otimes_{\mathcal{O}_{L_0}} \mathcal{O}_L$.
Let $\mathcal{G}$ be a Lubin-Tate group over L corresponding to a uniformizer $\pi_L \in L$. Fix a local co-ordinate X on $\mathcal{G}$ so that the formal Hopf algebra $\mathcal{O}_\mathcal{G}$ may be identified with $\mathcal{O}_L[\![X]\!]$. For $a \in \mathcal{O}_L$ we denote by $[a] \in \mathcal{O}_L[\![X]\!] = \mathcal{O}_\mathcal{G}$ the power series giving the endomorphism a of $\mathcal{G}$.
For $n \geq 1$, let $K_n \subset \bar{K}$ denote the subfield generated by the π_L^n-torsion points of $\mathcal{G}$. We set $K_\infty = \cup_n K_n$ and we write $\Gamma = \mathrm{Gal}(K_\infty/K)$ and $G_{K_\infty} = \mathrm{Gal}(\bar{K}/K_\infty)$. Let $T\mathcal{G}$ denote the p-adic Tate module of $\mathcal{G}$. Then $T\mathcal{G}$ is a free $\mathcal{O}_L$-module of rank 1, and the action of Γ induces a faithful character $\chi : \Gamma \to \mathcal{O}_L^\times$.
We let $R = \varprojlim \mathcal{O}_{\bar{K}}/p$ with the transition maps being given by Frobenius. We may also identify R with $\varprojlim \mathcal{O}_{\bar{K}}/\pi_L$ with the transition map being given by the q-Frobenius φ^r. Evaluation of X at π_L-torsion points then induces a map $\iota : T\mathcal{G} \to R$. Namely if $v = (v_n)_{n\geq 0} \in T\mathcal{G}$ with $v_n \in \mathcal{G}[\pi_L^n](\mathcal{O}_{\bar{K}})$ and $\pi_L \cdot v_{n+1} = v_n$, then $\iota(v) = (v_n^*(X))_{n\geq 0}$.

Lemma (1.2). *There is a unique map $\{\,\} : R \to W(R)_L$ such that $\{x\}$ is a lifting of x, and $\varphi^r(\{x\}) = [\pi_L](x)$. Moreover $\{\,\}$ respects the action of G_K, and for $v \in T\mathcal{G}$ we have*

(1) *If $a \in \mathcal{O}_L$ then $\{\iota(av)\} = [a](\{\iota(v)\})$.*

(2) *The action of G_K on $\{\iota(T\mathcal{G})\}$ factors through Γ and for $\gamma \in \Gamma$*

$$[\chi(\gamma)](\{\iota(v)\}) = \{\iota(\gamma v)\} = \{\gamma \cdot \iota(v)\} = \gamma \cdot \{\iota(v)\}$$

In particular, if $v \in T\mathcal{G}$ is an $\mathcal{O}_L$-generator, there is an embedding $W_L[\![u]\!] \hookrightarrow W(R)_L$ sending u to $\{\iota(v)\}$ which identifies $W_L[\![u]\!]$ with a G_K-stable, φ^r-stable subring of $W(R)_L$ such that $\{\iota(T\mathcal{G})\}$ lies in the image of $W_L[\![u]\!]$.

Proof. The existence and uniqueness of $\{\ \}$ is [Col 1, Lem. 9.3]. The map $\{x\}$ is given by

$$\{x\} = \lim_n [\pi_L^n](\varphi^{-rn}(\tilde{x}))$$

where $\tilde{x} \in W(R)_L$ is any lifting of x. That $\{\ \}$ respects the action of G_K follows by functoriality. In particular, the action of G_K on $\{\iota(T\mathcal{G})\}$ factors through Γ. For (1) note that

$$[\pi_L][a]\{\iota(v)\} = [a][\pi_L]\{\iota(v)\} = [a]\varphi^r\{\iota(v)\} = \varphi^r([a]\{\iota(v)\}).$$

Since $[a]\{\iota(v)\}$ and $\{\iota(av)\}$ both have image $[a](\iota(v))$ in R, this proves (1). (Here R is viewed as a $\mathcal{O}_K$ algebra via $\mathcal{O}_K \to k$.)
Now the first equality in (2) follows from (1), while the other two equalities follows from the compatibility of ι and $\{\ \}$ with the action of G_K.
Finally, since $\iota(v)$ has positive valuation with respect to the canonical valuation on R, $u \mapsto \{\iota(v)\}$ induces a well defined map $W_L[\![u]\!] \to W(R)_L$. Its image is φ-stable by definition of $\{\ \}$ and Γ-stable by (2). If this map had a non-trivial kernel, then so would its reduction modulo π_L. The latter map $k[\![u]\!] \to R$, sending u to $\iota(v)$ is easily seen to be injective, as $\iota(v)$ has positive valuation. $\square$

(1.3) Write $\mathfrak{S}_L = W_L[\![u]\!]$. We fix an $\mathcal{O}_L$-generator $v \in T_p\mathcal{G}$, and we identify $\mathfrak{S}_L$ with a subring of $W(R)_L$ by sending u to $\{\iota(v)\}$.
Let $\mathcal{O}_{\mathcal{E}}$ denote the p-adic completion of $\mathfrak{S}_L[1/u]$. Then $\mathcal{O}_{\mathcal{E}}$ is a complete discrete valuation ring with uniformizer π_L and residue field $k((u))$. We may view $\mathcal{O}_{\mathcal{E}}$ as a subring of $W(\mathrm{Fr}\, R)_L$. Let $\mathcal{O}_{\mathcal{E}^{\mathrm{ur}}} \subset W(\mathrm{Fr}\, R)_L$ denote the maximal integral, unramified extension of $\mathcal{O}_{\mathcal{E}}$. We denote by $\mathcal{O}_{\widehat{\mathcal{E}}^{\mathrm{ur}}}$ the p-adic completion of $\mathcal{O}_{\mathcal{E}^{\mathrm{ur}}}$, which is again naturally a subring of $W(\mathrm{Fr}\, R)_L$. We write $\mathcal{E}$, $\mathcal{E}^{\mathrm{ur}}$ and $\widehat{\mathcal{E}}^{\mathrm{ur}}$ for the fields of fractions of $\mathcal{O}_{\mathcal{E}}$, $\mathcal{O}_{\mathcal{E}^{\mathrm{ur}}}$ and $\mathcal{O}_{\widehat{\mathcal{E}}^{\mathrm{ur}}}$ respectively. These rings are all stable by φ^r, and by the action of G_K. Moreover the G_K-action on $\mathcal{O}_{\mathcal{E}}$ factors through Γ.

LEMMA (1.4). *The residue field of $\mathcal{O}_{\widehat{\mathcal{E}}^{\mathrm{ur}}}$ is a separable closure of $k((u))$. There is a natural isomorphism*

$$\mathrm{Gal}(\mathcal{E}^{\mathrm{ur}}/\mathcal{E}) \xrightarrow{\sim} \mathrm{Gal}(\bar{K}/K_\infty).$$

Proof. This is a consequence of the theory of norm fields [Wi]. Since Γ is a p-adic Lie group the theory of *loc. cit* applies [Wi, 1.2.2]. For any finite extension F/K write $X_K(F) = \varprojlim(F \cdot K_n)$ where the maps in the inverse limit are given by the norm. We set $X_K(\bar{K}) = \cup_F X_K(F)$ where the limit runs over finite extensions F/K in $\bar{K}$. Then $X_K(F)$ has the structure of a local field of characteristic p, which is a finite separable extension of $X_K(K)$, $X_K(\bar{K})$ is a separable closure of $X_K(K)$, and the functor X_K induces an isomorphism [Wi, 3.2.2]

$$\mathrm{Gal}(X_K(\bar{K})/X_K(K)) \xrightarrow{\sim} \mathrm{Gal}(\bar{K}/K_\infty).$$

On the other hand, there is a natural embedding $X_K(K) \hookrightarrow \mathrm{Fr}\, R$ [Wi ,§4]. To see this explicitly note that one has well defined maps of rings

$$\varprojlim \mathcal{O}_{K_n} \to \varprojlim \mathcal{O}_{K_n}/(v_1) \hookrightarrow \varprojlim \mathcal{O}_{\bar{K}}/\pi_L = R, \tag{1.4.1}$$

where the transition maps in the first two inverse limits are given by the norm, and the final inverse limit by $x \mapsto x^q$.
The image of (1.4.1) is easily seen to be $k[\![u]\!] \subset R$. Hence we may identify $\mathcal{O}_{\mathcal{E}}/\pi_L\mathcal{O}_{\mathcal{E}}$ with $X_K(K)$. It follows that $\mathcal{O}_{\mathcal{E}^{\mathrm{ur}}}/\pi_L\mathcal{O}_{\mathcal{E}^{\mathrm{ur}}} \subset \mathrm{Fr}\, R$ may be identified with $X_K(\bar{K})$. The lemma follows. $\square$

(1.5) Note that the above proof shows that the map ι induces a map

$$T\mathcal{G} \to \varprojlim K_\infty,$$

where the transition maps are given by the norm. This is Coleman's map [Co, Thm. A].
We will write φ_q for the q-Frobenius φ^r (for example on the ring $W(\mathrm{Fr}\, R)$). Now denote by $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}_{\mathcal{E}}}$ (resp. $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$) the category of finite free (resp. finite torsion) $\mathcal{O}_{\mathcal{E}}$-modules M, equipped with an isomorphism $(\varphi_q)^*M \xrightarrow{\sim} M$. We denote by $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}_{\mathcal{E}}}$ (resp. $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$) the category of consisting of a module M in $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}_{\mathcal{E}}}$ (resp. $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$) equipped with a continuous semi-linear action of Γ which commutes with the action of φ_q.
We denote by $\mathrm{Rep}_{G_{K_\infty}}$ (resp. $\mathrm{Rep}^{\mathrm{tor}}_{G_{K_\infty}}$) the category of finite free (resp. finite torsion) $\mathcal{O}_L$-modules V, equipped with a linear action of G_{K_∞} Similarly, we denote by Rep_{G_K} (resp. $\mathrm{Rep}^{\mathrm{tor}}_{G_K}$) the category of finite free (resp. finite torsion) $\mathcal{O}_K$-modules V, equipped with a linear action of G_K.
For M in $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}_{\mathcal{E}}}$ (resp. $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$, resp. $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}_{\mathcal{E}}}$, resp. $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$) we set

$$V(M) = (\mathcal{O}_{\widehat{\mathcal{E}}^{\mathrm{ur}}} \otimes_{\mathcal{O}_{\mathcal{E}}} M)^{\varphi_q=1}.$$

For V in $\mathrm{Rep}_{G_{K_\infty}}$ (resp. $\mathrm{Rep}^{\mathrm{tor}}_{G_{K_\infty}}$ resp. Rep_{G_K} resp. $\mathrm{Rep}^{\mathrm{tor}}_{G_K}$)) we set

$$M_{\mathcal{O}_{\mathcal{E}}}(V) = (V \otimes_{\mathcal{O}_K} \mathcal{O}_{\widehat{\mathcal{E}}^{\mathrm{ur}}})^{G_{K_\infty}}.$$

Theorem (1.6). *V and $M_{\mathcal{O}_{\mathcal{E}}}$ are quasi-inverse equivalences between the exact tensor categories* $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}_{\mathcal{E}}}$ *(resp.* $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$*, resp.* $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}_{\mathcal{E}}}$*, resp.* $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$*) and* $\mathrm{Rep}_{G_{K_\infty}}$ *(resp.* $\mathrm{Rep}^{\mathrm{tor}}_{G_{K_\infty}}$*, resp.* Rep_{G_K}*, resp.* $\mathrm{Rep}^{\mathrm{tor}}_{G_K}$*)*

Proof. The argument for this is identical to that in [Fo 1, 1.2.6, 3.4.3]. For the convenience of the reader we sketch it: It suffices to prove that V and M induce quasi-inverse, exact tensor equivalences between $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$ and $\mathrm{Rep}^{\mathrm{tor}}_{G_{K_\infty}}$.
We first remark that both functors are exact. It suffices to prove this for objects killed by p. For $M_{\mathcal{O}_{\mathcal{E}}}$ this follows from the fact that for M in $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$, $1-\varphi_q$ is étale locally (on $\operatorname{Spec} k((u))$) surjective. For V this is a consequence of Hilbert's theorem 90, and (1.4).
For M in $\mathrm{Mod}^{\varphi_q,\mathrm{tor}}_{/\mathcal{O}_{\mathcal{E}}}$, we have a natural map

$$(M \otimes_{\mathcal{O}_{\mathcal{E}}} \mathcal{O}_{\mathcal{E}^{\mathrm{ur}}})^{\varphi_q=1} \otimes_{\mathcal{O}_K} \mathcal{O}_{\mathcal{E}^{\mathrm{ur}}} \to M \otimes_{\mathcal{O}_{\mathcal{E}}} \mathcal{O}_{\mathcal{E}^{\mathrm{ur}}} \tag{1.6.1}$$

and taking G_{K_∞} invariants of both sides induces a map $M_{\mathcal{O}_{\mathcal{E}}}(V(M)) \to M$. To show that this map is an isomorphism one reduces to the case of objects killed by p, using the exactness proved above. In this case, (1.6.1) is an isomorphism, because étale locally M is spanned by its φ_q-invariants. Similarly, one obtains an isomorphism $V(M_{\mathcal{O}_{\mathcal{E}}}(V)) \to M$ for V in $\mathrm{Rep}^{\mathrm{tor}}_{G_{K_\infty}}$, using dévissage, Hilbert theorem 90, and (1.4). $\square$

§2 (φ_q,Γ)-modules and weakly admissible modules

(2.1) We keep the notation of the previous section, so in particular we write $K_{0,L}$ for the field $K_0 \otimes_{L_0} L = \operatorname{Fr} W_L \subset K$. In order not to overload notation we will write v_n for $(v_n)^*(X) \in \bar{K}$. We now also assume that $K \subset K_{0,L}(v_n)_{n\geq 0}$. Fix an integer $m \geq 1$ such that $K \subset K_{0,L}(v_m)$.
As in [Ki, 1.1.1], denote by $D[0,1)$ the rigid analytic disk of radius 1, over $K_{0,L}$, and denote by u the co-ordinate on $D[0,1)$. For $I \subset [0,1)$ an interval, denote by $D(I) \subset D[0,1)$ the open subspace whose $\bar{K}$ points consist of $x \in \bar{K}$ with $|x| \in I$. We denote by $\mathcal{O}_I$ the ring of rigid analytic functions on $D(I)$, and we write $\mathcal{O} = \mathcal{O}_{[0,1)}$. We will often use the fact that $D[0,1)$ is a p-adic Stein space, so that a coherent sheaf on $D[0,1)$ can be recovered from its global sections. In particular, we may regard a finite free $\mathcal{O}$-module as a coherent sheaf on $D[0,1)$.
We regard $\mathfrak{S}_L \subset \mathcal{O}$ by $u \mapsto u$. The action of φ_q and Γ on $\mathfrak{S}_L$ have a unique continuous extension to $\mathcal{O}$, regarded with its canonical Frechet topology.[1]
Let $Q = [\pi_L^m](u)/[\pi_L^{m-1}](u)$. Denote by $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}}$ the category of finite free $\mathcal{O}$-modules $\mathcal{M}$ equipped with an isomorphism $\varphi_q^*(\mathcal{M})[1/Q] \xrightarrow{\sim} \mathcal{M}[1/Q]$. We denote by $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}}$ the category whose objects consist of an object of $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}}$ equipped with a continuous semi-linear action of Γ such that Γ acts trivially

[1] Contrary to our usual conventions, the symbol $\mathcal{O}_L$ will continue to denote the ring of integers of L, rather than $\mathcal{O} \otimes_{L_0} L$.

on $\mathcal{M}/u\mathcal{M}$ and the φ_q-semi-linear map $1 \otimes \varphi_q : \mathcal{M} \to \mathcal{M}[1/Q]$ commutes with Γ.
We now explain how to differentiate the action of Γ on an object in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma}$ following [Be 1, §IV,V].

LEMMA (2.1.1). *The action of Γ on $\mathcal{O}$, defined above, is continuous. In particular, $\mathcal{O}$ with its action of Γ and φ_q is an object of $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma}$.*

Proof. For $r \in (0,1)$, denote by $|\cdot|_r$ the sup norm on $\mathcal{O}_{[0,r]}$. If $f(u) = \sum_{i \geq 0} a_i u^i \in \mathcal{O}_{[0,r]}$ then $|f|_r = \sup_i |a_i| r^i$. We have to show that, for any r, as $\gamma \to 1$, $|\gamma(f) - f|_r \to 0$, uniformly in f with $|f|_r \leqslant 1$.
Any $\gamma \in \Gamma$ acts on $\mathcal{O}$ by composition with $[\chi(\gamma)]$. Write

$$[\chi(\gamma)] = \sum_{i=1}^{\infty} b_i X^i = \exp_{\mathcal{G}}(\chi(\gamma) \log_{\mathcal{G}} X)$$

where $b_i \in \mathcal{O}_L$, $\log_{\mathcal{G}}$ denotes the logarithm of $\mathcal{G}$ and $\exp_{\mathcal{G}}$ denotes its inverse.[2] Then $b_1 = \chi(\gamma)$ and for $i > 1$, b_i is a polynomial in $\chi(\gamma)$, which vanishes at $\chi(\gamma) = 1$. Given $\epsilon > 0$, choose i_0 so that $r^{i_0} < \epsilon$. Then for γ sufficiently close to 1, $|b_i| < \epsilon$ for $1 < i < i_0$, and $|b_1 - 1| < \epsilon$ so $|[\chi(\gamma)](u) - u|_r < \epsilon$. Hence

$$|\gamma(f) - f|_r = |f([\chi(\gamma)](u)) - f(u)|_r \leqslant \sum_{i=1}^{\infty} |a_i([\chi(\gamma)](u)^i - u^i)|_r \leqslant \epsilon |f(u)|_r.$$

The lemma follows. $\square$

LEMMA (2.1.2). *Let $\mathcal{M}$ be in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma}$. For each $r \in (0,1)$ and $\gamma \in \Gamma$ sufficiently close to 1 (depending on r) the series*

$$\log\gamma = \sum_{i=1}^{\infty} (\gamma - 1)^i (-1)^{i-1}/i$$

induces a well defined operator on $\mathcal{M}|_{D[0,r)}$. This induces a well defined $\mathbb{Z}_p$-linear map of Lie algebras

$$d\Gamma_{\mathcal{M}} : \mathrm{Lie}\,\Gamma \to \mathrm{End}_{K_0}\mathcal{M}; \quad \beta \mapsto \log(\exp\beta).$$

such that for $\beta \in \mathrm{Lie}\,\Gamma$, $d\Gamma_{\mathcal{O}}(\beta)$ is a derivation and $d\Gamma_{\mathcal{M}}(\beta)$ is a differential operator over $d\Gamma_{\mathcal{O}}(\beta)$. That is, for $m \in \mathcal{M}$, $f \in \mathcal{O}$ and $\beta \in \mathrm{Lie}\,\Gamma$,

$$d\Gamma_{\mathcal{M}}(\beta)(fm) = d\Gamma_{\mathcal{O}}(\beta)(f)m + f d\Gamma_{\mathcal{M}}(\beta)(m).$$

Proof. Let $M_0 \subset \mathcal{M}$ be finite free W_L-submodule of rank equal to $d = \mathrm{rk}_{\mathcal{O}}\mathcal{M}$, which spans $\mathcal{M}$. Choosing a basis for M_0, we may identify $\mathcal{M}$ with $\mathcal{O}^d$. As

[2]So if $\mathcal{G} = \mathbb{G}_m$, then $\log_{\mathcal{G}}(X) = \log(1 + X)$.

in (2.1.1), choose $r \in (0,1)$ and denote by $|\,|_r$ the norm on $\mathcal{M}_{D[0,r]} = \mathcal{O}^d_{D[0,r]}$ induced by the sup norm on $\mathcal{O}_{D[0,r]}$. For any $\epsilon > 0$, and γ sufficiently small we have $|\gamma(m) - m|_r \leqslant \epsilon |m|_r$ for $m \in M_0$ and $|\gamma(f) - f|_r \leqslant \epsilon |f|_r$ for $f \in \mathcal{O}_{D[0,r]}$ by (2.1.1). Hence

$$|\gamma(mf) - mf|_r \leqslant |\gamma(m) - m|_r |\gamma(f)|_r + |\gamma(f) - f|_r |m|_r \leqslant 2\epsilon |m|_r |f|_r = 2\epsilon |fm|_r.$$

This shows that $\log\gamma$ is well defined.
It follows that the map

$$d\Gamma_{\mathcal{M}} =: \mathrm{Lie}\,\Gamma \to \mathrm{End}_{K_0}\mathcal{M}; \quad \beta \mapsto \log(\exp\beta)$$

is well defined for β sufficiently small, and we extend it to all of $\mathrm{Lie}\,\Gamma$ by $\mathbb{Z}_p$-linearity. That $d\Gamma_{\mathcal{O}}(\beta)$ is a derivation and $d\Gamma_{\mathcal{M}}(\beta)$ is a differential operator over $d\Gamma_{\mathcal{O}}(\beta)$ follows from a simple computation, as does the fact that $d\Gamma_{\mathcal{M}}$ is a map of Lie algebras. Note that the latter statement just means that the differential operators $d\Gamma_{\mathcal{M}}(\beta)$ for $\beta \in \mathrm{Lie}\,\Gamma$ commute. $\square$

(2.1.3) We say that $\mathcal{M}$ in $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}}$ is $\mathcal{O}_L$-*analytic* if the map $d\Gamma_{\mathcal{M}}$ is $\mathcal{O}_L$-linear, not just $\mathbb{Z}_p$-linear. We denote by $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathcal{O}}$ the full subcategory of $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}}$ consisting of $\mathcal{O}_L$-analytic objects. One checks easily that this is a $\otimes$-subcategory, which is stable under taking subobjects and quotients.

Lemma (2.1.4). *Let $\mathcal{M}$ be in $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathcal{O}}$.*

(1) *For each $r \in (0,1)$ the operator $N_\nabla =: \log\gamma/\log\chi(\gamma)$ is well defined for $\gamma \neq 1$ sufficiently close to 1, and is independent of γ.*
(2) *The operators in (1) induce a $K_{0,L}$-linear map $N_\nabla : \mathcal{M} \to \mathcal{M}$, which is a differential operator over the derivation $N_\nabla : \mathcal{O} \to \mathcal{O}$, and which commutes with φ_q on $\mathcal{M}$.*
(3) *There is a singular connection ∇ on $\mathcal{M}$ with simple poles at the zeroes of $[\pi_L^n]/u$ for $n \geq 1$ (that is at the non-trivial π_L-power torsion points of $\mathcal{G}$) such that $N_\nabla = \langle \nabla, \frac{\partial F}{\partial Y}(u,0)\log_{\mathcal{G}} u \cdot d/du \rangle$, where $F(X,Y)$ denotes the formal group law of $\mathcal{G}$ with respect to X, and $\frac{\partial F}{\partial Y}(u,0) \in \mathcal{O}_L[\![u]\!]^\times$.*

Proof. For γ sufficiently close to 1, we may write $\gamma = \exp\beta$ with $\beta \in \mathrm{Lie}\,\Gamma$. Since $\beta \mapsto \log(\exp\beta)$ is $\mathcal{O}_L$-linear by assumption. and $\beta \mapsto \log(\chi(\exp(\beta)))$ is obviously $\mathcal{O}_L$-linear, $\log\gamma/\log\chi(\gamma)$ is independent of γ. This proves (1) and (2) follows by viewing $\mathcal{M}$ as a coherent module on $D[0,1)$. The fact that N_∇ commutes with φ_q follows from the fact that φ_q commutes with the action of Γ.

To see (3), we first compute the derivation N_∇ on $\mathcal{O}$. For $\gamma \in \Gamma$ write $a_\gamma = \chi(\gamma) - 1$. Then

$$\begin{aligned} N_\nabla(u) &= \lim_{\gamma \mapsto 1} \frac{[\chi(\gamma)](u) - u}{\log\chi(\gamma)} = \lim_{\mathfrak{a}_\gamma \to 0} \frac{\exp_{\mathcal{G}}((1+a_\gamma)\log_{\mathcal{G}} u) - u}{\log\chi(\gamma)} \\ &= \lim_{\mathfrak{a}_\gamma \to 0} \frac{F(u, \exp_{\mathcal{G}}(a_\gamma \log_{\mathcal{G}} u)) - u}{\log\chi(\gamma)} = \frac{\partial F}{\partial Y}(u,0)\log_{\mathcal{G}}(u). \end{aligned}$$

Hence N_∇ is given on $\mathcal{O}$ by $N_\nabla(f) = \frac{\partial F}{\partial Y}(u,0)(\log_{\mathcal{G}} u)\frac{df}{du}$. As $\frac{\partial F}{\partial Y}(u,0)$ has constant term 1, and coefficients in $\mathcal{O}_L[\![u]\!]$, it is a unit $\mathcal{O}$.
Now for any $\mathcal{M}$, define $\nabla(m)$ for $m \in \mathcal{M}$ by $\nabla(m) = (\frac{\partial F}{\partial Y}(u,0)\log_{\mathcal{G}} u)^{-1} N_\nabla(m)$. Since N_∇ on $\mathcal{M}$ is a differential operator over the derivation $\frac{\partial F}{\partial Y}(u,0)(\log_{\mathcal{G}} u)\frac{df}{du}$ on $\mathcal{O}$, ∇ is a (singular) connection. A priori $\nabla(m)$ has a simple pole at each $[\pi_L]$-torsion point of $\mathcal{G}$, however since the action of Γ on $\mathcal{M}$ is trivial mod u, the operator N_∇ vanishes mod u, and $\nabla(m)$ has no pole at $u=0$. This proves (3). $\square$

(2.2) Denote by $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ the category of finite dimensional $K_{0,L}$-vector spaces D equipped with an isomorphism $\varphi_q^* D \xrightarrow{\sim} D$ and a decreasing, separated filtration on $D_K = D \otimes_{K_{0,L}} K$, indexed by $\mathbb{Z}$, by K-subspaces.
Our next task to to show that there is an exact $\otimes$-equivalence between $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ and $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathcal{O}}$. The construction is analogous to that in [Be 2] and [Ki, 1.2]. Since many of the proofs from [Ki] go over verbatim, we often only sketch the argument.[3]
For $n \geq 0$, denote by $\widehat{\mathfrak{S}}_n$ the complete local ring at the point x_n of $D[0,1)$, corresponding to $u = v_{m+n}$. That is, $\widehat{\mathfrak{S}}_n$ is the completion of the localization of $\mathcal{O}$ at the maximal ideal generated by $[\pi_L^{m+n}](u)/[\pi_L^{m+n-1}](u)$. Then $\widehat{\mathfrak{S}}_n$ is a discrete valuation ring with residue field $K_{m+n} = K(v_{m+n}) \supset K$, which is canonically a subfield of $\widehat{\mathfrak{S}}_n$. In particular, $u - v_{m+n}$ is a uniformizer for $\widehat{\mathfrak{S}}_n$. Let

$$\lambda = \prod_{n\geq 0} \varphi_q^n(Q(u)/Q(0))) = \prod_{n\geq 0} [\pi_L^{m+n}](u)/[\pi_L^{m+n-1}](u)\pi_L,$$

and write $\varphi_{q,W_L} : \mathcal{O} \to \mathcal{O}$ for the $\mathcal{O}_L$-linear automorphism given by applying φ^r to the coefficients of a series in $\mathcal{O}$.
Given D in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ and $n \geq 0$, we denote by ι_n the composite

$$\iota_n : D \otimes_{K_{0,L}} \mathcal{O}[1/\lambda] \overset{\varphi_q^{-n} \otimes \varphi_{q,W_L}^{-n}}{\hookrightarrow} D \otimes_{K_{0,L}} \widehat{\mathfrak{S}}_n[1/\lambda] \xrightarrow{\sim} D_K \otimes_K \widehat{\mathfrak{S}}_n[1/u - v_{m+n}].$$

We set

$$\mathcal{M}(D) = \{d \in D \otimes_{K_{0,L}} \mathcal{O}[1/\lambda] : \forall n \geq 0,\ \iota_n(d) \in \mathrm{Fil}^0(D_K \otimes_K \widehat{\mathfrak{S}}_n[1/u - v_{m+n}])\}.$$

Lemma (2.2.1). *For D in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$, $\mathcal{M}(D)$ is naturally an object of $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathcal{O}}$.*

Proof. That $\mathcal{M}(D)$ is a finite free $\mathcal{O}$-module, and the fact that φ_q on $D \otimes_{K_{0,L}} \mathcal{O}[1/\lambda]$ induces an isomorphism $\varphi_q^*(\mathcal{M}(D))[1/Q] \xrightarrow{\sim} \mathcal{M}(D)[1/Q]$ is proved exactly as in [Ki, 1.2.2].

[3] In fact they often simplify since one only has to consider the case when $N=0$ in [Ki].

Note that for $\gamma \in \Gamma$,

$$\gamma([\pi_L](u)) = [\pi_L] \circ [\chi(\gamma)](u) = [\pi_L \chi(\gamma)](u).$$

Hence $\gamma(\lambda) = \lambda \circ [\chi(\gamma)]$ has a simple zero at each $[\pi_L^m]$-torsion point which is not a $[\pi_L^{m-1}]$-torsion point. It follows that $\lambda/\gamma(\lambda) \in \mathfrak{S}_L[1/p]^\times$. In particular, if $\gamma \in \Gamma$ acts on $D \otimes_{K_{0,L}} \mathcal{O}$ by $1 \otimes \gamma$, then this induces an action of Γ on $D \otimes_{K_{0,L}} \mathcal{O}[1/\lambda]$. The same argument shows that γ induces an automorphism of $\widehat{\mathfrak{S}}_n$ for $n \geq 0$. As $\varphi_{q,W_L}[\chi(\gamma)] = [\chi(\gamma)]$, one sees that $\mathcal{M}(D)$ is stable by the action of Γ. Finally, this action is $\mathcal{O}_L$-analytic, as the action of Γ on $\mathcal{O}$ is $\mathcal{O}_L$-analytic. □

Lemma (2.2.2). *Let $\mathcal{M}$ be in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$. There exists a unique $K_{0,L}$-linear section $\xi : \mathcal{M}/u\mathcal{M} \to \mathcal{M}[1/\lambda]$ such that the elements of $\xi(\mathcal{M}/u\mathcal{M})$ are Γ-invariant. Moreover, we have*

(1) *ξ is φ_q-equivariant.*
(2) *ξ induces an isomorphism*

$$\mathcal{M}/u\mathcal{M} \otimes_{K_{0,L}} \mathcal{O}[1/\lambda] \xrightarrow{\sim} \mathcal{M}[1/\lambda].$$

(3) *The image of $\xi \otimes 1 : \mathcal{M}/u\mathcal{M} \otimes_{K_{0,L}} \mathcal{O} \to \mathcal{M}[1/\lambda]$ coincides with $(1 \otimes \varphi_q)(\varphi_q^*\mathcal{M})$ over an admissible open neighborhood of $u = v_m$.*

Proof. Consider the connection ∇ on $\mathcal{M}$ defined in (2.1.4)(3). For $r \in (0,1)$ sufficiently small there exists a unique ∇-parallel section $\xi_r : \mathcal{M}/u\mathcal{M} \to \mathcal{M}|_{D[0,r)}$. Since N_∇ commutes with φ_q, the section $\varphi_q \circ \xi_r \circ \varphi_q^{-1}$ is also ∇-parallel, and hence equal to ξ_r. Hence ξ_r is φ_q-invariant. Similarly $\gamma \circ \xi_r \circ \gamma^{-1}$ is ∇-parallel for $\gamma \in \Gamma$, so ξ_r is Γ-invariant.
Now ξ may be constructed from ξ_r exactly as in [Ki, 1.2.6], by repeatedly pulling ξ_r back by φ_q^* and using the isomorphism $\varphi_q^*\mathcal{M}[1/Q] \xrightarrow{\sim} \mathcal{M}[1/Q]$. The claims (2) and (3) also follow exactly as in *loc. cit.* □

(2.2.3) Suppose that $\mathcal{M}$ is in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$. For i an integer denote by $\mathrm{Fil}^i\varphi_q^*\mathcal{M}$ the preimage of $Q^i\mathcal{M}$ under $\varphi_q^*\mathcal{M}[1/Q] \xrightarrow{\sim} \mathcal{M}[1/Q]$. Note that this filtration is Γ-stable. Let $D(\mathcal{M}) = \mathcal{M}/u\mathcal{M}$. By (2.2.2), ξ induces an isomorphism $D(\mathcal{M}) \otimes_{K_{0,L}} \mathcal{O} \xrightarrow{\sim} \varphi^*(\mathcal{M})$ near the point $u = v_m$. Hence we obtain an isomorphism

$$D(\mathcal{M}) \otimes_{K_{0,L}} K(v_m) \xrightarrow{\sim} \varphi_q^*(\mathcal{M})/Q\varphi_q^*(\mathcal{M}). \tag{2.2.4}$$

Give the right hand side of (2.2.4) the filtration induced by that on $\varphi_q^*\mathcal{M}$, and pull this filtration back to $D(\mathcal{M}) \otimes_{K_{0,L}} K(v_m)$. This gives rise to a Γ-stable filtration on $D(\mathcal{M}) \otimes_{K_{0,L}} K(v_m)$, which necessarily descends to a filtration on $D(\mathcal{M})_K$. This gives $D(\mathcal{M})$ the structure of an object in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$.

Lemma (2.2.5). *Let $\mathcal{M}$ be in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$ and $D = D(\mathcal{M})$. Then for all $i \in \mathbb{Z}$ the map ξ induces an isomorphism*

$$\sum_{j\geq 0} Q^j \widehat{\mathfrak{S}}_0 \otimes_K \mathrm{Fil}^{i-j} D(\mathcal{M})_K \xrightarrow{\sim} \widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} \mathrm{Fil}^i \varphi^*(\mathcal{M}).$$

Proof. This is the analogue of [Ki, 1.2.12(4)] in our situation, and the proof is identical, so we only sketch it here. Since $D(\mathcal{M})_K \otimes \widehat{\mathfrak{S}}_0$ and $\widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} \varphi_q^*(\mathcal{M})$ induce the same filtration on their common quotient $D(\mathcal{M})_K$, one sees easily that it suffices to check that for all $i \in \mathbb{Z}$,

$$\xi(\mathrm{Fil}^i D(\mathcal{M})_K) \subset \widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} (1 \otimes \varphi_q)(\mathrm{Fil}^i \varphi^*(\mathcal{M})).$$

We will identify $\varphi_q^*\mathcal{M}$ with its image $(1 \otimes \varphi_q)(\varphi_q^*\mathcal{M})$ in $\mathcal{M}[1/Q]$. An element $d \in \xi(\mathrm{Fil}^i D(\mathcal{M})_K)$ can be written as $d = d_0 + d_1$ with $d_0 \in \widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} \mathrm{Fil}^i \varphi_q^*(\mathcal{M})$ and $d_1 \in Q\widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} \varphi_q^*(\mathcal{M})$. As $N_\nabla(d) = 0$, we have

$$N_\nabla(d_1) = -N_\nabla(d_0) \in \widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} (\mathrm{Fil}^i \varphi_q^*(\mathcal{M}) \cap Q\varphi_q^*(\mathcal{M})) =: M_i.$$

Thus it suffices to show that for all $i \in \mathbb{Z}$, N_∇ induces a bijection on M_i, for then $d_1 \in M_i \subset \widehat{\mathfrak{S}}_0 \otimes_{\mathcal{O}} \mathrm{Fil}^i \varphi_q^*(\mathcal{M})$. For i sufficiently small this follows from the isomorphism $D(\mathcal{M}) \otimes_{\mathcal{O}} \widehat{\mathfrak{S}}_0 \xrightarrow{\sim} \varphi_q^*\mathcal{M} \otimes_{\mathcal{O}} \widehat{\mathfrak{S}}_0$. The general case follows by descending induction on i and an application of the snake lemma. $\square$

Proposition (2.2.6). *The functors $\mathcal{M}$ and D between $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$ and $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$ are quasi-inverse, exact, $\otimes$-equivalences.*

Proof. Let D be in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$. From the definition of $\mathcal{M}(D)$, there is a natural Γ-equivariant inclusion $D \subset \mathcal{M}(D)[1/Q]$, which induces an isomorphism of D with $D(\mathcal{M}(D)) = \mathcal{M}(D)/u\mathcal{M}(D)$. Hence the image of this inclusion coincides with $\xi(D(\mathcal{M}(D)))$, and one sees from the definitions that the filtration on D_K coincides with the one on $D(\mathcal{M}(D))_K$. This produces a natural isomorphism $D \xrightarrow{\sim} D(\mathcal{M}(D))$.

Conversely, let $\mathcal{M}$ be in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$. Then both $\mathcal{M}$ and $\mathcal{M}(D(\mathcal{M}))$ may be identified with $\mathcal{O}$-submodules of $D(\mathcal{M}) \otimes_{K_{0,L}} \mathcal{O}[1/\lambda]$. At any point of $D[0,1)$ other $u = v_n$, $n \geq m$, both submodules coincide with $D(\mathcal{M}) \otimes_{K_{0,L}} \mathcal{O}[1/\lambda]$. Since both $\mathcal{M}$ and $\mathcal{M}(D(\mathcal{M}))$ are in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma}$, to show these submodules are equal it suffices to check that these two submodules coincide at $u = v_m$. This follows from (2.2.5). (cf. [Ki, 1.2.13]). Hence we have a natural isomorphism $\mathcal{M}(D(\mathcal{M})) \xrightarrow{\sim} \mathcal{M}$.

That $\mathcal{M}$ and D are exact follows from (2.2.5). One checks easily that $\mathcal{M}$ and D respect $\otimes$-products (cf. [Ki, 1.2.15]). $\square$

(2.3) We now apply Kedlaya's slope filtration as in [Be 2] and [Ki, §1.3] to show that an object $\mathcal{M}$ of $\mathrm{Mod}_{/\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$ can be descended to $\mathfrak{S}_L$ if and only if $D(\mathcal{M})$ is weakly admissible.[4] Again, as many of the arguments are identical to those of [Ki] we sometimes only sketch the proofs.

Let $\mathcal{R} = \lim_{r\to 1^-} \mathcal{O}_{(r,1)}$ denote the Robba ring, and $\mathcal{R}^b = \lim_{r\to 1^-} \mathcal{O}^b_{(r,1)}$, where $\mathcal{O}^b_{(r,1)} \subset \mathcal{O}_{(r,1)}$ denotes the subring of bounded functions. Then $\mathcal{R}^b$ is a discrete valuation field, with valuation

$$v_{\mathcal{R}^b}(f) = -\log_{\pi_L} \lim_{r\to 1^-} \sup_{x\in D(r,1)} |f(x)|$$

and uniformizer π_L. The endomorphism φ_q and the derivation N_∇ of $\mathcal{O}$ induce an automorphism and a derivation respectively of $\mathcal{R}$ and $\mathcal{R}^b$, which we will again denote by φ_q and N_∇.

Denote by $\mathrm{Mod}_{/\mathcal{R}}^{\varphi_q}$ (resp. $\mathrm{Mod}_{/\mathcal{R}^b}^{\varphi_q}$) the category of finite free $\mathcal{R}$-modules (resp. $\mathcal{R}^b$-modules) $\mathcal{M}$ equipped with an isomorphism $\varphi_q^*\mathcal{M} \xrightarrow{\sim} \mathcal{M}$. For an $\mathcal{M}$ in $\mathrm{Mod}_{/\mathcal{R}}^{\varphi_q}$, we denote by

$$0 = \mathcal{M}_0 \subset \mathcal{M}_1 \subset \cdots \subset \mathcal{M}_r = \mathcal{M}$$

Kedlaya's slope filtration [Ke 1], [Ke 2]. We write s_i for the slope of the pure slope quotient $\mathcal{M}_i/\mathcal{M}_{i-1}$, which is finite free over $\mathcal{R}$. The filtration is functorial for maps in $\mathrm{Mod}_{/\mathcal{R}}^{\varphi_q}$. One of Kedlaya's results about the filtration says that a module of pure slope s has a canonical descent to a module $\mathcal{M}^b$ in $\mathrm{Mod}_{/\mathcal{R}^b}^{\varphi_q}$ which has slope s in the sense of Dieudonné-Manin theory (and the valuation on $\mathcal{R}^b$ normalized so that $v(\pi_L) = 1$).

For $\mathcal{M}$ in $\mathrm{Mod}_{/\mathcal{O}}^{\varphi,\Gamma,\mathrm{an}}$ we write $\mathcal{M}_{\mathcal{R}} = \mathcal{M} \otimes_{\mathcal{O}} \mathcal{R}$. The operators φ_q and N_∇ on $\mathcal{M}$ induce operators φ_q and N_∇ on $\mathcal{M}_{\mathcal{R}}$.

Lemma (2.3.1). *Let $\mathcal{M}$ be in $\mathrm{Mod}_{\mathcal{O}}^{\varphi_q,\Gamma,\mathrm{an}}$. The slope filtration on $\mathcal{M}_{\mathcal{R}}$ is induced by a unique filtration on $\mathcal{M}$ by saturated, finite free $\mathcal{O}$-submodules. This filtration on $\mathcal{M}$ is stable by φ_q and the action of Γ.*

Proof. It is clear that a such filtration on $\mathcal{M}$, if it exists is unique and stable by φ_q. The functoriality of the slope filtration on $\mathcal{M}_{\mathcal{R}}$ implies that it is stable by Γ, and hence so is the filtration on $\mathcal{M}$.

It remains to show the existence of such a filtration. As $\varphi_q(\lambda) = \pi_L[\pi_L^{m-1}]/[\pi_L^m]\lambda$, for any integer s, the slope filtration on $\lambda^{-s}\mathcal{M}_{\mathcal{R}}$ is given by $\lambda^{-s}\mathcal{M}_{\mathcal{R},i}$, and the slopes of $\lambda^{-s}\mathcal{M}_{\mathcal{R}}$ are those of $\mathcal{M}_{\mathcal{R}}$ shifted by $-s$. Since $[\pi_L^m]/[\pi_L^{m-1}]\pi_L$ has a unique, simple zero on $D[0,1)$ at x_0, we may replace $\mathcal{M}$ by $\lambda^{-s}\mathcal{M}$ for s sufficiently large, and assume that φ_q induces a map $\varphi_q^*\mathcal{M} \to \mathcal{M}$. We first show that the slope filtration is induced by a filtration on $\mathcal{M}|_{D(0,1)}$ by saturated $\mathcal{O}_{(0,1)}$-submodules. For some r_0 sufficiently close to 1, the slope

[4]The idea of relating Kedlaya's slope filtration to the condition of weak admissibility comes from Berger's beautiful paper [Be 2], however our treatment here is closer to that of [Ki].

filtration on $\mathcal{M}_{\mathcal{R}}$ is induced by a filtration on $\mathcal{M}|_{D(r_0,1)}$ by saturated $\mathcal{O}_{(r_0,1)}$-submodules. Let $r_0 > r_1 > \dots$ be a sequence approaching 0, and such that $\varphi_q^{-1}(D(r_i,1)) \subset D(r_{i-1},1)$ for $i \geq 1$. The same argument as in [Ki, 1.3.4] shows that for $j \geq 0$, $\mathcal{M}_{\mathcal{R},i}$ is induced by a filtration on $\mathcal{M}|_{D(r_j,1)}$ by closed, saturated $\mathcal{O}_{(r_j,1)}$-submodules, and hence by a filtration on $\mathcal{M}|_{D(0,1)}$ by closed, saturated $\mathcal{O}_{(0,1)}$-submodules.

The filtration on $\mathcal{M}_{D(0,1)}$ is stable by Γ, by uniqueness, and hence it is stable by N_∇. Consider the operator $\partial = \langle \nabla, -u\frac{d}{du}\rangle$ on $\mathcal{M}$. This is well defined in a neighbourhood of 0, and preserves the filtration on $\mathcal{M}|_{D(0,1)}$ over this neighbourhood as N_∇ does. Hence the filtration on $\mathcal{M}|_{D(0,1)}$ is induced by a filtration on $\mathcal{M}$ by closed, saturated $\mathcal{O}$-submodules by [Ki, 1.3.5]. $\square$

(2.3.2) Let D in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ be 1-dimensional over $K_{0,L}$. Choose a basis vector v for D, and set $t_{N,L}(D) = v_{\pi_L}(\alpha)$ where $\alpha \in K_{0,L}$ satisfies $\varphi_q(v) = \alpha v$. We write $t_{H,L}(D)$ for the unique integer i such that $\mathrm{gr}^i D_K$ is non-zero. For D of arbitrary dimension d, we set $t_{N,L}(D) = t_{N,L}(\wedge^d D)$ and $t_{H,L}(D) = t_{H,L}(\wedge^d D)$. We will say that D is *weakly admissible* if the usual conditions of weak admissibility are satisfied with these invariants in place of the usual ones. That is if $t_{H,L}(D) = t_{N,L}(D)$ and $t_{H,L}(D') \leqslant t_{N,L}(D')$ for all φ_q-stable submodules $D' \subset D$.

Proposition (2.3.3). *Let D be in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ and $\mathcal{M} = \mathcal{M}(D)$. Then D is weakly admissible if and only if $\mathcal{M}_{\mathcal{R}}$ is pure of slope* 0.

Proof. Suppose first that $\dim_{K_{0,L}} D = 1$. Let v be a basis vector for D, and write $\varphi(v) = \alpha v$ for some $\alpha \in K_{0,L}^\times$. From the definition of $\mathcal{M}(D)$ one finds that $\mathcal{M}(D) = \lambda^{-t_{H,L}(D)}(D \otimes_{K_{0,L}} \mathcal{O})$, so

$$\varphi_q(\lambda^{-t_{H,L}(D)}e) = ([\pi_L^{m-1}]\pi_L/[\pi_L^m])^{-t_{H,L}(D)}\alpha\lambda^{-t_{H,L}(D)}e$$

As, $[\pi_L^m] = [\pi_L] \circ [\pi_L^{m-1}]$, we have that $[\pi_L^m]/[\pi_L^{m-1}] \in \mathfrak{S}_L$, is an element whose reduction modulo π_L is $u^{q^m - q^{m-1}}$. Hence $[\pi_L^m]/[\pi_L^{m-1}]$ is a unit in $\mathcal{R}^b$. It follows that $\mathcal{M}(D)$ has slope

$$v_{\pi_L}(\alpha) - t_{H,L}(D) = t_{N,L}(D) - t_{H,L}(D).$$

This proves the proposition when D has dimension 1. The general case follows from exactly the same argument as in [Ki, 1.3.8], using the equivalence (2.2.6) and (2.3.1). $\square$

(2.4) Denote by $\mathrm{Mod}^{\varphi_q}_{/\mathfrak{S}_L}$ the category consisting of finite free $\mathfrak{S}_L$-modules $\mathfrak{M}$ equipped with an isomorphisms $1 \otimes \varphi_q : \varphi_q^*\mathfrak{M}[1/Q] \xrightarrow{\sim} \mathfrak{M}[1/Q]$. We denote by $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathfrak{S}_L}$ the category whose objects consist of an object of $\mathrm{Mod}^{\varphi_q}_{/\mathfrak{S}_L}$ equipped a semi-linear action of Γ on $\mathfrak{M}$ which commutes with the action of φ_q, and such that Γ acts trivially on $\mathfrak{M}/u\mathfrak{M}$.

We denote by $\mathrm{Mod}^{\varphi_q,0}_{/\mathcal{O}}$ (resp. $\mathrm{Mod}^{\varphi_q,\Gamma,0}_{/\mathcal{O}}$) the full subcategory of $\mathrm{Mod}^{\varphi_q}_{/\mathcal{O}}$ (resp. $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}}$) consisting of objects $\mathcal{M}$ such that $\mathcal{M}_{\mathcal{R}} = \mathcal{M} \otimes_{\mathcal{O}} \mathcal{R}$ is pure of slope 0.

LEMMA (2.4.1). *There is an equivalence of $\otimes$-categories*

$$\mathrm{Mod}^{\varphi_q}_{/\mathfrak{S}_L} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \to \mathrm{Mod}^{\varphi_q,0}_{/\mathcal{O}}; \quad \mathfrak{M} \mapsto \mathfrak{M} \otimes_{\mathrm{W}_L[\![u]\!]} \mathcal{O}$$

where the left hand side means the category obtained from $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathfrak{S}_L}$ *by applying* $\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$ *to the* Hom *groups.*

Proof. This is identical to [Ki, 1.3.13]. □

COROLLARY (2.4.2). *There is an equivalence of exact $\otimes$-categories*

$$\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathfrak{S}_L} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \to \mathrm{Mod}^{\varphi_q,\Gamma,0}_{/\mathcal{O}}; \quad \mathfrak{M} \mapsto \mathfrak{M}/u\mathfrak{M}.$$

Proof. This follows from (2.4.1), since the action of $\gamma \in \Gamma$ can be thought of as an isomorphism $\gamma^*(\mathfrak{M}) \xrightarrow{\sim} \mathfrak{M}$ for $\mathfrak{M}$ in $\mathrm{Mod}^{\varphi_q}_{/\mathfrak{S}_L}$ or $\mathrm{Mod}^{\varphi_q,\Gamma,0}_{/\mathcal{O}}$. □

(2.4.3) We denote by $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathfrak{S}_L}$ the full subcategory of $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathfrak{S}_L}$ such that the corresponding object in $\mathrm{Mod}^{\varphi_q,\Gamma}_{/\mathcal{O}}$ is in $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathcal{O}}$

COROLLARY (2.4.4). *There is an exact, fully faithful $\otimes$-functor*

$$\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{/\mathfrak{S}_L} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \to \mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$$

whose essential image consists of the weakly admissible modules in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$.

Proof. This follows by combining (2.4.2), (2.2.6) and (2.3.3). □

§3 (φ_q, Γ)-MODULES AND CRYSTALLINE REPRESENTATIONS

(3.1) Recall the ring $R = \varprojlim \mathcal{O}_{\bar{K}}/p$ introduced in (1.1). Denote by $B_{\mathrm{dR}}, B_{\mathrm{cris}} \supset W(R)$ the usual rings introduced by Fontaine [Fo 2]. Write $B_{\mathrm{cris},L} = B_{\mathrm{cris}} \otimes_{L_0} L$. We write φ_q for the L-linear extension of the operator φ^r on B_{cris}. Note that we have an embedding

$$B_{\mathrm{cris},L} \otimes_{K_{0,L}} K \xrightarrow{\sim} B_{\mathrm{cris}} \otimes_{K_0} K \hookrightarrow B_{\mathrm{dR}}. \tag{3.1.1}$$

For D in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ this gives rise to an embedding

$$B_{\mathrm{cris},L} \otimes_{K_{0,L}} D_K \xrightarrow{\sim} B_{\mathrm{cris}} \otimes_{K_0} D_K \hookrightarrow B_{\mathrm{dR}} \otimes_K D_K. \tag{3.1.2}$$

We say an element in $B_{\mathrm{cris},L} \otimes_{K_{0,L}} D$ is in Fil^0 if its image in $B_{\mathrm{dR}} \otimes_K D_K$ under (3.1.2) is. Dually, we say $K_{0,L}$-linear map $D \to B_{\mathrm{cris},L}$ is compatible with filtrations if the map $D_K \to B_{\mathrm{dR}}$ induced by (3.1.1) is compatible with filtrations.

For D in $\mathrm{Mod}_{K_{0,L}}^{F,\varphi_q}$ let

$$V_L(D) = \mathrm{Fil}^0(B_{\mathrm{cris},L} \otimes_{K_{0,L}} D)^{\varphi_q=1}$$

and

$$V_L^*(D) = \mathrm{Hom}_{\varphi_q,\mathrm{Fil}}(D, B_{\mathrm{cris},L}).$$

There is a canonical isomorphism $V_L^*(D) \xrightarrow{\sim} V_L(D^*)$, where D^* denotes the dual of D in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$.

We will prove an analogue for the category $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$ of the result that weakly admissible modules are admissible.

Lemma (3.1.3). *Let D in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$ be of $K_{0,L}$-dimension 1 such that $t_{H,L}(D) \leqslant t_{N,L}(D)$. Then $V_L(D) = 0$ unless $t_{H,L}(D) = t_{N,L}(D)$, in which case $\dim_L V_L(D) = 1$.*

Proof. Give $D(\mathcal{G}) = K_{0,L}$ the structure of an object of $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$ by equipping it with a φ_q semi-linear automorphism given by sending 1 to π_L^{-1} and defining $\mathrm{Fil}^i D(\mathcal{G})_K = D(\mathcal{G})_K$ if $i \leqslant -1$ and 0 otherwise. Then by [Col 1, Prop. 9.19]

$$V_L(D(\mathcal{G})) = \mathrm{Fil}^0(D(\mathcal{G}) \otimes_{K_{0,L}} B_{\mathrm{cris},L})^{\varphi_q=1} = \mathrm{Fil}^1 B_{\mathrm{cris},L}^{\varphi_q=\pi_L} = t_L \cdot L,$$

where $t_L = \log_{\mathcal{G}} u$, is a unit in $B_{\mathrm{cris},L}$ [Col 1, Prop. 9.10, 9.17].

Let D in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$ be of $K_{0,L}$-dimension 1 such that $t_{H,L}(D) \leqslant t_{N,L}(D)$. Let $d \in D$ be a $K_{0,L}$-basis vector. Since multiplication by t_L^j induces a bijection of $V_L(D)$ with $V_L(D \otimes D(\mathcal{G})^j)$, we may assume that $\varphi_q(d) = \alpha d$ with $\alpha \in W_L^\times$. Furthermore, if $\bar{k}$ denotes the residue field of $\bar{K}$, then there exists $\beta \in W(\bar{k})_L^\times$ such that $\varphi_q(\beta) = \alpha\beta$, so we may assume that $\alpha = 1$. Then $t_{N,L}(D) = 0 \geq t_{H,L}(D)$,

$$V_L(D) = \mathrm{Fil}^{-t_{H,L}(D)} B_{\mathrm{cris},L}^{\varphi_q=1}$$

and the lemma follows from [Col 1, Lem. 9.14]. $\square$

Lemma (3.1.4). *Suppose that D in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$ is weakly admissible. Then*

$$\dim_L V_L(D) \leqslant \dim_{K_{0,L}} D.$$

Proof. This is very similar to [CF, Prop. 4.5]. Denote by $C_{\mathrm{cris},L}$ the field of fractions of $B_{\mathrm{cris},L}$. Let $\mathcal{V}$ denote the $C_{\mathrm{cris},L}$-span of $V_L(D) \subset C_{\mathrm{cris}} \otimes_{K_{0,L}} D$. Then $\mathcal{V}$ is invariant under the action of G_K, and [CF, Lem. 4.6] implies that there exists a unique $K_{0,L}$-subspace $D' \subset D$ such that $C_{\mathrm{cris},L} \otimes_{K_{0,L}} D'$ is equal to $\mathcal{V}$.

Let $s = \dim_{K_{0,L}} D'$ and $d_1, \ldots, d_s \in D'$ a $K_{0,L}$-basis. We also choose $v_1, \ldots, v_s \in V_L(D)$ which span $\mathcal{V}$. Then

$$w := v_1 \wedge \cdots \wedge v_s = b d_1 \wedge \cdots \wedge d_s$$

for some $b \in C_{\mathrm{cris},L}$. As $0 \neq w \in V_L(\wedge^s D')$, $t_{H,L}(D') = t_{N,L}(D')$ by (3.1.3) and $\dim_L V_L(\wedge^s D') = 1$. Moreover, (3.1.3) implies that there is a perfect pairing

$$V_L(\wedge^s D') \otimes_L V_L((\wedge^s D')^*) \to V_L(K_{0,L}) \xrightarrow{\sim} L,$$

so $V_L((\wedge^s D')^*) = b^{-1} d_1 \wedge \cdots \wedge d_s$ and b is a unit in $B_{\mathrm{cris},L}$.
Finally, if $v \in V_L(D)$, write $v = \sum_{i=1}^s b_i v_i$ with $b_i \in C_{\mathrm{cris}}$. Then for $1 \leqslant i \leqslant s$

$$v_1 \wedge \cdots \wedge v_{i-1} \wedge v \wedge v_{i+1} \cdots \wedge v_s = b_i w \in V_L(\wedge^s D').$$

Hence $b_i \in L$, which shows that $\dim_L V_L(D) = s \leqslant \dim_{K_{0,L}} D$. $\square$

(3.2) Write $\mathfrak{S}_L^{\mathrm{ur}} = \mathcal{O}_{\widehat{\mathcal{E}^{\mathrm{ur}}}} \cap W(R)_L \subset W(\operatorname{Fr} R)_L$. We set $P = [\pi_L^{m-1}](u)$. For $\gamma \in \Gamma$, we have $\gamma(P) = [\chi(\gamma)] \circ [\pi_L^{m-1}] = [\chi(\gamma)\pi_L^{m-1}]$, so $\gamma(P)/P$ is a unit in $\mathfrak{S}_L$. In particular $\mathfrak{S}_L^{\mathrm{ur}}[1/P]$ is G_K-stable.
Note that the embedding $\mathfrak{S}_L \hookrightarrow W(R)_L$, extends uniquely to a continuous embedding $\mathcal{O} \hookrightarrow B^+_{\mathrm{cris},L}$, where $B^+_{\mathrm{cris},L} = B^+_{\mathrm{cris}} \otimes_{\mathcal{O}_{L_0}} \mathcal{O}_L$, as usual.

Lemma (3.2.1). *Let $\mathfrak{M}$ be in $\mathrm{Mod}^{\varphi_q}_{/\mathfrak{S}_L}$. The natural map*

$$\text{(3.2.2)}\quad V_{\mathfrak{S}_L}(\mathfrak{M}) := \operatorname{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M}, \mathfrak{S}_L^{\mathrm{ur}}[1/P]) \to \operatorname{Hom}_{\mathcal{O}_{\mathcal{E}},\varphi_q}(\mathcal{O}_{\mathcal{E}} \otimes_{\mathfrak{S}_L} \mathfrak{M}, \widehat{\mathcal{O}}_{\mathcal{E}^{\mathrm{ur}}})$$

is an isomorphism, and both sides are free $\mathcal{O}_L$-modules of rank $\mathrm{rk}_{\mathfrak{S}_L}\mathfrak{M}$. Moreover, if φ_q on $\mathfrak{M}$ induces a map $\varphi_q^*\mathfrak{M} \to \mathfrak{M}$ then the left hand side of (3.2.2) is equal to $\operatorname{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M}, \mathfrak{S}_L^{\mathrm{ur}})$.

Proof. Suppose first that φ_q induces a map $\varphi_q^*\mathfrak{M} \to \mathfrak{M}$. In this case the proof of the lemma is identical to the proof of [Ki, 2.1.4], using [Fo 1, §A.1.2].
Next let $t_L = \log_{\mathcal{G}} u \in \mathcal{O}$ as in the proof of (3.1.3). Then $\varphi_q(\lambda^{-1}t_L) = Q(u)\lambda^{-1}t_L$, and the zeroes of $\lambda^{-1}t_L$ on $D[0,1)$ coincide with those of P. Hence $\lambda^{-1}t_L P^{-1} \in \mathfrak{S}_L[1/p]^\times$, and there exists $w \in \mathfrak{S}_L[1/P]^\times$ such that $\varphi_q(w) = Q(u)w$. Let $\mathfrak{M}(\mathcal{G}) = \mathfrak{S}_L$ equipped with a semi-linear action of φ_q which takes 1 to $Q(u)$. Then multiplication by w^i induces a bijection

$$V_{\mathfrak{S}_L}(\mathfrak{M}) \to V_{\mathfrak{S}_L}(\mathfrak{M} \otimes_{\mathfrak{S}_L} \mathfrak{M}(\mathcal{G})^{\otimes i}).$$

Hence the lemma for general $\mathfrak{M}$ in $\mathrm{Mod}^{\varphi_q}_{/\mathfrak{S}_L}$ follows from the case considered above. $\square$

Proposition (3.2.3). *Let $\mathfrak{M}$ be in $\mathrm{Mod}^{\varphi_q,\Gamma,\mathrm{an}}_{\mathfrak{S}_L}$ and D in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ the weakly admissible module associated to $\mathfrak{M}$ by (2.4.4). Then there is a a canonical G_K-equivariant bijection*

$$\operatorname{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M}, \mathfrak{S}_L^{\mathrm{ur}}[1/P]) \xrightarrow{\sim} \operatorname{Hom}_{\mathrm{Fil},\varphi_q}(D, B_{\mathrm{cris},L})$$

where the right hand side means maps compatible with filtrations in the sense explained in (3.1). In particular, both sides of the above isomorphism have dimension $\dim_{K_{0,L}} D$ over L.

Proof. The argument is similar to [Ki, 2.1.5]. Let $\mathcal{M} = \mathfrak{M} \otimes_{\mathfrak{S}_L} \mathcal{O}$. Note that $P^{-1} \in \lambda t_L^{-1}\mathfrak{S}_L[1/p]^\times \subset B_{\mathrm{cris},L}$. Similarly, $\lambda^{-1} \in P t_L^{-1}\mathfrak{S}_L[1/p]^\times \subset B_{\mathrm{cris},L}$.

Consider the composite
(3.2.4)
$$\mathrm{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M},\mathfrak{S}_L^{\mathrm{ur}}[1/P]) \to \mathrm{Hom}_{\mathcal{O},\varphi_q}(\mathcal{M},B_{\mathrm{cris},L}) \to \mathrm{Hom}_{\mathcal{O},\varphi_q}(\varphi_q^*\mathcal{M},B_{\mathrm{cris},L}).$$

We claim the image of composite map consists of morphisms respecting filtrations. To see this suppose first that φ_q on $\mathcal{M}$ induces a map $\varphi_q^*\mathcal{M} \to \mathcal{M}$. Then by (3.2.1), the left hand side of (3.2.4) is equal to $\mathrm{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M},\mathfrak{S}_L^{\mathrm{ur}})$. A $\mathfrak{S}_L$-linear map $f:\mathfrak{M}\to\mathfrak{S}_L^{\mathrm{ur}}$ induces an $\mathcal{O}$-linear map $f:\mathcal{M}\to B^+_{\mathrm{cris},L}$. If $m\in\varphi_q^*\mathcal{M}$ satisfies $(1\otimes\varphi_q)(m)\in Q^i\mathcal{M}$ for some integer i, then $f\circ(1\otimes\varphi_q)(m)\in Q^iB^+_{\mathrm{cris},L}\subset \mathrm{Fil}^iB_{\mathrm{dR}}$, as $Q(u)\in\mathrm{Fil}^1B_{\mathrm{dR}}$ [Col 1, Lem. 9.3]. This proves the claim when $\varphi_q^*\mathcal{M}$ maps to $\mathcal{M}$.
To prove the claim for general $\mathcal{M}$ we use the notation of the proof of (3.2.1). Let $\mathfrak{M}(i)=\mathfrak{M}\otimes_{\mathfrak{S}}\mathfrak{M}(\mathcal{G})^{\otimes i}$, and $\mathcal{M}(i)=\mathfrak{M}(i)\otimes_{\mathfrak{S}_L}\mathcal{O}$, where i is an integer which is large enough that φ_q induces a map $\varphi_q^*\mathfrak{M}(i)\to\mathfrak{M}(i)$. The underlying $\mathcal{O}$-module of $\mathcal{M}(i)$ may be identified with $\mathcal{M}$, and the induced identification $\varphi_q^*\mathcal{M}=\varphi_q^*\mathcal{M}(i)$ identifies $\mathrm{Fil}^j\varphi_q^*\mathcal{M}$ with $\mathrm{Fil}^{i+j}\varphi_q^*\mathcal{M}$ for all j. As we have a commutative diagram

$$\begin{tikzcd}
\mathrm{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M}(i),\mathfrak{S}_L^{\mathrm{ur}}[1/P]) \arrow[r] \arrow[d, "\sim"', "w^{-i}"] & \mathrm{Hom}_{\mathcal{O},\mathrm{Fil},\varphi_q}(\varphi_q^*\mathcal{M}(i),B_{\mathrm{cris},L}) \arrow[d, "\sim"', "w^{-i}"] \\
\mathrm{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M},\mathfrak{S}_L^{\mathrm{ur}}[1/P]) \arrow[r] & \mathrm{Hom}_{\mathcal{O},\varphi_q}(\varphi_q^*\mathcal{M},B_{\mathrm{cris},L})
\end{tikzcd}$$

the claim follows for general $\mathfrak{M}$.
We now compose (3.2.4) with the map

$$\mathrm{Hom}_{\mathcal{O},\mathrm{Fil},\varphi_q}(\varphi_q^*\mathcal{M},B_{\mathrm{cris},L}) \to \mathrm{Hom}_{K_{0,L},\mathrm{Fil},\varphi_q}(D,B_{\mathrm{cris},L}). \tag{3.2.5}$$

induced by extending a map in the left hand side to $\varphi_q^*\mathcal{M}[1/\lambda]$ and composing with $\varphi_q^*(\xi):D\to\varphi_q^*\mathcal{M}[1/\lambda]$, where ξ is the map of (2.2.2). Note that (3.2.5) respects filtrations as $\varphi_q^*(\xi)$ is compatible with filtrations by (2.2.5), and λ vanishes to order 1 at $u=v_m$. Combining (3.2.4) and (3.2.5) we obtain a canonical G_K-equivariant map

$$\mathrm{Hom}_{\mathfrak{S}_L,\varphi_q}(\mathfrak{M},\mathfrak{S}_L^{\mathrm{ur}}[1/P]) \to \mathrm{Hom}_{K_{0,L},\mathrm{Fil},\varphi_q}(D,B_{\mathrm{cris},L}). \tag{3.2.6}$$

It is easy to see that (3.2.6) is injective. By (3.2.1) the left hand side of (3.2.6) has L-dimension $d=\mathrm{rk}_{\mathfrak{S}_L}\mathfrak{M}$, while by (3.1.4) the right hand side has dimension $\leqslant d$. Hence (3.2.6) is an isomorphism. $\square$

(3.3) We now explain how to pass from φ_q-modules to φ-modules, using an induction procedure which was explained to us by Fontaine (cf. [FY, §7.3]). Suppose that D is a finite dimensional $K_{0,L}$-vector space. We set

$$\tilde{D} := \oplus_{i=0}^{r-1}\varphi^{i*}(D) = \oplus_{\sigma:L_0\hookrightarrow L}\sigma^*(D),$$

so that $\tilde{D}$ is a finite free $K_0 \otimes_{\mathbb{Q}_p} L$-module. Here we have denoted by φ^i the map $K_0 \otimes_{L_0} L \overset{\varphi^i \otimes 1}{\to} K_0 \otimes_{\varphi^i, L_0} L$. We put $\tilde{D}_K = \tilde{D} \otimes_{K_0} K$.

We denote by $\mathrm{Mod}^{F,\varphi}_{/K_0 \otimes_{\mathbb{Q}_p} L}$ the category of finite free $K \otimes_{\mathbb{Q}_p} L$-modules $\tilde{D}$ equipped with an isomorphism $\varphi^*(\tilde{D}) \xrightarrow{\sim} \tilde{D}$ and a decreasing separated filtration on $\tilde{D}_K = \tilde{D} \otimes_{K_0} K$, indexed by $\mathbb{Z}$, by $K \otimes_{\mathbb{Q}_p} L$-submodules. For any $\tilde{D}$ in $\mathrm{Mod}^{F,\varphi}_{K_O \otimes_{\mathbb{Q}_p} L}$ we denote by $t_N(\tilde{D})$ and $t_H(\tilde{D})$ the usual invariants when $\tilde{D}$ is considered as a filtered φ-module over K_0.

For any $\tilde{D}$ in $\mathrm{Mod}^{F,\varphi}_{/K_0 \otimes_{\mathbb{Q}_p} L}$ we may write $\tilde{D}_K = \oplus_{\mathfrak{m}} (\tilde{D}_K)_{\mathfrak{m}}$ where $\mathfrak{m}$ runs over the maximal ideals of $K \otimes_{\mathbb{Q}_p} L$. Denote by $\mathfrak{m}_0$ the kernel of the natural map $K \otimes_{\mathbb{Q}_p} L \to K$. Given D in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$ we give $\tilde{D}$ the structure of an object in $\mathrm{Mod}^{F,\varphi}_{/K_0 \otimes_{\mathbb{Q}_p} L}$ by noting that $(\tilde{D}_K)_{\mathfrak{m}_0}$ may be identified with D_K, and giving $\tilde{D}_K$ the direct sum of the filtration on D_K and the trivial filtration on $\oplus_{\mathfrak{m} \neq \mathfrak{m}_0} (\tilde{D}_K)_{\mathfrak{m}}$.

Lemma (3.3.1). *The functor $D \mapsto \tilde{D}$ induces an fully faithful $\otimes$-functor*

$$\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}} \xrightarrow{\sim} \mathrm{Mod}^{F,\varphi}_{K_0 \otimes_{\mathbb{Q}_p} L}$$

The essential image of this functor consists of those objects such that the filtration on $\oplus_{\mathfrak{m} \neq \mathfrak{m}_0} \tilde{D}_K$ is trivial.

Proof. Given D in $\mathrm{Mod}^{F,\varphi_q}_{K_{0,L}}$ there is a natural isomorphism

$$\varphi^*(\tilde{D}) = \oplus_{i=1}^{r} \varphi^{i*}(D) \xrightarrow{\sim} \oplus_{i=0}^{r-1} \varphi^{i*}(D) = \tilde{D},$$

which sends $\varphi^{i*}(D)$ identically to $\varphi^{i*}(D)$ for $i \neq r$ and maps $\varphi^{r*}(D) = \varphi_q^*(D)$ to D using the map φ_q on D. This defines the functor of the lemma.

To define a quasi-inverse on the essential image of the functor, let $\tilde{D}$ be in $\mathrm{Mod}^{F,\varphi}_{/K_0 \otimes_{\mathbb{Q}_p} L}$ be such that the filtration on $\oplus_{\mathfrak{m} \neq \mathfrak{m}_0} \tilde{D}_K$ is trivial, and set $D' = \tilde{D} \otimes_{K_0 \otimes_{\mathbb{Q}_p} L} (K_0 \otimes_{L_0} L)$. There is an isomorphism $\varphi^{*r}(D') \xrightarrow{\sim} D'$ induced by $\varphi^{*r}(\tilde{D}) \xrightarrow{\sim} \tilde{D}$. Using the decomposition

$$K_0 \otimes_{\mathbb{Q}_p} L \xrightarrow{\sim} \oplus_{i=0}^{r-1} K_0 \otimes_{\varphi^i, L_0} L$$

one sees that there is a canonical isomorphism $\tilde{D}' \xrightarrow{\sim} \tilde{D}$. In particular this makes D' into an object of $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$.

One checks immediately that these two functors are quasi-inverse. $\square$

Lemma (3.3.2). *Let D be in $\mathrm{Mod}^{F,\varphi_q}_{/K_{0,L}}$. Then*

$$t_{N,L}(D) = t_N(\tilde{D}) \text{ and } t_{H,L}(D) = t_H(\tilde{D}) \tag{3.3.3}$$

and D is weakly admissible if and only if $\tilde{D}$ is weakly admissible.

Proof. Since the functor of (3.3.1) respects $\otimes$-products, it suffices to prove (3.3.3) when D is 1-dimensional over $K_{0,L}$. Moreover, as the essential image of the functor in (3.3.1) is stable under subobjects, (3.3.3) implies the claim regarding weak admissibility.
Let D be 1-dimensional over $K_{0,L}$ with basis vector v, and $\varphi_q(v) = \alpha v$ for some $\alpha \in K_{0,L}$. Then for $i = 0, \dots, r-1$ the K_0-vector space $\wedge_{K_0}^{[L:L_0]} \varphi^{i*}(D)$ has a basis vector e_i such that $\varphi(e_i) = e_{i+1}$ if $i < r-1$, and $\varphi(e_{r-1}) = N_{K_{0,L}/K_0}(\alpha)e_0$. Hence φ takes

$$e_0 \wedge \cdots \wedge e_{r-1} \in \wedge_{K_0}^{[L:\mathbb{Q}_p]} \tilde{D} \xrightarrow{\sim} \otimes_{i=0}^{r-1} \wedge_{K_0}^{[L:L_0]} \varphi^{i*}(D)$$

to $(-1)^r N_{K_{0,L}/K_0}(\alpha) e_0 \wedge \cdots \wedge e_{r-1}$, and

$$t_N(\tilde{D}) = v_p(N_{K_{0,L}/K_0}(\alpha)) = [L : L_0] v_p(\alpha) = v_{\pi_L}(\alpha) = t_{N,L}(D).$$

On the other hand, $t_{H,L}(D) = t_H(\tilde{D})$, from the definition of the filtration on $\tilde{D}$. $\square$

Proposition (3.3.4). *Let D be a weakly admissible object in $\mathrm{Mod}_{/K_{0,L}}^{F,\varphi_q}$. Then there is a canonical G_K-equivariant isomorphism*

$$V_L^*(D) \xrightarrow{\sim} V^*(\tilde{D}) := \mathrm{Hom}_{\mathrm{Fil},\varphi}(\tilde{D}, B_{\mathrm{cris}}).$$

Proof. Suppose $\tilde{f} : \tilde{D} \to B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} L$ is a $K_0 \otimes_{\mathbb{Q}_p} L$-linear, φ-compatible map, such that

$$f_K : \tilde{D}_K \to K \otimes_{K_0} B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} L \hookrightarrow B_{\mathrm{dR}} \otimes_{\mathbb{Q}_p} L$$

is compatible with filtrations. Consider the composite

$$\theta(f) : D \hookrightarrow \tilde{D} \to B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} L \to B_{\mathrm{cris},L}.$$

This is a φ_q-compatible map, such that the composite

$$\theta(f)_K : D_K \to K \otimes_{K_{0,L}} B_{\mathrm{cris},L} \hookrightarrow B_{\mathrm{dR}}.$$

is obtained from f_K by localizing at $\mathfrak{m}_0$. In particular, $\theta(f)_K$ is compatible with filtrations. Note also that f can be recovered from $\theta(f)$: The decomposition

$$K_0 \otimes_{\mathbb{Q}_p} L \xrightarrow{\sim} \oplus_{i=0}^{r-1} K_0 \otimes_{\varphi^i, L_0} L$$

allows us to view $B_{\mathrm{cris},L}$ as a direct summand in $B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} L$. Then f is the unique φ-linear extension of the φ_q-linear map

$$D \overset{\theta(f)}{\to} B_{\mathrm{cris},L} \hookrightarrow B_{\mathrm{cris}} \otimes_{\mathbb{Q}_p} L.$$

Hence we have an injective map

$$\mathrm{Hom}_{\mathrm{Fil},\varphi,K_0\otimes_{\mathbb{Q}_p}L}(\tilde{D}, B_{\mathrm{cris}}\otimes_{\mathbb{Q}_p}L) \hookrightarrow \mathrm{Hom}_{\mathrm{Fil},\varphi_q,K_{0,L}}(D, B_{\mathrm{cris},L}) \tag{3.3.5}$$

On the other hand, the trace map $L \to \mathbb{Q}_p$ induces an isomorphism

$$\mathrm{Hom}_{\mathrm{Fil},\varphi,K_0\otimes_{\mathbb{Q}_p}L}(\tilde{D}, B_{\mathrm{cris}}\otimes_{\mathbb{Q}_p}L) \xrightarrow{\sim} \mathrm{Hom}_{\mathrm{Fil},\varphi,K_0}(\tilde{D}, B_{\mathrm{cris}}). \tag{3.3.6}$$

Composing the inverse of (3.3.6) with (3.3.5) gives an injective map $V^*(\tilde{D}) \to V_L^*(D)$. As $\tilde{D}$ is admissible by (3.3.2), $\dim_L V^*(\tilde{D}) = \dim_{K_{0,L}} D$, and this is equal to $\dim_L V_L^*(D)$ by (3.3.2). Hence we have $V^*(\tilde{D}) \xrightarrow{\sim} V_L^*(D)$. $\square$

(3.3.7) Denote by $\mathrm{Rep}_{G_K}^{L\text{-cris}}$ the full subcategory of Rep_{G_K} consisting of those objects V such that $V\otimes_{\mathcal{O}_L} L$ is a crystalline representation and, if $D_{\mathrm{dR}}(V) = (V\otimes_{\mathbb{Q}_p} B_{\mathrm{dR}})^{G_K}$, then the filtration on $D_{\mathrm{dR}}(V)_{\mathfrak{m}}$ is trivial for $\mathfrak{m} \neq \mathfrak{m}_0$ a maximal ideal of $K\otimes_{\mathbb{Q}_p} L$.

COROLLARY (3.3.8). *There is an exact equivalence of $\otimes$-categories*

$$\mathrm{Mod}_{/\mathfrak{S}_L}^{\varphi_q,\Gamma,\mathrm{an}} \xrightarrow{\sim} \mathrm{Rep}_{G_K}^{L\text{-cris}}; \quad \mathfrak{M} \mapsto V(\mathcal{O}_{\mathcal{E}}\otimes_{\mathfrak{S}_L}\mathfrak{M})$$

where V is the functor introduced in (1.5).

Proof. Using (3.2.3) and (3.3.4) one sees that the functor of (1.6) induces a fully faithful, exact $\otimes$-functor as in the corollary. To show that this functor is essentially surjective let V be in $\mathrm{Rep}_{G_K}^{L\text{-cris}}$, and let $M = M_{\mathcal{O}_{\mathcal{E}}}(V)$. By (2.4.4) and (3.3.1), there exists an $\mathfrak{M}'$ in $\mathrm{Mod}_{/\mathfrak{S}_L}^{\varphi_q,\Gamma,\mathrm{an}}$ such that $V(\mathcal{O}_{\mathcal{E}}\otimes_{\mathfrak{S}_L}\mathfrak{M}')$ is isomorphic to a G_K-stable $\mathcal{O}_L$-lattice $V' \subset V\otimes_{\mathcal{O}_L} L$. Thus, by the equivalence of (1.6) there is an isomorphism $\mathcal{E}\otimes_{\mathcal{O}_{\mathcal{E}}} M \xrightarrow{\sim} \mathcal{E}\otimes_{\mathfrak{S}_L}\mathfrak{M}'$. Then $\mathfrak{M} = M\cap\mathfrak{M}'[1/p] \subset M[1/p]$ is in $\mathrm{Mod}_{/\mathfrak{S}_L}^{\varphi_q,\Gamma,\mathrm{an}}$ and satisfies $\mathcal{O}_{\mathcal{E}}\otimes_{\mathfrak{S}_L}\mathfrak{M} \xrightarrow{\sim} M$. Hence $V(\mathcal{O}_{\mathcal{E}}\otimes_{\mathfrak{S}_L}\mathfrak{M}) \xrightarrow{\sim} V(M) \xrightarrow{\sim} V$. $\square$

REFERENCES

[Be] L. Berger, *Représentations p-adiques et équations différentielles*, Invent. Math. **148(2)** (2002), 219-284.

[Be 2] L. Berger, *Équations différentielles p-adiques et (φ, N)-modules filtrés*, Astérisque **319** (2008), 13-38.

[Be 3] L. Berger, *Limites de représentations cristallines*, Compos. Math **140** (2004), 1473-1498.

[CF] P. Colmez, J-M. Fontaine, *Construction des représentations p-adiques semi-stables*, Invent. Math. **140** (2000), 1-43.

[Co] R. Coleman, *Division values in local fields*, Invent. Math **53** (1979), 91-116.

[Col 1] P. Colmez, *Espaces de Banach de dimension finie*, J. Inst. Math. Jussieu **1** (2002), 331-439.

[Col 2] P. Colmez, *Représentations cristallines et représentations de hauteur finie*, J. reine angew. Math **514** (1999), 119-143.

[Fo] J-M. Fontaine, *Représentations p-adiques des corps locaux*, Grothendieck Festschrift II, Prog. Math. 87, Birkhauser, pp. 249-309, 1991.

[Fo 2] J. M. Fontaine, *Le corps de périodes p-adiques*, Périodes p-adiques (Bures-sur-Yvette 1988), Astérisque 233, SMF, 1994, pp. 59-111.

[Fou] L. Fourquaux, *Logarithme de Perrin-Riou pour des extensions associées à un groupe de Lubin-Tate*, thesis (2005).

[FY] J. M. Fontaine, Y. Ouyang, *Theory of p-adic Galois representations*, preprint.

[Ke 1] K. Kedlaya, *A p-adic local monodromy theorem*, Ann. of Math. (2004), 93-184.

[Ke 2] K. Kedlaya, *Slope filtrations revisited*, Documenta Math. **10** (2005), 447-525.

[Ki] M. Kisin, *Crystalline representations and F-crystals*, Algebraic geometry and number theory. In honour of Vladimir Drinfeld's 50^{th} birthday, Prog. Math. 253, Birkhäuser, pp. 459-496, 2006.

[Wa] N. Wach, *Représentations p-adiques potentiellement cristallines*, Bull. Soc. Math. France **124** (1996), 375-400.

[Wi] J-P. Wintenberger, *Le corps des normes de certaines extensions infinies de corps locaux; applications*, Annales. Sci. ENS **16** (1983), 59-89.

Mark Kisin
Department of Mathematics
University of Chicago
USA
kisin@ math.uchicago.edu
Current address:
Department of Mathematics
Harvard University
USA
kisin@math.harvard.edu

Wei Ren
Department of Mathematics
University of Chicago
USA

The Energy of Heavy Atoms According to Brown and Ravenhall: The Scott Correction

Rupert L. Frank, Heinz Siedentop, and Simone Warzel

Received: March 9, 2009

Communicated by Alfred K. Louis

Abstract. We consider relativistic many-particle operators which – according to Brown and Ravenhall – describe the electronic states of heavy atoms. Their ground state energy is investigated in the limit of large nuclear charge and velocity of light. We show that the leading quasi-classical behavior given by the Thomas-Fermi theory is raised by a subleading correction, the Scott correction. Our result is valid for the maximal range of coupling constants, including the critical one. As a technical tool, a Sobolev-Gagliardo-Nirenberg-type inequality is established for the critical atomic Brown-Ravenhall operator. Moreover, we prove sharp upper and lower bounds on the eigenvalues of the hydrogenic Brown-Ravenhall operator up to and including the critical coupling constant.

2000 Mathematics Subject Classification: Primary 81V45; Secondary 81V55, 35Q40, 46N50, 47N50.
Keywords and Phrases: Heavy atoms, ground state energy, relativistic Coulomb system, Scott correction, Brown-Ravenhall operator.

Contents

1 Introduction and main result

The description of atoms and molecules, in particular of their energies, has been a primer for the development of quantum mechanics. However, it became soon clear that atoms with more than one electron are not accessible to explicit solutions. This motivated the development of approximate models for large Coulomb systems. One of the most simple and – simultaneously – the most fundamental models was introduced by Thomas [73], Fermi [27, 28], and Lenz [41] who proposed the energy functional which we will also use here. It predicts that the ground state energy of atoms would decrease with the atomic number Z to leading order as $Z^{7/3}$. In order to get a refined description, Scott [54] conjectured that the electrons close to the nucleus should raise the energy by $Z^2/2$. Considerably later Schwinger [52] argued also for Scott's prediction; Schwinger [53] and Englert and Schwinger [13, 14, 15] even refined these considerations by adding more lower order terms [53] (see also Englert [12]). In fact, a contribution to the $Z^{5/3}$-term can be traced back to Dirac [10]. — The challenge to address the question whether the predicted formulae would yield asymptotically correct results when compared with the N-particle Schrödinger theory was for a long time unsuccessful. It was Lieb and Simon who proved in their seminal paper [44] that the prediction of Thomas, Fermi, and Lenz is indeed asymptotically correct. However, establishing the Scott correction resisted the mathematical efforts and became Problem 10B of Simon's 15 Problems in Mathematical Physics [62]. Eventually, the Scott correction was established mathematically by Hughes [37, 38] (lower bound), and Siedentop and Weikard [55, 56, 57, 58, 59] (lower and upper bound). In fact, even the existence of the $Z^{5/3}$-correction conjectured by Schwinger was proved by Fefferman and Seco [23, 24, 25, 18, 26, 21, 19, 20, 22]. Later these results were extended in various ways, e.g., to ions and molecules [1, 39, 66, 4].

Despite of the mathematical success in establishing the large Z asymptotics of the Schrödinger theory, these considerations remain questionable from a physical point of view, since large atoms force electrons into orbits that are close to the nucleus where the electrons move with high speed which should require a relativistic treatment. The atom is shrinking with increasing Z: already in non-relativistic quantum mechanics the bulk of the electrons has a distance $Z^{-1/3}$ from the nucleus; the electrons contributing to the Scott correction even live on the scale Z^{-1}. Schwinger [53] has estimated these effects concluding that a correction to the Scott correction occurs whereas the leading term should be unaffected by the change of model. Sørensen [51] was the first who proved that the latter is indeed the case for a simplified ad hoc naive relativistic model, the Chandrasekhar multi-particle operator, in the limit of large Z and large velocity of light c. The value of the Scott correction is again of order Z^2, a result which was announced [64] and proven [65] by Solovej, Sørensen, and Spitzer (see also Sørensen [50] for the non-interacting case). In a previous paper [31] we gave a short alternative proof, roling the problem back to the non-relativistic Scott correction. Nevertheless, a question from the physical point of view remains: Although the Chandrasekhar model is believed to represent some qualitative features of relativistic systems, there is no reason to assume that it should give quantitatively correct results. Therefore, to obtain not only qualitatively correct results it is interesting, in fact mandatory, to consider a Hamiltonian which – as the one by Brown and Ravenhall [6] – is derived from QED such that it yields the leading relativistic effects in a quantitative correct manner. (See also Sucher [69, 70, 71].) The first step in this direction was taken by Cassanas and Siedentop [7] who showed that, similarly to the Chandrasekhar case, the leading energy is not affected. To show in which way the Scott correction is changed for this model is our concern in this paper.

1.1 Relativistic energy form

According to Brown and Ravenhall [6] the energy of an atom with N electrons in a state $\psi \in \mathfrak{Q}_N^B$ is given by

$$\mathcal{E}_N^B(\psi) := \left\langle \psi, \left[\sum_{\nu=1}^N \left(c\,\boldsymbol{\alpha}_\nu \cdot \mathbf{p}_\nu + c^2\beta_\nu - c^2 - \frac{Z}{|\mathbf{x}_\nu|}\right) + \sum_{1\leq\mu<\nu\leq N} \frac{1}{|\mathbf{x}_\mu - \mathbf{x}_\nu|}\right] \psi \right\rangle. \tag{1.1}$$

This involves the free Dirac operator reduced by the rest mass, acting in $L^2(\mathbb{R}^3, \mathbb{C}^4)$, with the four Dirac matrices in standard representation,

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where $\boldsymbol{\sigma}$ are the three Pauli matrices in standard representation, i.e.,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We use atomic units in which $m = e^2 = \hbar = 1$. The parameter Z is the atomic number and c the velocity of light.
The Hilbert space of an electron is chosen as the positive spectral subspace of the Dirac operator,

$$\mathfrak{H}^B := \chi_{[c^2,\infty)}(c\,\boldsymbol{\alpha}\cdot\mathbf{p} + c^2\beta)\left(L^2(\mathbb{R}^3, \mathbb{C}^4)\right),$$

and, correspondingly, the Hilbert space of N electrons $\mathfrak{H}_N^B$ is the antisymmetric tensor product of the one-particle space, i.e., $\mathfrak{H}_N^B := \bigwedge_{\nu=1}^N \mathfrak{H}^B$. Finally, the form domain of (1.1) is $\mathfrak{Q}_N^B := \mathfrak{H}_N^B \cap \mathfrak{S}(\mathbb{R}^{3N}, \mathbb{C}^{4^N})$ with $\mathfrak{S}$ the Schwartz space of rapidly decreasing functions. As is shown in [17], the Brown-Ravenhall form $\mathcal{E}_N^B$ is closable and bounded from below if and only if

$$\kappa := \frac{Z}{c} \leq \kappa^B := \frac{2}{2/\pi + \pi/2}. \tag{1.2}$$

(See also Tix [74, 76], who improved the bound given in [17] to an explicit positive bound.) For the physical value, about 1/137, of the Sommerfeld fine structure, which equals $1/c$ in atomic units used here, the critical atomic number Z exceeds 124 slightly. This includes all known elements.

In the following we will assume that the atom described by (1.1) is neutral, i.e., $Z = N$, an assumption that we make mainly for the sake of brevity and clarity of presentation, since the Scott correction is independent of the ionization degree $N/Z \geq \text{const} > 0$. Similarly, it might seem that our treatment is restricted to spherically symmetric systems (atoms). However, on the energy scale considered here, molecular Hamiltonians essentially separate – in nature the distances between nuclei with charges $ZZ_1, ..., ZZ_K$ remain on a scale much larger than $Z^{-1/3}$ – into spherically symmetric one-center problems (atoms). Therefore, the molecular case follows from the atomic case by additional localization. However, for the sake of brevity and clarity, we will spare the reader the corresponding tedious technicalities, restrict to the atomic case, and freely use the resulting symmetry.
Thus, according to Friedrichs, the one-particle form $\mathcal{E}_1^B$ defines for $\kappa \leq \kappa^B$ a distinguished self-adjoint operator in $\mathfrak{H}^B$. Through a unitary transformation it may be represented as a self-adjoint operator in the Hilbert space $\mathfrak{H} := L^2(\mathbb{R}^3, \mathbb{C}^2)$ of two-spinors. More precisely, using the notation $p := |\mathbf{p}|$, $\boldsymbol{\omega}_{\mathbf{p}} := \mathbf{p}/p$ we set

$$E(p) := \sqrt{\mathbf{p}^2 + 1}, \qquad \phi_\nu(p) := \sqrt{\frac{E(p) + (-1)^\nu}{2E(p)}}, \ \nu = 0, 1, \tag{1.3}$$

and introduce the following bounded operators on $\mathfrak{H}$,

$$\Phi_0(\mathbf{p}) := \phi_0(p), \quad \Phi_1(\mathbf{p}) := \phi_1(p) \ \boldsymbol{\sigma} \cdot \boldsymbol{\omega}_{\mathbf{p}}. \tag{1.4}$$

The operator $\boldsymbol{\Phi}_c : \mathfrak{H} \to \mathfrak{H}^B$, $\psi \mapsto (\Phi_0(\mathbf{p}/c)\,\psi, \Phi_1(\mathbf{p}/c)\,\psi)$, maps $\mathfrak{H}$ unitarily onto $\mathfrak{H}^B$ [7]. Therefore, the form $\mathcal{E}_1^B$ defines the (two-spinor) Brown-Ravenhall operator in $\mathfrak{H}$,

$$\begin{aligned} B_c[Z/|\mathbf{x}|] :=& \boldsymbol{\Phi}_c^{-1} \left(c\boldsymbol{\alpha} \cdot \mathbf{p} + c^2\beta - c^2 - Z/|\mathbf{x}| \right) \boldsymbol{\Phi}_c \\ =& c^2 E(p/c) - c^2 - \mathcal{U}_c(Z/|\mathbf{x}|), \end{aligned} \tag{1.5}$$

where $\mathcal{U}_c(A) := \Phi_0(\mathbf{p}/c)\, A\, \Phi_0(\mathbf{p}/c) + \Phi_1(\mathbf{p}/c)\, A\, \Phi_1(\mathbf{p}/c)$. In the case $c = 1$ we denote this operator by B_Z. Further properties of B_Z and its relation to the corresponding Chandrasekhar operator and Schrödinger operator

$$C_Z := (\mathbf{p}^2 + 1)^{1/2} - 1 - Z/|\mathbf{x}|, \qquad S_Z := \tfrac{1}{2}\mathbf{p}^2 - Z/|\mathbf{x}| \tag{1.6}$$

all realized in $\mathfrak{H}$, can be found in Sections 2 and 3 below and in Appendix C.

1.2 Main result

We are interested in the ground state energy

$$E_c^B(Z) := \inf\{\mathcal{E}_Z^B(\psi) \mid \psi \in \mathfrak{Q}_Z^B, \ \|\psi\| = 1\}$$

of the energy form (1.1) for large atomic number Z and large velocity of light c satisfying (1.2). Note that we picked $N = Z$. It was shown in [7], that similarly to the Chandrasekhar case [51], the leading behavior of $E_c^B(Z)$ is not affected by relativistic effects and is, as in the Schrödinger case [44], given by the minimal Thomas-Fermi energy

$$E_{\rm TF}(Z) := \inf\{\mathcal{E}_{\rm TF}(\rho) \,|\, \rho \in L^{5/3}(\mathbb{R}^3),\ \rho \geq 0,\ D(\rho,\rho) < \infty\}. \tag{1.7}$$

The latter is defined in terms of the Thomas-Fermi energy functional

$$\mathcal{E}_{\rm TF}(\rho) := \int_{\mathbb{R}^3} \left[\frac{3}{5}\gamma_{\rm TF}\,\rho(\mathbf{x})^{5/3} - \frac{Z}{|\mathbf{x}|}\rho(\mathbf{x})\right] \mathrm{d}\mathbf{x} + D(\rho,\rho)$$

where, in our units, $\gamma_{\rm TF} = (3\pi^2)^{2/3}/2$ and

$$D(\rho,\sigma) := \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\overline{\rho(\mathbf{x})}\sigma(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\,\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

is the Coulomb scalar product. By scaling, one finds $E_{\rm TF}(Z) = E_{\rm TF}(1)\,Z^{7/3}$. This paper concerns the correction to the leading behavior. For the formulation of the main result, we abbreviate the negative part of an operator by $A_- := -A\chi_{(-\infty,0)}(A)$ and introduce for $0 < \kappa \leq \kappa^B$ the spectral shift

$$s(\kappa) := \kappa^{-2}\,\mathrm{tr}_{\mathfrak{H}}\left[(B_\kappa)_- - (S_\kappa)_-\right]. \tag{1.8}$$

(We use the term "spectral shift" for s for convenience although it is used in slightly different meaning otherwise.) It describes the shift of the Brown-Ravenhall bound state energies compared to those of the Schrödinger operator. In Section 3 we show that s is well-defined and discuss some of its properties. In particular, we prove that the function s is non-negative on the interval $(0,\kappa^B]$ and can be continuously extended to zero where it satisfies

$$s(\kappa) = \mathcal{O}(\kappa^2) \qquad \text{as } \kappa \to 0. \tag{1.9}$$

We are now ready to state our main result.

Theorem 1.1 (Scott correction). *There exists a constant $C > 0$ such that for all $Z \geq 1$ and all $c \geq Z/\kappa^B$ one has*

$$\left|E_c^B(Z) - E_{\rm TF}(Z) - \left(\tfrac{1}{2} - s(Z/c)\right)Z^2\right| \leq C Z^{47/24}. \tag{1.10}$$

Put differently, Theorem 1.1 asserts that in the limit $Z \to \infty$ we have uniformly in the quotient $\kappa = Z/c \in (0,\kappa^B]$

$$E_c^B(Z) = E_{\rm TF}(Z) + \left(\tfrac{1}{2} - s(\kappa)\right)Z^2 + o(Z^2). \tag{1.11}$$

(We do not claim that the error $Z^{47/24}$ in (1.10) is sharp, so we only write $o(Z^2)$ here.) The second term $\left(\frac{1}{2} - s(\kappa)\right)Z^2$ in (1.11) is the so-called Scott correction in the Brown-Ravenhall model. It does not exceed the Scott correction $Z^2/2$ in the non-relativistic model [55]. Indeed, if $\kappa = Z/c$ stays away from zero then there is a relativistic lowering of the ground state energy at order Z^2. On the other hand, in the non-relativistic limit $c \to \infty$ with $\kappa = Z/c \to 0$, one recovers – non-surprisingly – the value of the Schrödinger case. In this case (1.9) implies

$$E_c^B(Z) = E_{\mathrm{TF}}(Z) + \tfrac{1}{2}Z^2 + \mathcal{O}(c^{-2}Z^4 + Z^{47/24}). \tag{1.12}$$

The Scott correction in the Brown-Ravenhall model, however, exceeds the Scott correction predicted by the naive Chandrasekhar model treated in [65] and [31]. This follows from the fact that sums of bound state energies of the atomic Chandrasekhar operator are dominated by those of the Brown-Ravenhall operator, cf. the proof of Theorem 3.1 below.

1.3 Outline of the paper

The central strategy of our paper is to compare the ground state energy of the Brown-Ravenhall operator with that of the Schrödinger operator. The latter is known up to the required accuracy $o(Z^2)$ and the leading contribution agrees with the Brown-Ravenhall energy. The subtraction of the corresponding ground state energies results in a renormalized effective model which accurately describes the energy differences and is amenable to analysis. The germ of this idea has been presented in the simpler context of the Chandrasekhar model [31]. The full blown renormalization required is developed in this paper. A virtue of our approach is that it leads to an explicit formula for the spectral shift which can be evaluated numerically. We believe it would be interesting to compare this formula with experimental data.

We show that the difference between the Brown-Ravenhall and Schrödinger ground state energies on the multi-particle level coincides, up to the required accuracy, with a spectral shift on the one-particle level. A crucial step in our analysis is therefore a bound on the corresponding spectral shift for rather general spherically symmetric potentials. This is presented in Section 3, where we show that sums of differences of Brown-Ravenhall and Schrödinger eigenvalues decay rather rapidly as the angular momentum increases.

In Section 2 we address various aspects of hydrogenic Brown-Ravenhall operators. An essential feature and source of difficulties, which does not occur in the naive Chandrasekhar model, is the non-locality of the potential energy. In particular, instead of the usual Coulomb potential $|\mathbf{x}|^{-1}$ we face the 'twisted' non-local operator $\mathcal{U}_c(|\mathbf{x}|^{-1})$. Estimating the difference between the corresponding potential energies is the topic in Subsection 2.3. Since, in contrast to the Schrödinger case, the eigenvalues of the hydrogenic Brown-Ravenhall operator are not known explicitly, we prove upper and lower bounds in Subsection 2.1. Our bounds are sharp with respect to their dependence on the quantum numbers n and l. An upper bound is given by the Dirac eigenvalues, a consequence of the mini-max principle for eigenvalues in the gap. For the lower bound we overcome the non-locality of the potential by a non-trivial comparison argument with a super-critical Chandrasekhar operator. In Subsection 2.2 we prove a new Sobolev-type inequality, from which we derive estimates on the eigenfunctions of the hydrogenic Brown-Ravenhall operator. The technical challenge here is to prove such a result up to and including the critical coupling constant.

Finally, we present the proof of our main result, Theorem 1.1, in Section 4.

For the readers' convenience we collect various facts in the appendices. Appendix A recalls the partial wave decomposition of the Hilbert space of two-spinors, Appendix B establishes some useful properties of the twisting operators, and Appendix C collects basic facts on hydrogenic Brown-Ravenhall and Chandrasekhar operators. Appendix D fills in some details in the proof of Theorem 2.2 and, eventually, Appendix E defines the one-particle density matrix giving the main contribution of the energy.

2 The hydrogenic Brown-Ravenhall operator

In this section we set $c = 1$ and investigate the Brown-Ravenhall operator with Coulomb potential

$$B_\kappa = \sqrt{\mathbf{p}^2 + 1} - 1 - \kappa \mathcal{U}(|\mathbf{x}|^{-1}) \tag{2.1}$$

in the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C}^2)$ of two-spinors, where we recall that

$$\mathcal{U}(|\mathbf{x}|^{-1}) = \Phi_0(\mathbf{p})|\mathbf{x}|^{-1}\Phi_0(\mathbf{p}) + \Phi_1(\mathbf{p})|\mathbf{x}|^{-1}\Phi_1(\mathbf{p}) \tag{2.2}$$

with Φ_ν defined in (1.4). In Subsection 2.1 we prove sharp upper and lower bounds on the eigenvalues of B_κ. In Subsection 2.2 we prove L^p estimates on the eigenfunctions of this operator. Technically, this is expressed as a Sobolev-type inequality for the massless version of B_κ, which is a non-negative operator. Finally, in Subsection 2.3 we compare the potential energy of the operator B_κ, namely $\langle \psi, \mathcal{U}(|\mathbf{x}|^{-1})\psi \rangle$, with the corresponding local potential energy $\langle \psi, |\mathbf{x}|^{-1}\psi \rangle$. For comparison purpose also the corresponding Chandrasekhar and Schrödinger operator C_κ and S_κ occur (see (1.6)).

According to [17] and [40] the operators B_κ and C_κ are well-defined for all $\kappa \le \kappa^\#$ with $\# = B, C$ and

$$\kappa^B = \frac{2}{2/\pi + \pi/2}, \qquad \kappa^C := 2/\pi; \tag{2.3}$$

see also Appendix C. Of course, for the Schrödinger operator no upper bound on κ is needed.

2.1 Estimates on eigenvalues of the hydrogen atom

In contrast to the Schrödinger or Dirac models, the eigenvalues of B_κ and C_κ are not known explicitly. In order to obtain upper and lower bounds on these eigenvalues, we use that the spectra of B_κ, C_κ and S_κ may be classified in terms of angular momenta.

As usual write $\mathbf{L} := \mathbf{x} \times \mathbf{p}$ for the operators of orbital angular momentum and $\mathbf{J} := \mathbf{L} + \frac{1}{2}\boldsymbol{\sigma}$ for the operators of total angular momentum. The four operators

B_κ, $\mathbf{J}^2$, J_3, $\mathbf{L}^2$ commute pairwise, and this also holds, if C_κ or S_κ replace B_κ. This allows us to decompose the Hilbert space $\mathfrak{H}$ into orthogonal subspaces which reduce such a quadruple of operators, i.e.,

$$\mathfrak{H} = \bigoplus_{j\in\mathbb{N}_0+\frac{1}{2}} \bigoplus_{l=j\pm 1/2} \mathfrak{H}_{j,l}, \qquad \mathfrak{H}_{j,l} := \bigoplus_{m=-j}^{j} \mathfrak{H}_{j,l,m}. \tag{2.4}$$

Here $\mathfrak{H}_{j,l,m}$ is the maximal joint eigenspace of $\mathbf{J}^2$ with eigenvalues $j(j+1)$, of $\mathbf{L}^2$ with eigenvalue $l(l+1)$, and J_3 with eigenvalue m. More details concerning the partial wave decomposition (2.4) can be found in Appendix A.
We denote by $b_{j,l}(\kappa)$, $c_l(\kappa)$, and $s_l(\kappa)$ the reduced operators corresponding to fixed angular momenta j and l, where, strictly speaking, we consider $b_{j,l}(\kappa)$ and $c_l(\kappa)$ in momentum space whereas $s_l(\kappa)$ in position space. We refer to Appendix C for precise definitions of $b_{j,l}(\kappa)$ and $c_l(\kappa)$ and for further discussion. Of course, $s_l(\kappa) = -\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{l(l+1)}{2r^2} - \frac{\kappa}{r}$.
The main result of this subsection is that for large quantum numbers n, j, and l, the eigenvalues of $b_{j,l}(\kappa)$ and $c_l(\kappa)$ behave similarly to the explicitly known ones of the Schrödinger operator $s_l(\kappa)$.

Theorem 2.1 (Energies of Brown-Ravenhall hydrogen). *There is a constant $C < \infty$ such that for all $j \in \mathbb{N}_0 + \frac{1}{2}$, and $l = j \pm \frac{1}{2}$, $n \in \mathbb{N}$ and $\kappa \in (0, \kappa^B]$ one has*

$$-C\frac{\kappa^2}{(n+l)^2} \leq \lambda_n(b_{j,l}(\kappa)) \leq -\frac{\kappa^2}{2(n+l)^2}. \tag{2.5}$$

Here and below, we denote by $\lambda_1(A) \leq \lambda_2(A) \leq \dots$ the eigenvalues, repeated according to multiplicities, below the bottom of the essential spectrum of the self-adjoint, lower semi-bounded operator A. Note that $-\kappa^2(2(n+l)^2)^{-1} = \lambda_n(s_l(\kappa))$ on the right hand side of (2.5) is the n-th eigenvalue of the Schrödinger operator corresponding to angular momentum l. In particular, we conclude from (2.5) that the partial traces in $\mathfrak{H}_{j,l}$ (cf. (3.1)) satisfy

$$0 \leq \operatorname{tr}_{j,l}\left([B_\kappa + \mu]_- - [S_\kappa + \mu]_-\right) < \infty \tag{2.6}$$

for all $\mu \geq 0$. In the proof of Theorem 2.1 we use heavily the corresponding result for the Chandrasekhar case, which we state next.

Theorem 2.2 (Energies of Chandrasekhar hydrogen). *There is a constant $C < \infty$ such that for all $l \in \mathbb{N}_0$, $n \in \mathbb{N}$ and $\kappa \in (0, \kappa^C]$ one has*

$$-C\frac{\kappa^2}{(n+l)^2} \leq \lambda_n(c_l(\kappa)) \leq -\frac{\kappa^2}{2(n+l)^2}. \tag{2.7}$$

We break the proofs of Theorems 2.1 and 2.2 into three parts, corresponding to the upper bound and the lower bound for subcritical and, respectively, critical values of the coupling constant.

2.1.1 Upper bound on hydrogen eigenvalues

We begin with the Chandrasekhar case.

Proof of Theorem 2.2. Upper bound. The second inequality in (2.7) is an immediate consequence of the inequality $\sqrt{p^2+1}-1 \leq p^2/2$ and the known form of the Schrödinger eigenvalues in the subspace corresponding to fixed angular momentum l. $\square$

Next, we turn to the Brown-Ravenhall case.

Proof of Theorem 2.1. Upper bound. We first recall some facts about the eigenvalues of the hydrogenic Dirac operator $D_\kappa := \boldsymbol{\alpha}\cdot\mathbf{p}+\beta-\kappa|\mathbf{x}|^{-1}$; see Darwin [8], Gordon [32] and also Bethe and Salpeter [5] for a textbook presentation. The following subspaces of $L^2(\mathbb{R}^3,\mathbb{C}^4)$,

$$\tilde{\mathfrak{H}}_{j,l,m} = \left\{\mathbf{x}\mapsto \begin{pmatrix} ir^{-1}f(r)\Omega_{j,l,m}(\omega_\mathbf{x}) \\ -r^{-1}g(r)\Omega_{j,2j-l,m}(\omega_\mathbf{x})\end{pmatrix} :\ f,g\in L^2(\mathbb{R}_+)\right\},$$

reduce the Dirac operator D_κ with $\kappa\in(0,1)$. Under the natural identification of $\tilde{\mathfrak{H}}_{j,l,m}$ with $L^2(\mathbb{R}_+,\mathbb{C}^2)$ the part of D_κ in $\tilde{\mathfrak{H}}_{j,l,m}$ is unitarily equivalent to

$$d_{j,l}(\kappa) = \begin{pmatrix} 1-\frac{\kappa}{r} & -\frac{\mathrm{d}}{\mathrm{d}r}-\frac{(j-l)(2j+1)}{r} \\ \frac{d}{dr}-\frac{(j-l)(2j+1)}{r} & -1-\frac{\kappa}{r}\end{pmatrix}.$$

The non-decreasing sequence $\lambda_n(d_{j,l}(\kappa))$ of eigenvalues of $d_{j,l}(\kappa)$ in the gap $(-1,1)$ is independent of l and given explicitly by

$$\lambda_n(d_{j,l}(\kappa)) = \left(1-\frac{\kappa^2}{\left(n-1+\sqrt{(j+1/2)^2-\kappa^2}\right)^2+\kappa^2}\right)^{1/2}, \quad n\in\mathbb{N}. \tag{2.8}$$

The Dirac eigenvalues reduced by the rest energy are bounded from above by the Schrödinger eigenvalues: for all n, l, j, and $\kappa\in(0,1)$

$$1-\lambda_n(d_{j,l}(\kappa)) \geq \tfrac{\kappa^2}{2(n+l)^2} = -\lambda_n(s_l(\kappa)). \tag{2.9}$$

To show (2.9), we use

$$\sqrt{(j+1/2)^2-\kappa^2} \leq \sqrt{(l+1)^2-\kappa^2} \leq \sqrt{(n+l)^2-\kappa^2}+1-n$$

and expand the outer square root in (2.8) up to first order which gives an upper bound.

Hence the upper bound in Theorem 2.1 will follow if we can show that

$$\lambda_n(b_{j,l}(\kappa)) \leq -1+\lambda_n(d_{j,l}(\kappa)). \tag{2.10}$$

To prove this, we fix (j,l) and abbreviate $\Lambda_+ := \chi_{[1,\infty)}(d_{j,l}(0))$ and $\Lambda_- := 1-\Lambda_+$. It follows from the definition of the Brown-Ravenhall operator that $b_{j,l}(\kappa)$ is unitarily equivalent to the operator $\Lambda_+(d_{j,l}-1)\Lambda_+$ in the Hilbert space $\Lambda_+(L^2(\mathbb{R}_+,\mathbb{C}^2))$. The variational principle for eigenvalues in gaps by Griesemer et al. [34, 35] under the weakened hypotheses of Dolbeault et al. [11] states that

$$\begin{aligned}&\lambda_n(d_{j,l}(\kappa))\\ &\quad= \inf_{\substack{V\subset\Lambda_+(L^2(\mathbb{R}_+,\mathbb{C}^2)),\\ \dim V=n}} \sup\left\{\frac{(f,d_{j,l}(\kappa)f)}{\|f\|^2}:\ 0\neq f\in V\oplus\Lambda_-(L^2(\mathbb{R}_+,\mathbb{C}^2))\right\}.\end{aligned}$$

Since the supremum decreases when restricted to $0\neq f\in V$, one obtains (2.10). □

2.1.2 Lower bounds on hydrogen eigenvalues. Subcritical case

Proof of Theorem 2.2. Subcritical case. Since we will reduce the Brown-Ravenhall case in Theorem 2.1 to the Chandrasekhar case, we actually prove a slightly stronger statement. As explained in (C.8), the operators $c_l(\kappa)$ are lower bounded for all $l\geq 1$ up to $\kappa_l^C>\kappa^B$.

We assume that either $l\geq 1$ and $0<\kappa\leq\kappa^B$ or else that $l=0$ and $0<\kappa\leq \kappa^B\kappa^C/\kappa_1^C$. For any $0<\delta<1$ there exist $M_\delta>0$ and $c_\delta>0$ such that

$$\sqrt{p^2+1}-1\geq\begin{cases}(1-\delta)p & \text{if } p\geq M_\delta\\ c_\delta\,p^2/2 & \text{if } p\leq M_\delta.\end{cases}.$$

Denoting by χ_i the characteristic function of the centered ball in $\mathbb{R}^3$ with radius M_δ, and putting $\chi_o := 1-\chi_i$, the Schwarz inequality implies the operator inequality

$$|\mathbf{x}|^{-1}\leq(1+\delta^{-1})\chi_i(\mathbf{p})|\mathbf{x}|^{-1}\chi_i(\mathbf{p})+(1+\delta)\chi_o(\mathbf{p})|\mathbf{x}|^{-1}\chi_o(\mathbf{p}),$$

and hence

$$\begin{aligned}\sqrt{\mathbf{p}^2+1}-1-\kappa|\mathbf{x}|^{-1}\geq&\chi_i(\mathbf{p})\left(c_\delta\mathbf{p}^2/2-(1+\delta^{-1})\kappa|\mathbf{x}|^{-1}\right)\chi_i(\mathbf{p})\\ &+\chi_o(\mathbf{p})\left((1-\delta)|\mathbf{p}|-(1+\delta)\kappa|\mathbf{x}|^{-1}\right)\chi_o(\mathbf{p}).\end{aligned}\tag{2.11}$$

Now choose δ as the the unique solution of the equation $(1+\delta)/(1-\delta)=\kappa_1^C/\kappa^B$ in the interval $(0,1)$. Then the restrictions on κ imply that $(1+\delta)\kappa\leq(1-\delta)\kappa_1^C\leq(1-\delta)\kappa_l^C$ for $l\geq 1$ and $(1+\delta)\kappa\leq(1-\delta)\kappa^C$ for $l=0$. In any case, the second operator in the above sum is non-negative. The variational principle hence implies that the n-th eigenvalue of $c_l(\kappa)$ is greater or equal to the n-th eigenvalue of $\chi_i(\mathbf{p})\left(c_\delta\mathbf{p}^2/2-(1+\delta^{-1})\kappa|\mathbf{x}|^{-1}\right)\chi_i(\mathbf{p})$. Again by the variational principle, the latter is greater or equal to the n-th eigenvalue of $c_\delta\mathbf{p}^2/2-(1+\delta^{-1})\kappa|\mathbf{x}|^{-1}$, which is $-\mathrm{const}\,\kappa^2(n+l)^{-2}$. □

Proof of Theorem 2.1. Subcritical case. We assume that either $j \geq 3/2$ and $0 < \kappa \leq \kappa^B$ or else that $j = 1/2$ and $0 < \kappa \leq \kappa^B \kappa^C / \kappa_1^C$. We claim that

$$\lambda_n(c_l(\kappa)) = \lambda_{2n-1}(c_l(\kappa) \otimes 1_{\mathbb{C}^2}) \leq \lambda_{2n-1}(b_{j,l}(\kappa))\,. \tag{2.12}$$

Once we have proved this, the assertion follows easily from what we have shown in the proof of Theorem 2.2 above.
To establish (2.12) we use the same notation as in the proof of the upper bound in Theorem 2.1. By the variational principle,

$$\begin{aligned}
&\lambda_n(b_{j,l}(\kappa)) \\
&= \sup_{\substack{f_1,\dots,f_{n-1}\in \\ \Lambda_+(L^2(\mathbb{R}_+,\mathbb{C}^2))}} \inf\{\langle f, (d_{j,l}(\kappa)-1)f\rangle :\ \|f\|=1, f\in \Lambda_+(L^2(\mathbb{R}_+,\mathbb{C}^2)), f\perp f_\nu\} \\
&= \sup_{\substack{f_1,\dots,f_{n-1}\in \\ L^2(\mathbb{R}_+,\mathbb{C}^2)}} \inf\{\langle \mathcal{F}_l f, c_l(\kappa)\mathcal{F}_l f\rangle :\ \|f\|=1, f\in \Lambda_+(L^2(\mathbb{R}_+,\mathbb{C}^2)), f\perp f_\nu\}
\end{aligned}$$

with $\mathcal{F}_l$ the Fourier-Bessel transform, see (A.5). The infimum does not increase if the condition $f \in \Lambda_+(L^2(\mathbb{R}_+,\mathbb{C}^2))$ is relaxed to $f \in L^2(\mathbb{R}_+,\mathbb{C}^2)$. This gives the eigenvalues of the operator $c_l(\kappa)\otimes 1_{\mathbb{C}^2}$, proving (2.12). $\square$

2.1.3 Lower bounds on hydrogen eigenvalues. Critical case

Proof of Theorem 2.2. Critical case. It remains to prove that

$$\lambda_n(c_0(\kappa)) \geq -\mathrm{const}\, \kappa^2 n^{-2}$$

for $\kappa^B\kappa^C/\kappa_1^C \leq \kappa \leq \kappa^C$. We may assume that $\kappa = \kappa^C$ and will prove that for all $\tau > 0$

$$N(-\tau, c_0(\kappa^C)) := \operatorname{tr} \chi_{(-\infty,-\tau)}(c_0(\kappa^C)) \leq \mathrm{const}\, \tau^{-1/2}. \tag{2.13}$$

Let $\chi_i^2 + \chi_o^2 = 1$ be a smooth radial quadratic partition of unity with χ_i supported in the unit ball and χ_o supported outside the ball of radius 1/2 about the origin. It was shown in [31, Eq. (19)] that the localization error can be estimated by a bounded exponentially decaying potential $v(r) \leq \mathrm{const}\, e^{-r}$, i.e.,

$$\begin{aligned}
\sqrt{\mathbf{p}^2+1} - 1 - \kappa^C|\mathbf{x}|^{-1} \geq\ & \chi_i\left(\sqrt{\mathbf{p}^2+1} - 1 - \kappa^C|\mathbf{x}|^{-1} - v(|\mathbf{x}|)\right)\chi_i \\
&+ \chi_o\left(\sqrt{\mathbf{p}^2+1} - 1 - \kappa^C|\mathbf{x}|^{-1} - v(|\mathbf{x}|)\right)\chi_o.
\end{aligned}$$

By the variational principle it suffices to consider the eigenvalue counting function corresponding to the interior and exterior term separately. The interior term is further estimated according to

$$\begin{aligned}
&\chi_i\left(\sqrt{\mathbf{p}^2+1} - 1 - \kappa^C|\mathbf{x}|^{-1} - v(|\mathbf{x}|)\right)\chi_i \\
&\geq \chi_i\left(|\mathbf{p}| - \kappa^C|\mathbf{x}|^{-1} - \mathrm{const}\,\right)\chi_i.
\end{aligned}$$

As shown by Lieb and Yau [46] and explained in Corollary D.1, the number of negative eigenvalues of the latter operator acting in the subspace corresponding to $l=0$ is finite, i.e., for all $\tau>0$

$$N_{l=0}\left(-\tau, \chi_i\left(|\mathbf{p}|-\kappa^C|\mathbf{x}|^{-1}-\text{const}\ \right)\chi_i\right)\leq \text{const}\ . \tag{2.14}$$

For the exterior problem, we note that by the variational principle

$$\begin{aligned} N_{l=0}&\left(-\tau, \chi_o\left(\sqrt{\mathbf{p}^2+1}-1-\kappa^C|\mathbf{x}|^{-1}-v(|\mathbf{x}|)\right)\chi_o\right) \\ &\leq N_{l=0}\left(-\tau, \sqrt{\mathbf{p}^2+1}-1-\chi(\mathbf{x})(\kappa^C|\mathbf{x}|^{-1}+v(|\mathbf{x}|))\right) \end{aligned} \tag{2.15}$$

where χ denotes the characteristic function of the support of χ_o. With the singularity gone, the result follows as in the subcritical case. Namely, similarly as in (2.11) we cut in momentum space according to small and large momenta. Again, by the variational principle, the right-hand side of (2.15) is then bounded from above by

$$N_{l=0}(-\text{const}\ \tau, |\mathbf{p}|-w(|\mathbf{x}|))+N_{l=0}(-\text{const}\ \tau, p^2-w(|\mathbf{x}|)),$$

where $w(r)=\text{const}\ \chi(r)(\kappa^C r^{-1}+v(r))$. The first term is estimated with the help of Daubechies' inequality [9]

$$\begin{aligned} N_{l=0}(-\tau, |\mathbf{p}|-w(|\mathbf{x}|)) &\leq \tau^{-1/2}\operatorname{tr}_{l=0}(|\mathbf{p}|-w(|\mathbf{x}|))_-^{1/2} \\ &\leq \text{const}\ \tau^{-1/2}\int_0^\infty w(r)^{3/2}\,dr \end{aligned}$$

with the latter integral being finite. For the second term we estimate $w(r)\leq \text{const}\ r^{-1}$ and use that

$$N_{l=0}(-\tau, p^2-\text{const}\ |\mathbf{x}|^{-1})\leq \text{const}\ \tau^{-1/2}.$$

This concludes the proof of Theorem 2.2. □

Our proof of Theorem 2.1 in the critical Brown-Ravenhall case is based on a reduction to the Chandrasekhar case. The next lemma compares the number of eigenvalues of the critical operators $b_{1/2,l}(\kappa^B)$ with those of the two operators $c_{l'}(\kappa^C_{l'})$ with $l'=0,1$ and critical coupling constants $\kappa^C_0=2/\pi$ and $\kappa^C_1=\pi/2$, cf. (C.8).

Lemma 2.3. *There is a constant such that for $l=0,1$ and all $\tau>0$ one has*

$$N\left(-\tau, b_{1/2,l}(\kappa^B)\right)\leq \text{const}\ \left[N\left(-\tfrac{\tau}{\text{const}}, c_0(\kappa^C_0)\right)+N\left(-\tfrac{\tau}{\text{const}}, c_1(\kappa^C_1)\right)\right].$$

Proof. We start with the observation that $(\kappa^C_0)^{-1}+(\kappa^C_1)^{-1}=2(\kappa^B)^{-1}$. Using the explicit form of the reduced operators (cf. Appendix C), this implies the identities

$$\begin{aligned} b_{1/2,0}(\kappa^B) &= \kappa^B\left((\kappa^C_0)^{-1}\phi_0\tilde{b}_{0,0}\phi_0+(\kappa^C_1)^{-1}\phi_1\tilde{b}_{1,1}\phi_1\right), \\ b_{1/2,1}(\kappa^B) &= \kappa^B\left((\kappa^C_0)^{-1}\phi_1\tilde{b}_{0,1}\phi_1+(\kappa^C_1)^{-1}\phi_0\tilde{b}_{1,0}\phi_0\right), \end{aligned} \tag{2.16}$$

where the operators $\tilde{b}_{l,\nu}$ are defined in $L^2(\mathbb{R}_+)$ through quadratic forms

$$\langle f, \tilde{b}_{l,\nu} f\rangle := \int_0^\infty \frac{E(p)-1}{2\phi_\nu(p)^2}|f(p)|^2\,\mathrm{d}p - \kappa_l^C \int_0^\infty\int_0^\infty \overline{f(p)} k_l^C(p,q) f(q)\,\mathrm{d}p\,\mathrm{d}q.$$

In case $\nu = 1$ it hence follows from $2\phi_1(p)^2 \le 1$ that $\langle f, \tilde{b}_{l,1} f\rangle \ge \langle f, c_l(\kappa_l^C) f\rangle$. In case $\nu = 0$ we use the inequality

$$(E(p)-1)\phi_0(p)^{-2} \ge \sqrt{p^2+4} - 2 = 2(E(p/2)-1) \tag{2.17}$$

which is most easily seen by writing both sides in terms of $E(p)$. It implies

$$\langle f, \tilde{b}_{l,0} f\rangle \ge 2\langle uf, c_l(\kappa_l^C)\, uf\rangle$$

where the unitary scaling transformation u is defined through $(uf)(p) := \sqrt{2} f(2p)$. The proof is completed by the variational principle. □

We are now ready to give a

Proof of Theorem 2.1. Critical case. The previous lemma implies that it suffices to show that for $l = 0, 1$

$$N\left(-\tau, c_l(\kappa_l^C)\right) \le \text{const}\ \tau^{-1/2}.$$

In case $l = 0$ this was established in (2.13), and the case $l = 1$ follows similarly with the analogue of (2.14) given in Corollary D.1. □

2.2 Sobolev inequality for the critical Brown-Ravenhall operator

Having studied the eigenvalues of B_κ in the previous subsection, we now turn to integrability properties of its eigenfunctions. The L^q-norm of two-spinors ψ is given by

$$\|\psi\|_q := \left(\int_{\mathbb{R}^3} |\psi(\mathbf{x})|^q \mathrm{d}\mathbf{x}\right)^{1/q},$$

where the modulus, $|\cdot|$, refers to the Euclidean norm in $\mathbb{C}^2$. For $q = 2$ we drop the subscript. We aim at proving the following

Theorem 2.4 (L^q-properties of eigenfunctions). *Let $2 \le q < 3$. There exists a constant $C_q < \infty$ such that for any $\kappa \in (0, \kappa^B]$ and all $\psi \in \mathfrak{Q}(B_\kappa)$ with $\langle \psi, B_\kappa \psi\rangle \le 0$ one has $\psi \in L^q$ with*

$$\|\psi\|_q \le C_q \|\psi\|. \tag{2.18}$$

Note that (2.18) applies, in particular, to eigenfunctions of B_κ corresponding to negative eigenvalues. The proof of Theorem 2.4, which is spelled out below, relies on a Sobolev inequality for the massless atomic Brown-Ravenhall operator in $\mathfrak{H}$ given by

$$B_\kappa^{(0)} := |\mathbf{p}| - \frac{\kappa}{2}\left(|\mathbf{x}|^{-1} + \omega_\mathbf{p}\cdot\boldsymbol{\sigma}\ |\mathbf{x}|^{-1}\ \omega_\mathbf{p}\cdot\boldsymbol{\sigma}\right).$$

This operator is bounded below (in fact, non-negative) if and only if $\kappa \le \kappa^B$.

Theorem 2.5 (Sobolev inequality). *For any $2 \le q < 3$ there exists a constant $C_q > 0$ such that for all $\psi \in \mathfrak{Q}(B_{\kappa^B}^{(0)})$,*

$$\|\psi\|_q^2 \le C_q \left\langle \psi, B_{\kappa^B}^{(0)}\psi\right\rangle^\theta \|\psi\|^{2(1-\theta)}, \qquad \theta = 6(\tfrac{1}{2} - \tfrac{1}{q}). \tag{2.19}$$

It is illustrative to compare (2.19) with the 'standard' Sobolev-Gagliardo-Nirenberg inequalities,

$$\|\psi\|_q^2 \le C_q'\ \langle \psi, |\mathbf{p}|\psi\rangle^\theta\ \|\psi\|^{2(1-\theta)}, \qquad \theta = 6(\tfrac{1}{2} - \tfrac{1}{q}), \qquad 2 \le q \le 3, \tag{2.20}$$

see, e.g., [43, Thm. 8.4]. Hence Theorem 2.5 says that, if the endpoint $q = 3$ is avoided, an inequality of the same form remains true after subtracting the maximal possible multiple of $\frac{1}{2}\left(|\mathbf{x}|^{-1} + \omega_\mathbf{p}\cdot\boldsymbol{\sigma}\ |\mathbf{x}|^{-1}\ \omega_\mathbf{p}\cdot\boldsymbol{\sigma}\right)$ from $|\mathbf{p}|$. Moreover, one can show that (2.19) does not hold with $q = 3$, not even if the L^3-norm is replaced by the weak L^3-norm.

Note that if $\kappa < \kappa^B$ then (2.19) with $B_\kappa^{(0)}$ instead of $B_{\kappa^B}^{(0)}$ follows from (2.20) – but with a constant that deteriorates as $\kappa \to \kappa^B$. The main point is to derive an inequality which holds uniformly in κ up to and including the critical constant. Our proof is based on the somewhat surprising fact that the Brown-Ravenhall operator with coupling constant κ^B can be bounded from below by the Chandrasekhar operator with *smaller* coupling constant κ^C.

Before we start the proof of (2.19), we provide the

Proof of Theorem 2.4. The Sobolev inequality (2.19) implies

$$\begin{aligned}\|\psi\|_q^2 &\le C_q \left\langle \psi, B_\kappa^{(0)}\psi\right\rangle^\theta \|\psi\|^{2(1-\theta)} \le C_q \left\langle \psi, \left[B_\kappa^{(0)} - B_\kappa\right]\psi\right\rangle^\theta \|\psi\|^{2(1-\theta)} \\ &\le C_q \|B_\kappa^{(0)} - B_\kappa\|^\theta \|\psi\|^2.\end{aligned}$$

Tix showed [75, Thm. 1] (see also Balinsky and Evans [3]) that the difference $B_\kappa^{(0)} - B_\kappa$ extends to a bounded operator with norm uniformly bounded for any $\kappa \in (0, \kappa^B]$. □

2.2.1 Comparison of critical operators

The first step in the proof of the Sobolev inequality (2.19) is a comparison of $B_\kappa^{(0)}$ with the massless atomic Chandrasekhar operator in $\mathfrak{H}$, which is given by

$$C_\kappa^{(0)} := |\mathbf{p}| - \kappa|\mathbf{x}|^{-1}.$$

It is bounded below if and only if $\kappa \leq \kappa^C$. As discussed in Appendix C the parts of $B_\kappa^{(0)}$ and $C_\kappa^{(0)}$ in the subspace $\mathfrak{H}_{j,l,m}$ are unitarily equivalent to operators $b_j^{(0)}(\kappa)$ and $c_l^{(0)}(\kappa)$ in $L^2(\mathbb{R}_+)$, which depend only on j in the Brown-Ravenhall case and only on l in the Chandrasekhar case. For the comparison argument it is important to note that the reduced operators $b_j^{(0)}(\kappa)$ and $c_l^{(0)}(\kappa)$ are lower bounded for κ up to and including the critical coupling constants κ_j^B and κ_l^C respectively. They are defined in (C.7) and, as is explained there, exceed κ^B and κ^C, if $j \geq 3/2$ or $l \geq 1$.
We begin by observing that all the critical operators $b_j^{(0)}(\kappa_j^B)$ and $c_l^{(0)}(\kappa_l^C)$ have the same 'generalized ground state', namely p. The corresponding ground state representation formula (in momentum space) is given in

Lemma 2.6 (Ground state representation). *If $f \in \mathfrak{Q}(b_j^{(0)}(\kappa_j^B))$ and $g(p) = pf(p)$, then*

$$\langle f, b_j^{(0)}(\kappa_j^B) f\rangle = \frac{\kappa_j^B}{2}\int_0^\infty\int_0^\infty |g(p)-g(q)|^2 k_j^B(p,q)\,\frac{\mathrm{d}p}{p}\frac{\mathrm{d}q}{q}. \tag{2.21}$$

Similarly, if $f \in \mathfrak{Q}(c_l^{(0)}(\kappa_l^C))$ and $g(p) = pf(p)$, then

$$\langle f, c_l^{(0)}(\kappa_l^C) f\rangle = \frac{\kappa_l^C}{2}\int_0^\infty\int_0^\infty |g(p)-g(q)|^2 k_l^C(p,q)\,\frac{\mathrm{d}p}{p}\frac{\mathrm{d}q}{q}. \tag{2.22}$$

where k_j^B and k_l^C are defined in (C.4).

Proof. We write k for one of the functions k_j^B or k_l^C and κ for the corresponding constant κ_j^B or κ_l^C. Expanding the square and using $k(p,q) = k(q,p)$, we find

$$\begin{aligned}
&\frac{1}{2}\int_0^\infty\int_0^\infty |g(p)-g(q)|^2 k(p,q)\,\frac{\mathrm{d}p}{p}\frac{\mathrm{d}q}{q}\\
&=\int_0^\infty |g(p)|^2\left(\int k(p,q)\,\frac{\mathrm{d}q}{q}\right)\frac{\mathrm{d}p}{p} - \int_0^\infty\int_0^\infty \overline{g(p)}k(p,q)g(q)\frac{\mathrm{d}p}{p}\frac{\mathrm{d}q}{q}\\
&=\int_0^\infty p|f(p)|^2\left(\int k(p,q)\,\frac{\mathrm{d}q}{q}\right)\mathrm{d}p - \int_0^\infty\int_0^\infty \overline{f(p)}k(p,q)f(q)\,\mathrm{d}p\,\mathrm{d}q.
\end{aligned}$$

By definitions (C.7) and (C.8) of κ we have

$$\int_0^\infty k(p,q)\,\frac{\mathrm{d}q}{q} = \kappa^{-1},$$

which implies the assertion. ☐

Next, we bound $B_{\kappa^B}^{(0)}$ from below by $C_{\kappa^C}^{(0)}$.

Lemma 2.7 (Comparison of critical operators). *There is a positive constant such that for any $\psi \in \mathfrak{Q}(B_{\kappa^B}^{(0)}) \cap \mathfrak{H}_{1/2,1}^\perp$*

$$\left\langle \psi, B_{\kappa^B}^{(0)}\psi\right\rangle \geq \text{const}\ \left\langle \psi, C_{\kappa^C}^{(0)}\psi\right\rangle. \tag{2.23}$$

An inequality of the form (2.23) cannot hold for $\psi \in \mathfrak{H}_{1/2,1}$, since the right hand side is bounded from below by a constant times $\langle \psi, |\mathbf{p}|\psi\rangle$ while the left hand side is not.

Proof. By orthogonality it suffices to prove the inequality on each subspace $\mathfrak{H}_{j,l}$. First let $(j,l) = (1/2,0)$. We may also fix $m = \pm 1/2$ and choose $\psi \in \mathfrak{H}_{1/2,0,m}$. Its Fourier transform is of the form $\hat\psi(\mathbf{p}) = p^{-1} f(p)\Omega_{\frac{1}{2},0,m}(\omega_{\mathbf{p}})$, see Appendix A. Setting $f(p) =: pg(p)$ one finds using (C.2) and Lemma 2.6

$$\langle \psi, B^{(0)}_{\kappa^B}\psi\rangle = \langle f, b^{(0)}_{1/2}(\kappa^B) f\rangle = \frac{\kappa^B}{2}\int_0^\infty\int_0^\infty |g(p)-g(q)|^2\, k^B_{1/2}(p,q)\,\frac{\mathrm{d}p}{p}\frac{\mathrm{d}q}{q}.$$

Similarly, using (C.3)

$$\langle \psi, C^{(0)}_{\kappa^C}\psi\rangle = \langle f, c^{(0)}_{0}(\kappa^C) f\rangle = \frac{\kappa^C}{2}\int_0^\infty\int_0^\infty |g(p)-g(q)|^2\, k^C_{0}(p,q)\,\frac{\mathrm{d}p}{p}\frac{\mathrm{d}q}{q}.$$

Recalling the explicit expressions (C.4) and (C.5) for $k^B_{1/2}$ and k^C_0 and estimating $Q_1 \geq 0$ we conclude that

$$\langle \psi, B^{(0)}_{\kappa^B}\psi\rangle \geq (1+(2/\pi)^2)^{-1}\langle \psi, C^{(0)}_{\kappa^C}\psi\rangle,$$

which proves the assertion on the subspace $\mathfrak{H}_{1/2,0}$. Now assume that $\psi \in \left(\mathfrak{H}_{1/2,0}\oplus\mathfrak{H}_{1/2,1}\right)^{\perp}$. Since κ^B_j is monotone increasing in j, see Appendix C, we have on that subspace

$$|\mathbf{p}| \geq \frac{\kappa^B_{3/2}}{2}\left(|\mathbf{x}|^{-1} + \omega_{\mathbf{p}}\cdot\boldsymbol{\sigma}\;|\mathbf{x}|^{-1}\;\omega_{\mathbf{p}}\cdot\boldsymbol{\sigma}\right).$$

We conclude that

$$\langle \psi, B^{(0)}_{\kappa^B}\psi\rangle \geq \frac{\kappa^B_{3/2}-\kappa^B_{1/2}}{\kappa^B_{3/2}}\langle \psi, |\mathbf{p}|\psi\rangle \geq \frac{\kappa^B_{3/2}-\kappa^B_{1/2}}{\kappa^B_{3/2}}\langle \psi, C^{(0)}_{\kappa^C}\psi\rangle,$$

proving the assertion. □

2.2.2 Proof of the Sobolev inequality

We are now ready to give a

Proof of Theorem 2.5. By scaling, (2.19) is equivalent to the inequality

$$\|\psi\|_q^2 \leq C'_q\,\left(\left\langle \psi, B^{(0)}_{\kappa^B}\psi\right\rangle + \|\psi\|^2\right).$$

This, together with the triangle inequality, shows that it is enough to prove the inequality separately on the subspaces $\mathfrak{H}_{1/2,1}$ and $\mathfrak{H}^{\perp}_{1/2,1}$. On the latter subspace, the claim follows immediately from Lemma 2.7 above and the Sobolev inequality for the critical Chandrasekhar operator [30, Corollary 2.5]. We now reduce the claim for the subspace $\mathfrak{H}_{1/2,1}$ to that for $\mathfrak{H}_{1/2,0}$. For this purpose, we note that the helicity operator $\mathrm{H} = \omega_{\mathbf{p}} \cdot \boldsymbol{\sigma}$, cf. (B.1), commutes with $B_\kappa^{(0)}$ and, by (B.2), maps $\mathfrak{H}_{j,l}$ into $\mathfrak{H}_{j,2j-l}$. Hence if $\psi \in \mathfrak{H}_{1/2,1}$ then by the Sobolev inequality on $\mathfrak{H}_{1/2,0}$

$$\left\langle \psi, B_{\kappa^B}^{(0)} \psi \right\rangle + \|\psi\|^2 = \left\langle \mathrm{H}\psi, B_{\kappa^B}^{(0)} \mathrm{H}\psi \right\rangle + \|\mathrm{H}\psi\|^2 \geq \mathrm{const}\, \|\mathrm{H}\psi\|_q^2.$$

By Lemma B.1 the helicity $\mathrm{H} = \mathrm{H}^{-1}$ is bounded on $L^q(\mathbb{R}^3, \mathbb{C}^2)$. □

2.3 Estimates on the electric potential

The goal of this subsection is to compare twisted and untwisted electric potentials. We begin with an estimates for point charges and then turn to smeared out charges.

Lemma 2.8. *Let $l \geq 1$ and $\psi \in \mathfrak{H}_{j,l}$. Then*

$$\left|\left\langle \psi, \left(|\mathbf{x}|^{-1} - \mathcal{U}(|\mathbf{x}|^{-1})\right) \psi \right\rangle\right| \leq \frac{\mathrm{const}}{l^2} \langle \psi, \mathbf{p}^2 \psi \rangle. \tag{2.24}$$

Proof. By orthogonality it suffices to prove the assertion for $\psi \in \mathfrak{H}_{j,l,m}$. Its Fourier transform is of the form $\hat{\psi}(\mathbf{p}) = f(p) p^{-1} \Omega_{j,l,m}(\omega_{\mathbf{p}})$, cf. Appendix A, and we compute similarly as in (C.11)

$$\begin{aligned}
&\langle \psi, \left(|\mathbf{x}|^{-1} - \mathcal{U}(|\mathbf{x}|^{-1})\right) \psi \rangle \\
&= \frac{1}{\pi} \int_0^\infty \mathrm{d}p\, \overline{f(p)} \int_0^\infty \mathrm{d}q\, f(q) \Big\{ [1 - \phi_0(p)\phi_0(q)]\, Q_l\left(\tfrac{1}{2}\left(\tfrac{q}{p} + \tfrac{p}{q}\right)\right) \\
&\qquad\qquad - \phi_1(p)\phi_1(q) Q_{2j-l}\left(\tfrac{1}{2}\left(\tfrac{q}{p} + \tfrac{p}{q}\right)\right)\Big\} \\
&= \frac{1}{2\pi}(A_1 + A_2)
\end{aligned}$$

with

$$\begin{aligned}
A_1 :=& \int_0^\infty \mathrm{d}p\, \overline{f(p)} \int_0^\infty \mathrm{d}q\, f(q) \sum_{\nu=0}^{1} \left(\phi_\nu(p) - \phi_\nu(q)\right)^2 Q_l\left(\tfrac{1}{2}\left(\tfrac{q}{p} + \tfrac{p}{q}\right)\right), \\
A_2 :=& 2 \int_0^\infty \mathrm{d}p\, \overline{f(p)} \int_0^\infty \mathrm{d}q\, f(q)\phi_1(p)\phi_1(q) \\
&\qquad \times \left[Q_l\left(\tfrac{1}{2}\left(\tfrac{q}{p} + \tfrac{p}{q}\right)\right) - Q_{2j-l}\left(\tfrac{1}{2}\left(\tfrac{q}{p} + \tfrac{p}{q}\right)\right)\right].
\end{aligned}$$

We estimate these terms separately. For the first term we use (B.5) and (B.6) together with Abel's argument to turn Hermitian integral operators into multiplication operators by means of the Schwarz inequality (see also [46, Ineq.

(6.9)]). Since the Q_l are positive, we obtain

$$\begin{aligned}A_1 \leq &\int_0^\infty \mathrm{d}p \frac{|f(p)|^2}{E(p)^4} \int_0^\infty \mathrm{d}q \left(\frac{p}{q}\right)^2 E(p)^2 \\ &\times \sum_{\nu=0}^{1} \left(\phi_\nu(p) - \phi_\nu(q)\right)^2 E(q)^2 Q_l\left(\tfrac12\left(\tfrac{q}{p}+\tfrac{p}{q}\right)\right) \\ \leq &\frac58 \int_0^\infty \mathrm{d}p \frac{|f(p)|^2}{E(p)^4} \int_0^\infty \mathrm{d}q \left(\frac{p}{q}\right)^2 (p-q)^2 Q_l\left(\tfrac12\left(\tfrac{q}{p}+\tfrac{p}{q}\right)\right) \\ = &\frac58 \int_0^\infty \mathrm{d}p \frac{|f(p)|^2}{E(p)^4} p^3 \int_0^\infty \frac{\mathrm{d}q}{q^2} (1-q)^2 Q_l\left(\tfrac12\left(q+q^{-1}\right)\right).\end{aligned}$$

We now use the bounds $p^3/E(p)^4 \leq p^2$ and, for $q \geq 1$, $(1-q)^2 \leq q^2 - 1$ which yield

$$\begin{aligned}\int_0^\infty \frac{\mathrm{d}q}{q^2}(1-q)^2 Q_l\left(\tfrac12\left(q+q^{-1}\right)\right) &= 2\int_1^\infty \frac{\mathrm{d}q}{q^2}(1-q)^2 Q_l\left(\tfrac12\left(q+q^{-1}\right)\right) \\ &\leq 4\int_1^\infty \mathrm{d}x\, Q_l(x) = \frac{4}{l(l+1)},\end{aligned}$$

where the last step involved [16, 324(18)]. Thus,

$$A_1 \leq \frac{5}{2l(l+1)} \int_0^\infty \mathrm{d}p\, p^2 |f(p)|^2 = \frac{5}{2l(l+1)} \langle \psi, \mathbf{p}^2 \psi \rangle.$$

We estimate the term A_2 similarly by the Schwarz inequality,

$$\begin{aligned}A_2 \leq &2\int_0^\infty \mathrm{d}p |f(p)|^2 \, |\phi_1(p)|^2 \\ &\times \int_0^\infty \mathrm{d}q \, \frac{p}{q} \left| Q_l\left(\tfrac12\left(\tfrac{q}{p}+\tfrac{p}{q}\right)\right) - Q_{2j-l}\left(\tfrac12\left(\tfrac{q}{p}+\tfrac{p}{q}\right)\right)\right| \\ \leq &4\int_0^\infty \mathrm{d}p |f(p)|^2 p^2 \\ &\times \int_1^\infty \frac{\mathrm{d}q}{q} \left| Q_l\left(\tfrac12\left(q+q^{-1}\right)\right) - Q_{2j-l}\left(\tfrac12\left(q+q^{-1}\right)\right)\right|.\end{aligned}$$

Due to the pointwise monotonicity (C.10) the difference inside the modulus is of definite sign. Without loss of generality, we may therefore assume $2j = 2l+1$. Using the integral representation (C.1) we can bound

$$\begin{aligned}&\int_1^\infty \frac{\mathrm{d}q}{q} \left[Q_l\left(\tfrac12\left(q+q^{-1}\right)\right) - Q_{l+1}\left(\tfrac12\left(q+q^{-1}\right)\right)\right] \\ = &\int_1^\infty \mathrm{d}z z^{-l-2}(z-1) \int_1^{\frac12(z+z^{-1})} \frac{\mathrm{d}x}{\sqrt{x^2-1}} \frac{1}{\sqrt{1-2xz+z^2}} \\ \leq &\frac{\pi}{\sqrt2} \int_1^\infty \mathrm{d}z z^{-l-5/2}(z-1) = \frac{\pi}{\sqrt2 (l+\frac12)(l+\frac32)}.\end{aligned}$$

Adding the estimates for A_1 and A_2 we arrive at (2.24). $\square$

Note that our proof shows that one can choose different powers of $|\mathbf{p}|$ on the right hand side of (2.24).

Lemma 2.9. *There exists a constant such that for any electric potential v of a spherically symmetric non-negative charge density*

$$|\langle\psi,(v-\mathcal{U}(v))\,\psi\rangle| \leq \mathrm{const}\ v(0)\,\langle\psi,\mathbf{p}^2\psi\rangle.$$

Proof. We denote by $\tau:\mathbb{R}^3\to[0,\infty)$ the spherically symmetric, non-negative charge density corresponding to v, i.e., $v(\mathbf{x})=\int\tau(\mathbf{x}-\mathbf{y})\,|\mathbf{y}|^{-1}\mathrm{d}\mathbf{y}$. The Fourier transform of τ obeys the estimates

$$|\hat\tau(\mathbf{p})| = \sqrt{\frac{2}{\pi\,\mathbf{p}^2}}\left|\int_0^\infty r\,\sin(|\mathbf{p}|r)\,\tau(r)\mathrm{d}r\right| \leq \frac{v(0)}{(2\pi)^{3/2}\,|\mathbf{p}|}.$$

By Fourier transform the scalar product on the left side of the assertion becomes

$$\begin{aligned}&\langle\psi,(v-\mathcal{U}(v))\,\psi\rangle\\ &\quad=\sqrt{\frac{2}{\pi}}\iint\hat\psi(\mathbf{p})^*\frac{\hat\tau(\mathbf{p}-\mathbf{q})}{|\mathbf{p}-\mathbf{q}|^2}\,(1-\Phi_0(\mathbf{p})\Phi_0(\mathbf{q})-\Phi_1(\mathbf{p})\Phi_1(\mathbf{q}))\,\hat\psi(\mathbf{q})\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}.\end{aligned}$$

Using Lemma B.2 we estimate the absolute value of the preceding expression from above by the sum of two terms, B_1 and B_2. The first term can be further bounded as follows,

$$\begin{aligned}B_1 &= \mathrm{const}\iint|\hat\tau(\mathbf{p}-\mathbf{q})||\hat\psi(\mathbf{p})||\hat\psi(\mathbf{q})|\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\\ &\leq \mathrm{const}\ v(0)\int\mathrm{d}\mathbf{p}\,|\hat\psi(\mathbf{p})|^2\int\left(\frac{|\mathbf{p}|}{|\mathbf{q}|}\right)^{5/2}\frac{1}{|\mathbf{p}-\mathbf{q}|}\,\mathrm{d}\mathbf{q}\\ &\leq \mathrm{const}\ v(0)\int|\hat\psi(\mathbf{p})|^2\mathbf{p}^2\,\mathrm{d}\mathbf{p},\end{aligned}$$

where we use the Schwarz inequality in the second step. The second term is estimated similarly

$$\begin{aligned}B_2 &= \mathrm{const}\iint|\hat\tau(\mathbf{p}-\mathbf{q})|\frac{\sqrt{|\mathbf{p}|\,|\mathbf{q}|}}{|\mathbf{p}-\mathbf{q}|}|\hat\psi(\mathbf{p})||\hat\psi(\mathbf{q})|\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\\ &\leq \mathrm{const}\ v(0)\int\mathrm{d}\mathbf{p}\,|\hat\psi(\mathbf{p})|^2\int\left(\frac{|\mathbf{p}|}{|\mathbf{q}|}\right)^{2}\frac{\sqrt{|\mathbf{p}|\,|\mathbf{q}|}}{|\mathbf{p}-\mathbf{q}|^2}\,\mathrm{d}\mathbf{q}\\ &\leq \mathrm{const}\ v(0)\int|\hat\psi(\mathbf{p})|^2|\mathbf{p}|^2\,\mathrm{d}\mathbf{p}.\end{aligned}$$

This proves the assertion. $\square$

3 Spectral shift from Schrödinger to Brown-Ravenhall operators

The main theme of this section is the (integrated) spectral shift, i.e., the difference of sums of eigenvalues of the Brown-Ravenhall or Chandrasekhar operator

$$B[v] := \sqrt{\mathbf{p}^2+1} - 1 - \mathcal{U}(v), \qquad C[v] := \sqrt{\mathbf{p}^2+1} - 1 - v,$$

(cf. (2.2)) and the Schrödinger operator $S[v] := \frac{1}{2}\mathbf{p}^2 - v$, all acting in the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}^3, \mathbb{C}^2)$ of two-spinors. We have set $c = 1$.
Concerning the potential $v : \mathbb{R}^3 \to \mathbb{R}$ we will always assume that the above operators can be defined through the Friedrichs extension starting from $\mathfrak{S}(\mathbb{R}^3, \mathbb{C}^2)$. For example, the condition $0 \le v(\mathbf{x}) \le \kappa^{\#} |\mathbf{x}|^{-1}$ with $\# = B,\ C$ (cf. (2.3)) ensures that the Brown-Ravenhall, respectively the Chandrasekhar operator are well-defined and bounded from below (see [17] and [40]).
We assume throughout that the potential v is radially symmetric which allows us to investigate the spectral shift on each subspace $\mathfrak{H}_{j,l}$ in the decomposition (2.4) separately. We write $\Lambda_{j,l}$ for the orthogonal projection onto $\mathfrak{H}_{j,l}$. For the reduced traces we use the notations

$$\mathrm{tr}_{j,l}(A) := \mathrm{tr}(\Lambda_{j,l} A), \qquad \mathrm{tr}_j(A) := \mathrm{tr}_{j,j+1/2}(A) + \mathrm{tr}_{j,j-1/2}(A). \tag{3.1}$$

3.1 Estimate on the spectral shift

One of the key observations in our proof of the Scott correction is that the spectral shift between the one-particle Brown-Ravenhall and the Schrödinger operator decreases sufficiently fast for high angular momenta.

Theorem 3.1 (Spectral shift: Brown-Ravenhall case). *There exists a constant $C < \infty$ such that for any $\kappa \le \kappa^B$, any $v : [0,\infty) \to [0,\infty)$ satisfying*

$$v(r) \le \kappa\, r^{-1}, \tag{3.2}$$

any $\mu > 0$ and any $j \in \mathbb{N}_0 + 1/2$ one has

$$\mathrm{tr}_j\left([B[v]+\mu]_- - [S[v]+\mu]_-\right) \le C\,\kappa^4\, j^{-2}. \tag{3.3}$$

We derive this result from a corresponding theorem for the Chandrasekhar operator. For a proof of the latter we need to strengthen [31, Thm. 2]. In particular, we need to consider $C[v]$ for potentials v satisfying (3.2) also in case $\kappa^C < \kappa \le \kappa^B$. Those operators are *not* densely defined in the Hilbert space $\mathfrak{H}$. However, according to (C.8) below, they *are* densely defined in the subspaces $\mathfrak{H}_{j,l}$ with $j \ge 3/2$. Another new aspect is that we trace the dependence on the coupling constant.

Theorem 3.2 (Spectral shift: Chandrasekhar case). *There exists a constant $C < \infty$ such that for all $l \in \mathbb{N}_0$, $j = l \pm \frac{1}{2}$, for all κ satisfying*

$$\kappa \le \begin{cases} \kappa^C & \text{if } l = 0, \\ \kappa^B & \text{if } l \ge 1, \end{cases}$$

for all $\mu \geq 0$ *and for all* $v : [0,\infty) \to [0,\infty)$ *satisfying* (3.2)*, one has*

$$0 \leq \mathrm{tr}_{j,l}\left([C[v]+\mu]_- - [S[v]+\mu]_-\right) \leq C\frac{\kappa^4}{(l+\frac{1}{2})^2}. \tag{3.4}$$

One of the key points to be appreciated in the above theorems is an effective cancellation in the differences in (3.4) and (3.3). This can already be seen for Coulomb potentials $v(r) = \kappa r^{-1}$, where

$$\mathrm{tr}_{j,l}\,[S_\kappa]_- = (2j+1)\frac{\kappa^2}{2}\sum_{n=1}^{\infty}\frac{1}{(n+l)^2},$$

which does not decay at all as $j \to \infty$. Moreover, for fixed j and l the above trace vanishes only like κ^2 as $\kappa \to 0$. It is rather remarkable that such cancellations occur uniformly for all attractive potential v satisfying (3.2).
The following proof of Theorem 3.2 follows the ideas of [31, Thm. 2]. It is not only included to render the paper self-contained, but also to establish the above mentioned improvement, which is important for the present paper.

Proof of Theorem 3.2. We note that both traces $\mathrm{tr}_{j,l}\,[C[v]+\mu]_-$ and $\mathrm{tr}_{j,l}\,[S[v]+\mu]_-$ are finite. This follows by the variational principle from the case $v(r) = \kappa r^{-1}$, cf. Theorem 2.2 in the Chandrasekhar case. Thus, for $l < 3$ say, it is enough to show the claim for κ in a neighborhood of 0. More precisely, we can assume $\kappa \leq \frac{1}{\sqrt{8}}(l+\frac{1}{2})$ which covers all $\kappa \leq \kappa^B$ for $l \geq 3$.
Moreover, by an approximation argument it is sufficient to consider $\mu > 0$ and bounded potentials v, cf. [31].
We denote by $\gamma_{j,l}$ the orthogonal projection onto the eigenspace of $C[v]$ corresponding to angular momenta j, l and eigenvalues less or equal than $-\mu$. The identity

$$\tfrac{1}{2}p^2 = C_0 + \tfrac{1}{2}C_0^2 \tag{3.5}$$

and the variational principle (cf. [43, Thm. 12.1]) imply

$$0 \leq 2\,\mathrm{tr}_{j,l}\left([C[v]+\mu]_- - [S[v]+\mu]_-\right) \leq \mathrm{tr}\left[C_0^2\gamma_{j,l}\right]. \tag{3.6}$$

Using the eigenvalue equation and the bound (3.2) on the potential we estimate this term further as follows.

$$\mathrm{tr}\left[C_0^2\gamma_{j,l}\right] \leq \mathrm{tr}_{j,l}\,[C[v]]_-^2 + \mathrm{tr}\left[v^2\gamma_{j,l}\right] \leq \mathrm{tr}_{j,l}\,[C_\kappa]_-^2 + \kappa^2\,\mathrm{tr}\left[|\mathbf{x}|^{-2}\gamma_{j,l}\right].$$

Using Hardy's inequality and (3.5)

$$\begin{aligned}\mathrm{tr}\left[|\mathbf{x}|^{-2}\gamma_{j,l}\right] &\leq \left(l+\tfrac{1}{2}\right)^{-2}\mathrm{tr}\left[\mathbf{p}^2\gamma_{j,l}\right]\\ &= \left(l+\tfrac{1}{2}\right)^{-2}\left(\mathrm{tr}\left[C_0^2\gamma_{j,l}\right] + 2\,\mathrm{tr}\left[C_0\gamma_{j,l}\right]\right).\end{aligned}$$

Since $\kappa < l + \frac{1}{2}$, the last two estimates may be summarized as

$$\mathrm{tr}\left[C_0^2\gamma_{j,l}\right] \leq \left(1 - \frac{\kappa^2}{(l+\frac{1}{2})^2}\right)^{-1}\left(\mathrm{tr}_{j,l}\,[C_\kappa]_-^2 + \frac{2\kappa^2}{(l+\frac{1}{2})^2}\,\mathrm{tr}\left[C_0\gamma_{j,l}\right]\right). \tag{3.7}$$

We shall estimate the two terms on the right hand side separately. From [31, Lemma 3] we recall the following angular momentum barrier inequality on $\mathfrak{H}_{j,l}$,

$$C_0 \geq 2\kappa r^{-1}\chi_{\{r\leq R_l(\kappa)\}}, \qquad R_l(\kappa) = \tfrac{1}{8\kappa}(l+\tfrac12)^2. \tag{3.8}$$

(Here we use that $\kappa \leq \frac{1}{\sqrt{8}}(l+\frac12)$.) This implies

$$\begin{aligned}\operatorname{tr}[C_0\gamma_{j,l}] &\leq \kappa \operatorname{tr}\left[|\mathbf{x}|^{-1}\gamma_{j,l}\right] \leq \frac12 \operatorname{tr}[C_0\gamma_{j,l}] + \frac14 \operatorname{tr}[w_l\gamma_{j,l}] \\ &= \frac34 \operatorname{tr}[C_0\gamma_{j,l}] - \frac14 \operatorname{tr}[C[w_l]\gamma_{j,l}]\end{aligned}$$

where $w_l(r) := 4\kappa r^{-1}\chi_{\{r\geq R_l(\kappa)\}}$. Hence, using the variational principle followed by Daubechies' inequality [9] (cf. also [31, Prop. 1])

$$\begin{aligned}\operatorname{tr}[C_0\gamma_{j,l}] &\leq \operatorname{tr}_{j,l}[C[w_l]]_- \\ &\leq \text{const}\,(2l+1)\left(\int_0^\infty w_l(r)^{3/2}\,dr + \int_0^\infty w_l(r)^2\,dr\right) \leq \text{const}\,\kappa^2.\end{aligned} \tag{3.9}$$

In order to estimate the first term on the right hand side of (3.7) we use (3.8) to obtain on $\mathfrak{H}_{j,l}$

$$C_\kappa \geq \tfrac12 C_0 - \kappa r^{-1}\chi_{\{r\geq R_l(\kappa)\}} \geq \tfrac12 C[w_l].$$

with w_l as above. Hence again by Daubechies' inequality

$$\begin{aligned}\operatorname{tr}_{j,l}[C_\kappa]_-^2 &\leq \text{const}\,(2l+1)\left(\int_0^\infty w_l(r)^{5/2}\,dr + \int_0^\infty w_l(r)^3\,dr\right) \\ &\leq \text{const}\,\kappa^4(l+\tfrac12)^{-2}.\end{aligned}$$

Combing this with (3.9), (3.7), and (3.6) completes the proof. ☐

Having finished the proof of Theorem 3.2 it is easy to give the

Proof of Theorem 3.1. Since the trace $\operatorname{tr}_j[B[v]+\mu]_-$ is finite according to Theorem 2.1 we may assume that either $\kappa \leq \kappa^C$ and $j = 1/2$, or else that $j \geq 3/2$. In this case, the claim essential boils down to Theorem 3.2. To see this, we note the identity

$$B[v] = \mathcal{U}(C[v]) = \tfrac12\left(U(\mathbf{p})^* C[v] U(\mathbf{p}) + U(\mathbf{p}) C[v] U(\mathbf{p})^*\right) \tag{3.10}$$

involving the unitary operator $U(\mathbf{p}) := \Phi_0(\mathbf{p}) + i\Phi_1(\mathbf{p})$ (see also (2.1)). Equality (3.10) as well as the unitarity of $U(\mathbf{p})$ are easily derived from the fact that $\Phi_0^2(\mathbf{p}) + \Phi_1^2(\mathbf{p}) = 1$.
Even if v satisfies (3.2) only with a $\kappa^C < \kappa \leq \kappa^B$, identity (3.10) remains valid on all subspaces $\mathfrak{H}_j$ with $j \geq 3/2$. Hence by the concavity of the sum of

negative eigenvalues [72] of $B[v]+\mu$ one has for any $\mu \geq 0$

$$\begin{aligned} & \operatorname{tr}_j [B[v]+\mu]_- \\ \leq & \frac{1}{2}\operatorname{tr}_j [U^*(\mathbf{p})C[v]U(\mathbf{p})+\mu]_- + \frac{1}{2}\operatorname{tr}_j [U(\mathbf{p})C[v]U^*(\mathbf{p})+\mu]_- \\ = & \operatorname{tr}_j [C[v]+\mu]_- . \end{aligned} \tag{3.11}$$

By (3.4) the trace in (3.3) is thus bounded from above by

$$\operatorname{tr}_j \left([C[v]+\mu]_- - [S[v]+\mu]_-\right) \leq \text{const}\, \kappa^4 j^{-2},$$

as claimed. ☐

3.2 Properties of the spectral shift

In this subsection we discuss some properties of the spectral shift $s(\kappa)$ defined in (1.8).

Lemma 3.3 (Properties of the spectral shift). *The spectral shift s is a continuous, non-negative function on $(0, \kappa^B]$ satisfying $s(\kappa) = \mathcal{O}(\kappa^2)$ as $\kappa \downarrow 0$.*

Proof. According to (2.6) and Theorem 3.1 one has

$$0 \leq s_j(\kappa) := \kappa^{-2} \operatorname{tr}_j \left([B_\kappa]_- - [S_\kappa]_-\right) \leq \text{const}\, \kappa^2 j^{-2}.$$

Therefore the sum $s(\kappa) = \sum_j s_j(\kappa)$ converges, is non-negative and satisfies the claimed asymptotic estimate as $\kappa \downarrow 0$. By the min-max principle each eigenvalue depends continuously on κ. Thus the continuity of their sum follows from the estimates in Theorem 2.1 and the Weierstraß criterion for uniform convergence. ☐

4 Proof of the Scott correction

The strategy of the proof of the main results is similar to the one used for the Chandrasekhar operator [31]. We employ the Schrödinger operator as a regularization for the relativistic problem, i.e., we will use it to eliminate the main contribution to the energy (the Thomas-Fermi energy) and focus only on the energy shift of the low lying states. For these the electron-electron interaction plays no role and the unscreened problem remains. We define

$$E^S(Z) := \inf\{\mathcal{E}^S_Z(\psi) \,|\, \psi \in \mathfrak{Q}^S_Z, \|\psi\| = 1\}$$

to be the ground state energy in the Schrödinger case,

$$\mathcal{E}^S_N(\psi) := \left\langle \psi, \left[\sum_{\nu=1}^N \left(\frac{1}{2}\mathbf{p}_\nu^2 - Z|\mathbf{x}_\nu|^{-1}\right) + \sum_{1\leq\mu<\nu\leq N} |\mathbf{x}_\mu - \mathbf{x}_\nu|^{-1}\right] \psi \right\rangle.$$

It is defined on $\mathfrak{Q}_N^S := \mathfrak{H}_N^S \cap \mathfrak{S}(\mathbb{R}^{3N}, \mathbb{C}^{2^N})$, where $\mathfrak{H}_N^S := \bigwedge_{\nu=1}^N \mathfrak{H}$ is the Hilbert space of anti-symmetric two-spinors. We recall that we suppose neutrality, i.e., $N = Z$.

The asymptotics of the Schrödinger ground-state energy up to Scott correction reads [55]

$$E^S(Z) = E_{\mathrm{TF}}(Z) + \tfrac{1}{2} Z^2 + \mathcal{O}(Z^{47/24}). \tag{4.1}$$

For our purpose this remainder estimate is sufficient. However, even the coefficient of the $Z^{5/3}$-term in the asymptotic expansion is known [23, 24, 25, 18, 26, 21, 19, 20, 22].

Our main result, Theorem 1.1, will follow from (4.1) if we can show that in the limit $Z \to \infty$ the difference of the Schrödinger and Brown-Ravenhall ground-state energy satisfies

$$E^S(Z) - E_c^B(Z) = s(Z/c)\, Z^2 + \mathcal{O}(Z^{47/24}) \tag{4.2}$$

uniformly in $\kappa = Z/c \in (0, \kappa^B]$. We break the proof of this assertion into an upper and lower bound.

4.1 Upper bound on the energy difference

The Thomas-Fermi functional (1.7) has a unique minimizer ϱ_Z, the Thomas-Fermi density (Lieb and Simon [44]). It scales as $\varrho_Z(\mathbf{x}) := Z^2 \varrho_1(Z^{1/3}\mathbf{x})$. We set

$$\phi_{\mathrm{TF}}(\mathbf{x}) := Z|\mathbf{x}|^{-1} - \int_{\mathbb{R}^3} \frac{\varrho_Z(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\, \mathrm{d}\mathbf{y}, \tag{4.3}$$

the Thomas-Fermi potential, and

$$L_{\mathrm{TF}}(\mathbf{x}) := \int_{|\mathbf{x}-\mathbf{y}|<R_Z(\mathbf{x})} \frac{\varrho_Z(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\, \mathrm{d}\mathbf{y},$$

the exchange hole potential. Here $R_Z(\mathbf{x})$ is defined as the (unique) minimal radius for which $\int_{|\mathbf{x}-\mathbf{y}|\le R_Z(\mathbf{x})} \varrho_Z(\mathbf{y})\mathrm{d}\mathbf{y} = \frac{1}{2}$. The corresponding one-particle operators – self-adjointly realized in $\mathfrak{H}$ – are

$$S_{\mathrm{TF}} = S[\phi_{\mathrm{TF}} + L_{\mathrm{TF}}], \qquad B_{\mathrm{TF}} = B_c[\phi_{\mathrm{TF}} + L_{\mathrm{TF}}].$$

Here we use a notation analogous to that in (1.5).

We shall express the many-particle ground-state energies $E^S(Z)$ and $E_c^B(Z)$ in terms of quantities involving the above one-particle operators. In the Schrödinger case, this was achieved in [55, 58] in terms of the Thomas-Fermi potential ϕ_{TF}. Our point in the proof of the following proposition is to replace ϕ_{TF} by the exchange hole reduced potential $\phi_{\mathrm{TF}} + L_{\mathrm{TF}}$.

Proposition 4.1. *Let* $J := \left[Z^{1/9}\right] + \frac{1}{2}$. *Then, as* $Z \to \infty$,

$$\begin{aligned} &E^S(Z) \\ &= -\sum_{j=1/2}^{J-1} \mathrm{tr}_j \left[S[Z|\mathbf{x}|^{-1}]\right]_- - \sum_{j=J}^{Z+1/2} \mathrm{tr}_j \left[S_{\mathrm{TF}}\right]_- - D(\varrho_Z, \varrho_Z) + O(Z^{47/24}). \end{aligned} \tag{4.4}$$

Since $\phi_{\mathrm{TF}} + L_{\mathrm{TF}}$ has a Coulomb tail, the trace $\mathrm{tr}_j\,[S_{\mathrm{TF}}]_-$ is finite for each j, but not summable with respect to j. It is therefore essential to restrict the second sum to a finite number of angular momenta. However, the value of the cut-off, $j \le Z + 1/2$, is not chosen optimally here, since for our argument it is largely arbitrary.

Proof of Proposition 4.1. According to the correlation inequality [47]

$$E^S(Z) \ge -\sum_{j=1/2}^{Z+1/2} \mathrm{tr}_j \left[S_{\mathrm{TF}}\right]_- - D(\varrho_Z, \varrho_Z).$$

Note that the Z electrons can certainly be accommodated in the first Z angular momentum channels (which is a very crude bound). Estimating $\phi_{\mathrm{TF}} + L_{\mathrm{TF}}$ from above by the Coulomb potential for small angular momenta, we obtain

$$E^S(Z) \ge -\sum_{j=1/2}^{J-1} \mathrm{tr}_j \left[S[Z|\mathbf{x}|^{-1}]\right]_- - \sum_{j=J}^{Z+1/2} \mathrm{tr}_j \left[S_{\mathrm{TF}}\right]_- - D(\varrho_Z, \varrho_Z). \tag{4.5}$$

Moreover, see [55, 58],

$$\begin{aligned} &E^S(Z) \\ &\le -\sum_{j=1/2}^{J-1} \mathrm{tr}_j \left[S[Z|\mathbf{x}|^{-1}]\right]_- - \sum_{j=J}^{\infty} \mathrm{tr}_j \left[S[\phi_{\mathrm{TF}}]\right]_- - D(\varrho_Z, \varrho_Z) + \mathrm{const}\, Z^{47/24}. \end{aligned}$$

Hence it suffices to prove that

$$-\sum_{j=J}^{Z+1/2} \mathrm{tr}_j \left[S_{\mathrm{TF}}\right]_- \ge -\sum_{j=J}^{\infty} \mathrm{tr}_j \left[S[\phi_{\mathrm{TF}}]\right]_- - \mathrm{const}\, Z^{5/3} \tag{4.6}$$

(Note that the lower bound in [31] contains an error by estimating [31, Equation (43)] too generously. Really, only the first Z lowest negative eigenvalues need to occur on the right hand side instead of all. In particular, there will be never more than Z total angular momentum channels occupied. This fact is taken into account here yielding a suitable lower bound. The problem in [31] can be circumvented in exactly the same way.) We decompose $L_{\mathrm{TF}} = L_< + L_>$ where

$$L_< = \chi_{\{|\mathbf{x}|<R\}} L_{\mathrm{TF}}, \qquad L_> = \chi_{\{|\mathbf{x}|\ge R\}} L_{\mathrm{TF}},$$

with a constant R (independent of Z) to be chosen below. For $\varepsilon > 0$ to be specified later we estimate using the variational principle for sums of eigenvalues

$$\begin{aligned}&\operatorname{tr}_j [S_{\mathrm{TF}}]_- \\ &\le \operatorname{tr}_j(\tfrac{1}{2}(1-2\varepsilon^2)p^2 - \phi_{\mathrm{TF}})_- + \varepsilon^2 \operatorname{tr}_j(\tfrac{1}{2}p^2 - \varepsilon^{-2}L_<)_- + \varepsilon^2 \operatorname{tr}_j(\tfrac{1}{2}p^2 - \varepsilon^{-2}L_>)_- .\end{aligned} \tag{4.7}$$

By the subsequent lemma the first (and main) term is bounded according to

$$\begin{aligned}&\sum_{j=J}^{Z+1/2} \operatorname{tr}_j(\tfrac{1}{2}(1-2\varepsilon^2)p^2 - \phi_{\mathrm{TF}})_- - \sum_{j=J}^{\infty} \operatorname{tr}_j(\tfrac{1}{2}p^2 - \phi_{\mathrm{TF}})_- \\ &\le \operatorname{tr}(\tfrac{1}{2}(1-2\varepsilon^2)p^2 - \phi_{\mathrm{TF}})_- - \operatorname{tr}(\tfrac{1}{2}p^2 - \phi_{\mathrm{TF}})_- \le \mathrm{const}\, \varepsilon^2 Z^{7/3}.\end{aligned}$$

For the second term on the right side of (4.7) we use the Lieb-Thirring inequality [45] to obtain

$$\begin{aligned}&\varepsilon^2 \sum_{j=J}^{Z+1/2} \operatorname{tr}_j(\tfrac{1}{2}p^2 - \varepsilon^{-2}L_<)_- \le \varepsilon^2 \operatorname{tr}(\tfrac{1}{2}p^2 - \varepsilon^{-2}L_<)_- \\ &\le \mathrm{const}\, \varepsilon^{-3} \int L_<(\mathbf{x})^{5/2}\, d\mathbf{x} \le \mathrm{const}\, \varepsilon^{-3} Z^{2/3}.\end{aligned}$$

In the last inequality we used a bound of Siedentop and Weikard [58, Proof of Lemma 2]. It is at this point that R is chosen. The penultimate inequality in [58, Proof of Lemma 2] asserts after scaling that $L_>(\mathbf{x}) \le \mathrm{const}\, |\mathbf{x}|^{-1}$. Hence by comparison with the exact hydrogen solution

$$\begin{aligned}&\varepsilon^2 \sum_{j=J}^{Z+1/2} \operatorname{tr}_j(\tfrac{1}{2}p^2 - \varepsilon^{-2}L_>)_- \le \varepsilon^2 \sum_{j=1/2}^{Z+1/2} \operatorname{tr}_j(\tfrac{1}{2}p^2 - \varepsilon^{-2}\mathrm{const}\, |\mathbf{x}|^{-1})_- \\ &= \mathrm{const}\, \varepsilon^{-2} \sum_{j=1/2}^{Z+1/2} \sum_{n=1}^{\infty} \frac{2j+1}{(n+j-1/2)^2} \le \mathrm{const}\, \varepsilon^{-2} Z.\end{aligned}$$

Choosing $\varepsilon = Z^{-1/3}$ all the error terms are $\mathcal{O}(Z^{5/3})$, proving (4.6). □

In the previous proof we used

LEMMA 4.2. *For all* $0 < \varepsilon \le 1/2$, *as* $Z \to \infty$,

$$\operatorname{tr}(\tfrac{1}{2}(1-\varepsilon^2)p^2 - \phi_{\mathrm{TF}})_- \le \operatorname{tr}(\tfrac{1}{2}p^2 - \phi_{\mathrm{TF}})_- + \mathrm{const}\, \varepsilon^2 Z^{7/3}. \tag{4.8}$$

Note that there are only a finite number of eigenvalues, since ϕ_{TF} decays like $|\mathbf{x}|^{-4}$.

Proof. Let $d^{\varepsilon}_{\mathrm{TF}}$ be the projection onto the negative eigenvalues of $\frac{1}{2}(1-\epsilon^2)p^2 - \phi_{\mathrm{TF}}$. Then, by the variational principle

$$\begin{aligned}&\operatorname{tr}(\tfrac{1}{2}(1-\varepsilon^2)p^2-\phi_{\mathrm{TF}})_- - \operatorname{tr}(\tfrac{1}{2}p^2-\phi_{\mathrm{TF}})_-\\ &\quad\le -\operatorname{tr} d^{\varepsilon}_{\mathrm{TF}}(\tfrac{1}{2}(1-\varepsilon^2)p^2-\phi_{\mathrm{TF}}) + \operatorname{tr} d^{\varepsilon}_{\mathrm{TF}}(\tfrac{1}{2}p^2-\phi_{\mathrm{TF}}) = \tfrac{\varepsilon^2}{2}\operatorname{tr} d^{\varepsilon}_{\mathrm{TF}}p^2. \qquad (4.9)\end{aligned}$$

Hence the claim will follow, if we show that $\operatorname{tr} d^{\varepsilon}_{\mathrm{TF}}p^2 \le \text{const } Z^{7/3}$. Note that $d^{\varepsilon}_{\mathrm{TF}}$ depends on both ε and Z, and by rescaling one may get rid of the ε dependence at the expense of changing Z. We may therefore assume that $\varepsilon = 0$ and write $d_{\mathrm{TF}} = d^0_{\mathrm{TF}}$.
Thus, it remains to prove

$$\operatorname{tr} d_{\mathrm{TF}}p^2 \le \text{const } Z^{7/3}. \qquad (4.10)$$

Note that this says that the *kinetic* energy is bounded by the order of the *total* energy $\operatorname{tr} d_{\mathrm{TF}}(\frac{1}{2}p^2 - \phi_{\mathrm{TF}})$, which is well-known to be of order $Z^{7/3}$. Using that ϕ_{TF} is bounded by a constant times $\min\{Z|\mathbf{x}|^{-1}, |\mathbf{x}|^{-4}\}$ (see [44]) we get for any $R > 0$

$$\begin{aligned}&\tfrac{1}{2}\operatorname{tr} d_{\mathrm{TF}}p^2 \le \operatorname{tr} d_{\mathrm{TF}}\phi_{\mathrm{TF}}\\ &\quad\le \text{const}\left(\left(\int_{\{|\mathbf{x}|<R\}}(Z|\mathbf{x}|^{-1})^{5/2}\,\mathrm{d}\mathbf{x}\right)^{2/5}\left(\int d_{\mathrm{TF}}(\mathbf{x},\mathbf{x})^{5/3}\,\mathrm{d}\mathbf{x}\right)^{3/5}\right.\\ &\qquad\qquad\left. + R^{-4}\int d_{\mathrm{TF}}(\mathbf{x},\mathbf{x})\,\mathrm{d}\mathbf{x}\right).\end{aligned}$$

The Cwikel-Lieb-Rozenblum inequality (for a textbook presentation, see, e.g., [63]) guarantees that

$$\int d_{\mathrm{TF}}(\mathbf{x},\mathbf{x})\,\mathrm{d}\mathbf{x} \le \text{const } \int \phi_{\mathrm{TF}}(\mathbf{x})^{3/2}\mathrm{d}\mathbf{x} = \text{const } Z.$$

Moreover, by the Lieb-Thirring inequality [45]

$$\int d_{\mathrm{TF}}(\mathbf{x},\mathbf{x})^{5/3}\,\mathrm{d}\mathbf{x} \le \text{const } \operatorname{tr} d_{\mathrm{TF}}p^2.$$

We can estimate for any $\delta > 0$

$$\begin{aligned}&\left(\int_{\{|\mathbf{x}|<R\}}(Z|\mathbf{x}|^{-1})^{5/2}\,\mathrm{d}\mathbf{x}\right)^{2/5}\left(\int d_{\mathrm{TF}}(\mathbf{x},\mathbf{x})^{5/3}\,\mathrm{d}\mathbf{x}\right)^{3/5}\\ &\qquad\le \text{const } ZR^{1/5}\left(\operatorname{tr} d_{\mathrm{TF}}p^2\right)^{3/5}\\ &\qquad\le \delta\operatorname{tr} d_{\mathrm{TF}}p^2 + \text{const } \delta^{-3/2}Z^{5/2}R^{1/2}.\end{aligned}$$

In summary, we have shown that

$$\left(\tfrac{1}{2} - \text{const}\,\delta\right) \operatorname{tr} d_{\mathrm{TF}} p^2 \le \text{const}\,\left(\delta^{-3/2} Z^{5/2} R^{1/2} + R^{-4} Z\right).$$

Choosing δ small (of order one) and $R = Z^{-1/3}$ we obtain (4.10). □

Next, we bound the many-particle ground state energy of the Brown-Ravenhall operator from below by one-body quantities which match the corresponding quantities in the Schrödinger case (4.4).

Lemma 4.3. *For all $J \in \mathbb{N}_0 + 1/2$ and $Z \in \mathbb{N}$*

$$E_c^B(Z) \ge -\sum_{j=1/2}^{J-1} \operatorname{tr}_j \left[B_c[Z|\mathbf{x}|^{-1}]\right]_- - \sum_{j=J}^{Z+1/2} \operatorname{tr}_j \left[B_{\mathrm{TF}}\right]_- - D(\varrho_Z, \varrho_Z).$$

Proof. This follows by the same argument leading to (4.5). □

We are now ready to give a

Proof of Theorem 1.1 – first part. Choosing $J = \left[Z^{1/9}\right] + \frac{1}{2}$ and combining Proposition 4.1 and Lemma 4.3 we obtain

$$\begin{aligned} E^S(Z) - E_c^B(Z) \le &\sum_{j=1/2}^{J-1} \operatorname{tr}_j \left(\left[B_c[Z|\mathbf{x}|^{-1}]\right]_- - \left[S[Z|\mathbf{x}|^{-1}]\right]_-\right) \\ &+ \sum_{j=J}^{Z+1/2} \operatorname{tr}_j \left([B_{\mathrm{TF}}]_- - [S_{\mathrm{TF}}]_-\right) + \mathcal{O}(Z^{47/24}). \end{aligned} \tag{4.11}$$

We note that by scaling $\mathbf{x} \mapsto \mathbf{x}/c$, the operators $S[Z|\mathbf{x}|^{-1}]$ and $B_c[Z|\mathbf{x}|^{-1}]$ are unitarily equivalent to the operators $Z^2\kappa^{-2}S_\kappa$ and $Z^2\kappa^{-2}B_\kappa$ where $\kappa = Z/c$. Similarly, S_{TF} and B_{TF} are unitarily equivalent to the operators $Z^2\kappa^{-2}S[\kappa|\mathbf{x}|^{-1} - \chi_c]$ and $Z^2\kappa^{-2}B[\kappa|\mathbf{x}|^{-1} - \chi_c]$ acting in $\mathfrak{H}$, where

$$\chi_c(\mathbf{x}) := c^{-4} \int_{|\mathbf{x}-\mathbf{y}|>cR_Z(c^{-1}\mathbf{x})} \frac{\varrho_Z(c^{-1}\mathbf{y})}{|\mathbf{x}-\mathbf{y}|} \,\mathrm{d}\mathbf{y}.$$

This implies that the first two terms on the right-hand side of (4.11), which we denote by $\Sigma_1(Z,c)$ and $\Sigma_2(Z,c)$, can be rewritten as

$$\begin{aligned} \Sigma_1(Z,c) &= Z^2\kappa^{-2} \sum_{j=1/2}^{J-1} \operatorname{tr}_j \left([B_\kappa]_- - [S_\kappa]_-\right), \\ \Sigma_2(Z,c) &= Z^2\kappa^{-2} \sum_{j=J}^{Z+1/2} \operatorname{tr}_j \left(\left[B[\kappa|\mathbf{x}|^{-1} - \chi_c]\right]_- - \left[S[\kappa|\mathbf{x}|^{-1} - \chi_c]\right]_-\right). \end{aligned}$$

Inequality (2.6) and Theorem 3.1 guarantee that the terms in the first sum are non-negative and that the terms in both sums are bounded from above by a constant times $\kappa^4 j^{-2}$ independently of Z and c. Therefore, the first sum can be bounded from above by an absolutely convergent series,

$$\Sigma_1(Z,c) \leq Z^2\kappa^{-2} \sum_{j=1/2}^{\infty} \operatorname{tr}_j \left([B_\kappa]_- - [S_\kappa]_-\right) = Z^2\, s(\kappa).$$

By the same token

$$\Sigma_2(Z,c) \leq \text{const } Z^2\kappa^2 \sum_{j=J}^{\infty} j^{-2} = \mathcal{O}(Z^{17/9}),$$

uniformly in c. This concludes the proof of the upper bound on the energy difference. □

4.2 Lower bound on the energy difference

Similarly to [55] we define one-particle density matrices d^S and d^B on $\mathfrak{H}$ as sums

$$d^{\#} = d^{\#}_{<} + d_{>}, \quad \# = S, B. \tag{4.12}$$

The contribution of small total angular momenta, $d^{\#}_{<} = \sum_{l<L} d^{\#}_l$, is defined in Appendix E.1. It comes from the eigenspinors of the atomic problems. The contribution of large angular momentum, $d_> = \sum_{l=L}^{\infty} d_l$, is defined in Appendix E.2. It corresponds to the Macke orbitals of [55] and, in particular, coincides for the Schrödinger and Brown-Ravenhall case. The angular-momentum cut-off L will be chosen in a Z-dependent way, namely,

$$L := [Z^{1/12}].$$

Important properties of the density matrices, whose construction is explained in more detail in Appendix E, are:

- The densities

$$\rho^{\#}(\mathbf{x}) := \operatorname{tr}_{\mathbb{C}^2}\left(d^{\#}(\mathbf{x},\mathbf{x})\right), \quad \rho^{\#}_l(\mathbf{x}) := \operatorname{tr}_{\mathbb{C}^2}\left(d^{\#}_l(\mathbf{x},\mathbf{x})\right),$$
$$\rho^{\#}_{<}(\mathbf{x}) := \sum_{l<L} \rho^{\#}_l(\mathbf{x}), \quad \rho_>(\mathbf{x}) := \sum_{l\geq L} \rho_l(\mathbf{x}).$$

  of $d^{\#}$, $d^{\#}_l$, and $d_>$ are all spherically symmetric.

- The dimension of the ranges of the density matrices d^S and d^B is at most Z, in particular $\operatorname{tr} d^{\#} \leq Z$. Moreover,

$$\operatorname{tr} d^{\#}_l = \int \rho^{\#}_l(\mathbf{x})\,\mathrm{d}\mathbf{x} = 2(2l+1)(K-l), \qquad 0 \leq l < L, \tag{4.13}$$

  with $K = [\text{const } Z^{1/3}]$ and a suitable constant.

For a lower bound on the ground state energy in the Schrödinger case, we recall from [55] and [31, Proposition 4] the following

Proposition 4.4. *For large Z,*

$$E^S(Z) = \operatorname{tr}\left[S[Z|\mathbf{x}|^{-1}]\, d^S\right] + D(\rho^S, \rho^S) + \mathcal{O}(Z^{47/24}).$$

To obtain an upper bound on the ground state energy in the Brown-Ravenhall case, we use the reduced Hartree-Fock variational principle. It involves the density

$$\rho_U^B(\mathbf{x}) := \operatorname{tr}_{\mathbb{C}^2}\left(\mathcal{U}_c(d^B)(\mathbf{x},\mathbf{x})\right)$$

of the twisted density matrix $\mathcal{U}_c(d^B)$.
For further reference, we also set

$$\rho_{U,l}^B(\mathbf{x}) := \operatorname{tr}_{\mathbb{C}^2}\left(\mathcal{U}_c(d_l^B)(\mathbf{x},\mathbf{x})\right), \quad \rho_{U,<}^B(\mathbf{x}) := \sum_{l<L} \rho_{U,l}^B(\mathbf{x}),$$
$$\rho_{U,>}(\mathbf{x}) := \sum_{l\geq L} \rho_{U,l}(\mathbf{x}).$$

Applying to (1.1) the Hartree-Fock variational principle – in the strengthened version of Lieb [42] (see also Bach [2]) – and omitting the manifestly negative exchange energy we arrive at

Proposition 4.5. *For all Z and c,*

$$E_c^B(Z) \leq \operatorname{tr}[B_c[Z|\mathbf{x}|^{-1}]\, d^B] + D(\rho_U^B, \rho_U^B).$$

Combining Propositions 4.4 and 4.5 we find

$$\begin{aligned} &E_c^B(Z) - E^S(Z) \\ &\leq \operatorname{tr}[B_c[Z|\mathbf{x}|^{-1}]d^B] - \operatorname{tr}[S[Z|\mathbf{x}|^{-1}]d^S] + D(\rho_U^B - \rho^S, \rho_U^B + \rho^S) + \text{const}\; Z^{47/24}. \end{aligned}$$

Now we use the inequality $\mathbf{p}^2 \geq 2c^2(E(p/c)-1)$ for the kinetic energy corresponding to $d_>$. Moreover, we write

$$\begin{aligned} D(\rho_U^B - \rho^S, \rho_U^B + \rho^S) = D(\rho_{U,>} - \rho_>, \rho_{U,>} + \rho_>) + 2\, D(\rho_{U,<}^B, \rho_U^B + \rho^S) \\ - D(\rho_{U,<}^B + \rho_<^S, \rho_{U,<}^B + \rho_<^S) - 2\, D(\rho_{U,<}^B + \rho_<^S, \rho_>) \end{aligned}$$

and drop the two negative terms on the right hand side. We arrive at

$$\begin{aligned} E_c^B(Z) - E^S(Z) \leq \operatorname{tr}\left[B_c[\tfrac{Z}{|\mathbf{x}|}]d_<^B\right] - \operatorname{tr}\left[S[\tfrac{Z}{|\mathbf{x}|}]d_<^S\right] + \underbrace{\operatorname{tr}\left[\left(\tfrac{Z}{|\mathbf{x}|} - \mathcal{U}_c(\tfrac{Z}{|\mathbf{x}|})\right) d_>\right]}_{=:\mathcal{R}_1} \\ + \underbrace{D(\rho_{U,>} - \rho_>, \rho_{U,>} + \rho_>)}_{=:\mathcal{R}_2} + 2\underbrace{D(\rho_{U,<}^B, \rho_U^B + \rho^S)}_{=:\mathcal{R}_3} + \text{const}\; Z^{47/24}. \end{aligned} \tag{4.14}$$

As we shall see, the first two terms will yield the Scott correction. In the following subsections we prove that $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_3$ are relatively small remainder terms. Hence, we wish to control the effects of the twisting operation $\mathcal{U}_c$, which stems from the electronic projection, on the electrostatic Coulomb energy.

4.2.1 Controlling the electron projection for high angular momenta

Our task in this subsection is to prove that for large angular momenta, the twisted and untwisted electrostatic energy are asymptotically equal.
We start by comparing the electric potential energy with or without electron projection for large angular momentum. This will imply that the term $\mathcal{R}_1$ in (4.14) is relatively small.

Lemma 4.6. *In the limit $Z \to \infty$ one has uniformly in $\kappa = Z/c \in (0, \kappa^B]$*

$$\int (\rho_>(\mathbf{x}) - \rho_{U,>}(\mathbf{x})) \frac{\mathrm{d}\mathbf{x}}{|\mathbf{x}|} = \mathrm{tr}\left[\left(|\mathbf{x}|^{-1} - \mathcal{U}_c(|\mathbf{x}|^{-1})\right) d_>\right] = \mathcal{O}(Z^{11/12}).$$

Proof. Let $\{\psi_\alpha\}$ stand for the Macke orbitals building up $d_>$ which we label by $\alpha = (j, l, m, n)$; see (E.1) and preceding equations in Appendix E.2. By the scaling $\mathbf{x} \mapsto \mathbf{x}/c$ one has the relation

$$\left\langle \psi_\alpha, \left[|\mathbf{x}|^{-1} - \mathcal{U}_c(|\mathbf{x}|^{-1})\right] \psi_\alpha \right\rangle = c \left\langle \psi_\alpha^{(c)}, \left[|\mathbf{x}|^{-1} - \mathcal{U}_1(|\mathbf{x}|^{-1})\right] \psi_\alpha^{(c)} \right\rangle$$

where $\psi_\alpha^{(c)}(\mathbf{x}) := c^{-3/2}\psi_\alpha(\mathbf{x}/c)$. Assuming that α corresponds to a fixed (large) (j, l) we may use Lemma 2.8 to estimate the right-hand side by a constant times

$$\frac{c}{l^2} \left\langle \psi_\alpha^{(c)}, \mathbf{p}^2 \psi_\alpha^{(c)} \right\rangle = \frac{1}{l^2\, c} \left\langle \psi_\alpha, \mathbf{p}^2 \psi_\alpha \right\rangle.$$

Using that $Z/c \leq \kappa^B$ we obtain the estimate

$$\mathrm{tr}\left[\left(|\mathbf{x}|^{-1} - \mathcal{U}_c(|\mathbf{x}|^{-1})\right) d_>\right] \leq \mathrm{const}\, \frac{\kappa^B}{Z} \sum_{l=L}^{\infty} \frac{1}{l^2} \sum_{j=l\pm 1/2} \mathrm{tr}_{j,l}\left[\mathbf{p}^2 d_>\right].$$

The proof is completed using Lemma E.1 from Appendix E.3. $\square$

Next, we estimate the difference of Coulomb energies corresponding to large total angular momenta. This shows that the term $\mathcal{R}_2$ in (4.14) may be neglected.

Lemma 4.7. *In the limit $Z \to \infty$,one has uniformly in $\kappa = Z/c \in (0, \kappa^B]$*

$$\mathcal{R}_2 = D(\rho_{U,>} - \rho_>, \rho_> + \rho_{U,>}) = \mathcal{O}(Z^{5/3}).$$

Proof. We define $v := (\rho_> + \rho_{U,>}) * |\cdot|^{-1}$ to be the electric potential generated by $\rho_> + \rho_{U,>}$ which is obviously spherically symmetric and obeys

$$\begin{aligned} v(0) &= \mathrm{tr}\left[d_> \left(|\mathbf{x}|^{-1} + \mathcal{U}_c(|\mathbf{x}|^{-1})\right)\right] \\ &= 2\,\mathrm{tr}\left[d_>|\mathbf{x}|^{-1}\right] - \mathrm{tr}\left[d_> \left(|\mathbf{x}|^{-1} - \mathcal{U}_c(|\mathbf{x}|^{-1})\right)\right]. \end{aligned}$$

According to [55] (see also (E.8)) the first term on the right side is $\mathcal{O}(Z^{4/3})$. Moreover the second term is $\mathcal{O}(Z^{11/12})$ by Lemma 4.6, hence much smaller than the first term. Now,

$$D(\rho_> - \rho_{U,>}, \rho_> + \rho_{U,>}) = \tfrac{1}{2}\operatorname{tr}\left[d_> \left(v - \mathcal{U}_c(v)\right)\right]. \tag{4.15}$$

Decomposing the trace in (4.15) into the orbitals contributing to $d_>$ and scaling $\mathbf{x} \mapsto \mathbf{x}/c$ enables us to employ Lemma 2.9 to obtain the bound

$$\operatorname{tr}\left[d_> \left(v - \mathcal{U}_c(v)\right)\right] \leq \frac{\text{const}}{c^2} v(0)\ \operatorname{tr}\left[d_> \mathbf{p}^2\right].$$

This concludes the proof, since again from [55] (see (E.8)) we conclude that the trace on the right-hand side is $\mathcal{O}(Z^{7/3})$. □

4.2.2 Contribution from low angular momenta to the Coulomb energy

We now show that the term $\mathcal{R}_3$ in (4.14) is negligible.

Lemma 4.8. *In the limit $Z \to \infty$ one has uniformly in $\kappa = Z/c \in (0, \kappa^B]$*

$$\mathcal{R}_3 = D(\rho^B_{U,<}, \rho^B_U + \rho^S) = \mathcal{O}(Z^{11/6} \log Z).$$

Proof. We first treat the term $D(\rho^B_{U,<}, \rho_{U,>} + \rho_>)$. By construction the densities $\rho^B_{U,j}$ are spherically symmetric and satisfy according to (4.13)

$$\int \rho^B_{U,l}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int \rho^B_l(\mathbf{x})\,\mathrm{d}\mathbf{x} = 2(2l+1)(K-l), \quad 0 \leq l < L. \tag{4.16}$$

Recalling the choice of K and L we see that

$$\int \rho^B_{U,<}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathcal{O}(Z^{1/2}). \tag{4.17}$$

It follows from (E.8) and Lemma 4.6 that

$$\int \frac{\rho_{U,>}(\mathbf{x}) + \rho_>(\mathbf{x})}{|\mathbf{x}|}\mathrm{d}\mathbf{x} = \mathcal{O}(Z^{4/3}).$$

Hence Newton's theorem [49] yields

$$D(\rho^B_{U,<}, \rho_{U,>} + \rho_>) \leq \frac{1}{2}\int \rho^B_{U,<}(\mathbf{x})\mathrm{d}\mathbf{x} \int \frac{\rho_{U,>}(\mathbf{y}) + \rho_>(\mathbf{y})}{|\mathbf{y}|}\mathrm{d}\mathbf{y} = \mathcal{O}(Z^{11/6}).$$

In the remainder of the proof we are concerned with the term $D(\rho^B_{U,<}, \rho^B_{U,<} + \rho^S_<)$. Noting that

$$D(\rho^B_{U,<}, \rho^B_{U,<} + \rho^S_<) \leq \frac{3}{2} D(\rho^B_{U,<}, \rho^B_{U,<}) + \frac{1}{2} D(\rho^S_<, \rho^S_<).$$

and that according to [55, Prop. 3.5] $D(\rho^S_<, \rho^S_<) = \mathcal{O}(Z^{11/6})$, it suffices to consider $D(\rho^B_{U,<}, \rho^B_{U,<})$. We split the lowest angular momentum corresponding to $l \leq 2Z/c - 1/4 =: l_0$ off and define

$$d^B_\vdash := \sum_{l\leq l_0} d^B_l, \qquad d^B_\dashv := \sum_{l>l_0}^{L-1} d^B_l,$$

and

$$\rho^B_{U,\vdash} := \mathrm{tr}_{\mathbb{C}^2}\left(\mathcal{U}_c(d^B_\vdash)(\mathbf{x},\mathbf{x})\right), \qquad \rho^B_{U,\dashv} := \mathrm{tr}_{\mathbb{C}^2}\left(\mathcal{U}_c(d^B_\dashv)(\mathbf{x},\mathbf{x})\right).$$

Note that in case $l_0 < 0$ there is no need for this procedure. Accordingly, we estimate

$$D(\rho^B_{U,<}, \rho^B_{U,<}) \leq 2\, D(\rho^B_{U,\vdash}, \rho^B_{U,\vdash}) + 2\, D(\rho^B_{U,\dashv}, \rho^B_{U,\dashv}).$$

For an estimate of the second part corresponding to $l_0 < l < L$, we apply the following angular momentum barrier inequality

$$B_c[0] \geq \mathcal{U}_c\left(\frac{2Z}{|\mathbf{x}|}\,\chi_{\{|\mathbf{x}|\leq r_l\}}\right) \tag{4.18}$$

on $\mathfrak{H}_{j,l}$, where $r_l = \left((l+1/2)^2c^2 - 4Z^2\right)/(4Zc^2)$ and $l > 2Z/c$. This bound follows by applying $\mathcal{U}_1$ to the inequality in [31, Lemma 2.6] with $R_l = [(l+1/2)^2 - 4\kappa^2]/(4\kappa)$ and scaling $\mathbf{x} \mapsto \mathbf{x}/c$.
Inequality (4.18) implies

$$\begin{aligned}\mathrm{tr}\left[\mathcal{U}_c\left(|\mathbf{x}|^{-1}\right) d^B_l\right] &\leq \frac{1}{2Z}\,\mathrm{tr}\left[B_c[0]\, d^B_l\right] + \mathrm{tr}\left[\mathcal{U}_c\left(|\mathbf{x}|^{-1}\chi_{\{|\mathbf{x}|>r_l\}}\right) d^B_l\right]\\ &\leq \frac{1}{2}\,\mathrm{tr}\left[\mathcal{U}_c\left(|\mathbf{x}|^{-1}\right) d^B_l\right] + \frac{4Z}{(l+1/2)^2 - 4Z^2/c^2}\,\mathrm{tr}[d^B_l].\end{aligned}$$

Here the last inequality used the fact that eigenfunctions of d^B_l are eigenfunctions of $B_c[Z|\mathbf{x}|^{-1}]$ with negative eigenvalue. Now, note that

$$(l+1/2)^2 - 4Z^2/c^2 = (l+1/2+2Z/c)(l+1/2-2Z/c) \geq \mathrm{const}\,(l+1/2)^2$$

for $l \geq l_0$. Hence, using (4.16) and summing over l we obtain

$$\begin{aligned}\int \frac{\rho^B_{U,\dashv}(x)}{|\mathbf{x}|}\mathrm{d}\mathbf{x} &= \sum_{l>l_0}^{L-1} \mathrm{tr}\left[\mathcal{U}_c(|\mathbf{x}|^{-1})d^B_l\right]\\ &\leq \mathrm{const}\, Z\sum_{l=0}^{L-1}(l+1/2)^{-2}\int \rho^B_l(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathcal{O}(Z^{4/3}\log Z).\end{aligned}$$

Accordingly, Newton's theorem and (4.17) yield

$$D(\rho^B_{U,\dashv}, \rho^B_{U,\dashv}) \leq \frac{1}{2}\int \rho^B_{U,\dashv}(\mathbf{x})\,\mathrm{d}\mathbf{x}\int \frac{\rho^B_{U,\dashv}(\mathbf{x})}{|\mathbf{x}|}\,\mathrm{d}\mathbf{x} = \mathcal{O}(Z^{11/6}\log Z).$$

Finally, we consider the contribution from $l \leq l_0$. Note that then $l \leq 2\kappa^B - 1/4 < 2$. We claim that the electrostatic energy corresponding to the electrons in this subspace is bounded by

$$D\big(\rho^B_{U,\vdash}, \rho^B_{U,\vdash}\big) \leq \text{const } cK^2. \tag{4.19}$$

Since by the choice of l_0 one has $2Z/c \geq l + 1/4 \geq 1/4$, estimate (4.19) will imply that $D\big(\rho^B_{U,\vdash}, \rho^B_{U,\vdash}\big) \leq \text{const } ZK^2 = \mathcal{O}(Z^{5/3})$ and hence complete the proof of Lemma 4.8. By scaling it suffices to prove (4.19) for $c = 1$, which we will assume in the sequel. The Hardy-Littlewood-Sobolev inequality (see, e.g., [43, Thm. 4.3]) implies that

$$D\big(\rho^B_{U,\vdash}, \rho^B_{U,\vdash}\big) \leq \text{const } \big\|\rho^B_{U,\vdash}\big\|^2_{6/5}. \tag{4.20}$$

The triangle inequality together with the definition of $\mathcal{U}$ and (B.2) yields

$$\|\rho_{U,\vdash}\|_{6/5} \leq \sum_{\alpha\in\mathcal{A}} \sum_{\nu=0,1} \|\Phi_\nu \psi_\alpha\|^2_{12/5}, \tag{4.21}$$

where $\{\psi_\alpha | \alpha \in \mathcal{A}\}$ stands for the collection of normalized eigenfunctions building up $d^B_\vdash$, i.e., the corresponding sum ranges over all indices (j, l, m, n). We further estimate with the help of Lemma B.1 and Theorem 2.4,

$$\|\Phi_\nu \psi_\alpha\|^2_{12/5} \leq \text{const } \|\psi_\alpha\|^2_{12/5} \leq \text{const }.$$

This, together with (4.20), (4.21) and the fact that the number of indices in $\mathcal{A}$ is bounded by a constant times K proves (4.19). $\square$

4.2.3 Finishing the proof

We repeat (4.14),

$$E^S(Z) - E^B_c(Z) \geq \text{tr}[S[\tfrac{Z}{|\mathbf{x}|}]d^S_<] - \text{tr}[B_c[\tfrac{Z}{|\mathbf{x}|}]d^B_<] - \mathcal{R}_1 - \mathcal{R}_2 - \mathcal{R}_3 - \text{const } Z^{\frac{47}{24}}.$$

By Lemmata 4.6, 4.7, and 4.8 we have uniformly in $\kappa = Z/c \in (0, \kappa^B]$

$$\mathcal{R}_1 = \mathcal{O}(Z^{23/12}), \quad \mathcal{R}_2 = \mathcal{O}(Z^{5/3}), \quad \mathcal{R}_3 = \mathcal{O}(Z^{11/6}\log Z),$$

so these terms are of lower order than $Z^{47/24}$. Next, we scale $\mathbf{x} \mapsto \mathbf{x}/c$ and obtain

$$\text{tr}[S[Z|\mathbf{x}|^{-1}]d^S_<] - \text{tr}[B_c[Z|\mathbf{x}|^{-1}]d^B_<] = Z^2 s(\kappa) - \mathcal{R}_4$$

where $s(\kappa)$ is introduced in (1.8) and

$$\mathcal{R}_4 := Z^2\kappa^{-2}\sum_{l=0}^{L-1}(2l+1)\sum_{j=l\pm 1/2}\sum_{n=K-l+1}^{\infty}\left(\lambda_n(s_l(\kappa)) - \lambda_n(b_{j,l}(\kappa))\right).$$

By Theorem 2.1 there is a constant such that for all $0 < \kappa \le \kappa^B$

$$\begin{aligned}0 \le \mathcal{R}_4 &\le Z^2\kappa^{-2}\sum_{l=0}^{L-1}(2l+1)\sum_{j=l\pm 1/2}\sum_{n=K-l+1}^{\infty}|\lambda_n(b_{j,l}(\kappa)|\\ &\le \text{const } Z^2\sum_{l=0}^{L-1}(2l+1)\sum_{n=K-l+1}^{\infty}(n+l)^{-2}\\ &\le \text{const } Z^2L^2K^{-1} = \mathcal{O}(Z^{11/6}).\end{aligned}$$

This concludes the proof of the lower bound and hence of our main result. □

Acknowledgments. We thank Elliott Lieb and Robert Seiringer for supportive discussions. H.S. thanks the Departments of Mathematics and Physics of Princeton University and R.F. and S.W. thank the Department of Mathematics of LMU Munich for hospitality while parts of this work were done. We also thank Volker Bach for stimulating questions resulting in several improvements. The work has been partially supported by the Deutscher Akademischer Austauschdienst, grant D/06/49117 (R.F.), the U.S. National Science Foundation, grant PHY 01 39984 (H.S.), the Deutsche Forschungsgemeinschaft, grant SI 348/13-1 (H.S.), and a Sloan Fellowship (S.W.).

Note added in proof. After this paper has been submitted, one of us (R. F.) found an easier proof of a stronger inequality than (2.19). This proof is based on an inequality from [65] as well as Lemma 2.6 in the present paper; see [29].

A Partial wave analysis

For the convenience of the reader and for normalization of the notation we gather some fact on the partial wave analysis of the Brown-Ravenhall operator. We denote by $Y_{l,m}$ the normalized spherical harmonics on the unit sphere $\mathbb{S}^2$ (see, e.g., [48], p. 421) with the convention that $Y_{l,m} \equiv 0$ if $|m| > l$, and we define for $j \in \mathbb{N}_0 + \frac{1}{2}$, $l \in \mathbb{N}_0$, and $m = -j, \ldots, j$ the spherical spinors

$$\Omega_{j,l,m}(\omega) := \begin{cases} \begin{pmatrix} \sqrt{\frac{j+m}{2j}}\,Y_{l,m-\frac{1}{2}}(\omega) \\ \sqrt{\frac{j-m}{2j}}\,Y_{l,m+\frac{1}{2}}(\omega) \end{pmatrix} & \text{if } j = l + \frac{1}{2}, \\ \begin{pmatrix} -\sqrt{\frac{j-m+1}{2j+2}}\,Y_{l,m-\frac{1}{2}}(\omega) \\ \sqrt{\frac{j+m+1}{2j+2}}\,Y_{l,m+\frac{1}{2}}(\omega) \end{pmatrix} & \text{if } j = l - \frac{1}{2}. \end{cases} \tag{A.1}$$

The set of admissible indices is $\mathcal{I} := \{(j,l,m) : j \in \mathbb{N} - 1/2,\ l = j \pm 1/2,\ m = -j, ..., j\}$. It is known that the functions $\Omega_{j,l,m}$, $(j,l,m) \in \mathcal{I}$, form an orthonormal basis of the Hilbert space $L^2(\mathbb{S}^2;\mathbb{C}^2)$. They are joint eigenfunctions of $\mathbf{J}^2$,

J_3, and $\mathbf{L}^2$ with eigenvalues given by $j(j+1)$, $l(l+1)$, and m. The subspace $\mathfrak{H}_{j,l,m}$ corresponding to the joint eigenspace of total angular momentum $\mathbf{J}^2$ with eigenvalue $j(j+1)$ and angular momentum $\mathbf{L}^2$ with eigenvalue $l(l+1)$ is then given by

$$\mathfrak{H}_{j,l,m} = \mathrm{span}\{\mathbf{x} \mapsto |\mathbf{x}|^{-1} f(|\mathbf{x}|)\, \Omega_{j,l,m}(\omega_{\mathbf{x}}) \mid f \in L^2(\mathbb{R}_+)\}$$

where $\omega_{\mathbf{x}} := \mathbf{x}/|\mathbf{x}|$. This leads to the orthogonal decomposition

$$\mathfrak{H} = \bigoplus_{j\in\mathbb{N}_0+\frac{1}{2}} \bigoplus_{l=j\pm 1/2} \mathfrak{H}_{j,l}, \qquad \mathfrak{H}_{j,l} = \bigoplus_{m=-j}^{j} \mathfrak{H}_{j,l,m}, \tag{A.2}$$

of the Hilbert space of two spinors.

We note that the Fourier transform,

$$\hat{\psi}(\mathbf{p}) := (2\pi)^{-3/2} \int_{\mathbb{R}^3} e^{-i\mathbf{p}\cdot\mathbf{x}} \psi(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{A.3}$$

leaves the spaces $\mathfrak{H}_{j,l}$ invariant. Namely, if we decompose ψ according to (A.2),

$$\psi(\mathbf{x}) = \sum_{(j,l,m)\in\mathcal{I}} r^{-1} \psi_{j,m,l}(r)\, \Omega_{j,l,m}(\omega_{\mathbf{x}}),$$

then

$$\hat{\psi}(\mathbf{p}) = \sum_{(j,l,m)\in\mathcal{I}} p^{-1} \left(\mathcal{F}_l \psi_{j,m,l}\right)(p)\, \Omega_{j,l,m}(\omega_{\mathbf{p}}) \tag{A.4}$$

with the Fourier-Bessel transform

$$(\mathcal{F}_l f)(p) = i^{-l} \sqrt{\frac{2}{\pi}} \int_0^\infty f(r) j_l(rp) rp\,\mathrm{d}r. \tag{A.5}$$

Here j_l is a spherical Bessel function. Moreover,

$$\|\psi\|^2 = \sum_{(j,l,m)\in\mathcal{I}} \int_0^\infty |\psi_{j,m,l}(r)|^2 \mathrm{d}r = \sum_{(j,l,m)\in\mathcal{I}} \int_0^\infty |(\mathcal{F}_l \psi_{j,m,l})(p)|^2 \mathrm{d}p = \|\hat{\psi}\|^2.$$

B Properties of the twisting operators

We define the helicity operator $\mathrm{H} = \omega_{\mathbf{p}} \cdot \boldsymbol{\sigma}$ on $\mathfrak{H}$ by

$$\widehat{\mathrm{H}\psi}(\mathbf{p}) := \boldsymbol{\sigma} \cdot \boldsymbol{\omega}_{\mathbf{p}} \hat{\psi}(\mathbf{p}). \tag{B.1}$$

It follows from the pointwise identity

$$(\omega_{\mathbf{p}} \cdot \boldsymbol{\sigma}) \Omega_{j,l,m}(\omega_{\mathbf{p}}) = -\Omega_{j,2j-l,m}(\omega_{\mathbf{p}}), \tag{B.2}$$

see, e.g., Greiner [33, p. 171, (12)], that H is an isomorphism between $\mathfrak{H}_{j,l}$ and $\mathfrak{H}_{j,2j-l}$. Moreover, since $(\boldsymbol{\sigma}\cdot\boldsymbol{a})(\boldsymbol{\sigma}\cdot\boldsymbol{b}) = \boldsymbol{a}\cdot\boldsymbol{b} + i\boldsymbol{\sigma}\cdot(\boldsymbol{a}\times\boldsymbol{b})$ for any $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^3$, we infer that H is an involution on $\mathfrak{H}$, i.e., $\mathrm{H}=\mathrm{H}^{-1}$.
We shall need to consider H on L^p spaces with $p\neq 2$. The relevant properties are summarized in the next lemma, together with those of the operators

$$\widehat{\Phi_\nu\psi}(\mathbf{p}) := \Phi_\nu(\mathbf{p})\,\hat\psi(\mathbf{p}), \tag{B.3}$$

introduced in (1.4). Note that while Φ_0 acts trivially on the spin, Φ_1 involves the helicity H.

Lemma B.1 (L^p-properties of H and Φ_ν). *The operators* H *and* Φ_ν, $\nu = 0,1$, *extend to bounded operators from* $L^p(\mathbb{R}^3,\mathbb{C}^2)$ *to* $L^p(\mathbb{R}^3,\mathbb{C}^2)$ *for any* $p\in(1,\infty)$.

Proof. The L^p-boundedness of H follows from that of the Riesz transformation, see [68, Ch. II-III]. Therefore, to prove the statement about the operators Φ_ν, it suffices to consider the operators ϕ_ν defined analogously as in (B.3) on $L^2(\mathbb{R}^3)$. Since $\mathbf{p}\mapsto\phi_\nu(p)$ is smooth away from the origin and $p^k\partial^k\phi_\nu$ is bounded for $k=0,1,2$, the Hörmander-Mihlin multiplier theorem [68, Thm. IV.3] implies that ϕ_ν extend to bounded operators from $L^p(\mathbb{R}^3)$ to $L^p(\mathbb{R}^3)$ for any $p\in(1,\infty)$. □

Lemma B.2. *For all* $\mathbf{p},\mathbf{q}\in\mathbb{R}^3$

$$\begin{aligned} 1 &- \Phi_0(\mathbf{p})\Phi_0(\mathbf{q}) - \Phi_1(\mathbf{p})\Phi_1(\mathbf{q}) \\ &= \frac12\sum_{\nu=0}^1(\Phi_\nu(\mathbf{p})-\Phi_\nu(\mathbf{q}))^2 + \frac12(\Phi_1(\mathbf{q})\Phi_1(\mathbf{p}) - \Phi_1(\mathbf{p})\Phi_1(\mathbf{q}))\,. \end{aligned} \tag{B.4}$$

and furthermore

$$\begin{aligned} |\Phi_0(\mathbf{p})-\Phi_0(\mathbf{q})|^2 &\le \frac{|\mathbf{p}-\mathbf{q}|^2}{8E(\mathbf{p})^2E(\mathbf{q})^2}\\ |\Phi_1(\mathbf{p})-\Phi_1(\mathbf{q})|^2 &\le \frac{|\mathbf{p}-\mathbf{q}|^2}{E(\mathbf{p})E(\mathbf{q})}\\ |\Phi_1(\mathbf{q})\Phi_1(\mathbf{p})-\Phi_1(\mathbf{p})\Phi_1(\mathbf{q})| &\le \frac{\sqrt{|\mathbf{p}||\mathbf{q}|}|\mathbf{p}-\mathbf{q}|}{E(\mathbf{p})E(\mathbf{q})} \end{aligned}$$

Proof. The first equality is an immediate consequence of the definition of Φ_0 and Φ_1. From this definition we also conclude by an explicit calculation that

$$|\Phi_0(\mathbf{p})-\Phi_0(\mathbf{q})|^2 = (\phi_0(\mathbf{p})-\phi_0(\mathbf{q}))^2 \le \frac{|\mathbf{p}-\mathbf{q}|^2}{8E(\mathbf{p})^2E(\mathbf{q})^2}. \tag{B.5}$$

Moreover, for a proof of the next inequality we write

$$|\Phi_1(\mathbf{p})-\Phi_1(\mathbf{q})|^2 = (\phi_1(\mathbf{p})-\phi_1(\mathbf{q}))^2 + \phi_1(\mathbf{p})\phi_1(\mathbf{q})|\omega_{\mathbf{p}}-\omega_{\mathbf{q}}|^2,$$

and estimate the last two terms with the help of the inequalities

$$(\phi_1(\mathbf{p}) - \phi_1(\mathbf{q}))^2 \leq \frac{(|\mathbf{p}| - |\mathbf{q}|)^2}{2E(\mathbf{p})^2 E(\mathbf{q})^2} \leq \frac{|\mathbf{p} - \mathbf{q}|^2}{2E(\mathbf{p})^2 E(\mathbf{q})^2}, \tag{B.6}$$

and

$$\phi_1(\mathbf{p}) \leq \frac{1}{\sqrt{2}} \frac{|\mathbf{p}|}{E(\mathbf{p})} \quad \text{and} \quad |\omega_{\mathbf{p}} - \omega_{\mathbf{q}}|^2 \leq \frac{|\mathbf{p} - \mathbf{q}|^2}{|\mathbf{p}||\mathbf{q}|}. \tag{B.7}$$

Finally, for a proof of the last inequality we use

$$\begin{aligned} |\Phi_1(\mathbf{q})\Phi_1(\mathbf{p}) - \Phi_1(\mathbf{p})\Phi_1(\mathbf{q})| &= 2\phi_1(\mathbf{p})\phi_1(\mathbf{q})\, |\boldsymbol{\sigma} \cdot (\omega_{\mathbf{p}} \times \omega_{\mathbf{q}})| \\ &\leq 2\phi_1(\mathbf{p})\phi_1(\mathbf{q}) |\omega_{\mathbf{p}} - \omega_{\mathbf{q}}|. \end{aligned}$$

Using again (B.7) concludes the proof of the third inequality. □

C Basics of relativistic hydrogenic operators

In this section we collect – following [17] – some basic properties of the operators B_κ and C_κ which describe hydrogenic atoms in the Brown-Ravenhall respectively Chandrasekhar model. For pedagogical reasons we first discuss their massless analogues,

$$B_\kappa^{(0)} := |\mathbf{p}| - \frac{\kappa}{2} \left(|\mathbf{x}|^{-1} + \omega_{\mathbf{p}} \cdot \boldsymbol{\sigma}\ |\mathbf{x}|^{-1}\ \omega_{\mathbf{p}} \cdot \boldsymbol{\sigma} \right), \quad C_\kappa^{(0)} := |\mathbf{p}| - \kappa |\mathbf{x}|^{-1}. \tag{C.1}$$

C.1 Massless case

Expanding $\hat{\psi}$ as in (A.4) and using (B.2) yields [17] the following partial diagonalization of the massless operators,

$$\langle \psi, B_\kappa^{(0)} \psi \rangle = \sum_{(l,m,s) \in \mathcal{I}} \langle \mathcal{F}_l \psi_{j,m,l}, b_j^{(0)}(\kappa)\, \mathcal{F}_l \psi_{j,m,l} \rangle, \tag{C.2}$$

$$\langle \psi, C_\kappa^{(0)} \psi \rangle = \sum_{(l,m,s) \in \mathcal{I}} \langle \mathcal{F}_l \psi_{j,m,l}, c_l^{(0)}(\kappa)\, \mathcal{F}_l \psi_{j,m,l} \rangle. \tag{C.3}$$

Here the operators $b_j^{(0)}(\kappa)$ and $c_l^{(0)}(\kappa)$ are densely defined in $L^2(\mathbb{R}_+)$ through their quadratic forms,

$$\langle f, b_j^{(0)}(\kappa) f \rangle := \int_0^\infty p\, |f(p)|^2 \mathrm{d}p - \kappa \int_0^\infty \int_0^\infty \overline{f(p)}\, k_j^B(p,q)\, f(q)\, \mathrm{d}q\, \mathrm{d}p,$$

$$\langle f, c_l^{(0)}(\kappa) f \rangle := \int_0^\infty p\, |f(p)|^2 \mathrm{d}p - \kappa \int_0^\infty \int_0^\infty \overline{f(p)}\, k_l^C(p,q)\, f(q)\, \mathrm{d}q\, \mathrm{d}p,$$

with maximal form domain denoted by $\mathfrak{Q}(b_j^{(0)}(\kappa))$ and $\mathfrak{Q}(c_l^{(0)}(\kappa))$. In the above expression, the integral kernels k_j^B and k_l^C are given by

$$k_j^B(p,q) := \frac{1}{2\pi} \left[Q_{j-1/2} \left(\tfrac{1}{2} \left(\tfrac{p}{q} + \tfrac{q}{p} \right) \right) + Q_{j+1/2} \left(\tfrac{1}{2} \left(\tfrac{p}{q} + \tfrac{q}{p} \right) \right) \right], \tag{C.4}$$

$$k_l^C(p,q) := \frac{1}{\pi}\, Q_l \left(\tfrac{1}{2} \left(\tfrac{p}{q} + \tfrac{q}{p} \right) \right), \tag{C.5}$$

where Q_l are the Legendre functions of the second kind, i.e.,

$$Q_l(z) = \tfrac{1}{2}\int_{-1}^{1} P_l(t)(z-t)^{-1}\mathrm{d}t \tag{C.6}$$

with P_l standing for Legendre polynomials; see Stegun [67] for the notation and some properties of these special functions.
It was proved in [17] and [40, Eq. (5.33)] that the operators (C.1) are self-adjoint and lower bounded if and only if $\kappa \le \kappa^{\#}$, $\# = B, C$, cf. (2.3). More can be said about the reduced operators $b_j^{(0)}(\kappa)$ and $c_l^{(0)}(\kappa)$. They are lower bounded (in fact, non-negative) if and only if

$$\frac{1}{\kappa} \ge \frac{1}{\kappa_j^B} := \int_0^\infty k_j^B(1,t)\,\frac{\mathrm{d}t}{t}, \tag{C.7}$$

$$\frac{1}{\kappa} \ge \frac{1}{\kappa_l^C} := \int_0^\infty k_l^C(1,t)\,\frac{\mathrm{d}t}{t}. \tag{C.8}$$

This follows by the same lines of reasoning as in [17].
Since [67, (8.4)] $P_0(t) = 1$, $P_1(t) = t$, we have

$$Q_0(t) = \frac{1}{2}\log\frac{t+1}{t-1}, \quad Q_1(t) = \frac{t}{2}\log\frac{t+1}{t-1} - 1, \tag{C.9}$$

such that $\kappa_0^C = 2/\pi$, $\kappa_1^C = \pi/2$ and thus $\kappa_{1/2}^B = 2/(2/\pi + \pi/2)$.
The critical coupling constants κ_j^B and κ_l^C are strictly increasing in j and l and, in particular, $\kappa_{1/2}^B = \kappa^B$ and $\kappa_0^C = \kappa^C$. This follows from the pointwise monotonicity

$$Q_l(t) \ge Q_{l'}(t) \qquad \text{for } l' \ge l \text{ and } t > 1 \tag{C.10}$$

which, in turn, is evident from the integral representation

$$Q_l(x) = \int_{x+\sqrt{x^2-1}}^\infty \frac{z^{-l-1}}{\sqrt{1-2xz+z^2}}\mathrm{d}z, \qquad x > 1;$$

see Whittaker and Watson [77, p. 334, Chap. X, Sec. 3.2].

C.2 Massive case

Similarly as in the previous subsection, one obtains the following partial diagonalization of the massive hydrogenic Brown-Ravenhall and Chandrasekhar operators,

$$\langle \psi, B_\kappa \psi\rangle = \sum_{(l,m,s)\in\mathcal{I}} \langle \mathcal{F}_l\psi_{j,m,l}, b_{j,l}(\kappa)\,\mathcal{F}_l\psi_{j,m,l}\rangle, \tag{C.11}$$

$$\langle \psi, C_\kappa \psi\rangle = \sum_{(j,l,m)\in\mathcal{I}} \langle \mathcal{F}_l\psi_{j,m,l}, c_l(\kappa)\,\mathcal{F}_l\psi_{j,m,l}\rangle. \tag{C.12}$$

Here the operators $b_{j,l}(\kappa)$ and $c_l(\kappa)$ are densely defined in $L^2(\mathbb{R}_+)$ through their quadratic forms,

$$\begin{aligned}
&\langle f, b_{j,l}(\kappa) f\rangle \\
&:= \int_0^\infty (E(p)-1)|f(p)|^2 \mathrm{d}p - \kappa \int_0^\infty \int_0^\infty \overline{f(p)}\, k_{j,l}^B(p,q)\, f(q)\, \mathrm{d}q\, \mathrm{d}p, \qquad \text{(C.13)}\\
&\langle f, c_l(\kappa) f\rangle \\
&:= \int_0^\infty (E(p)-1)|f(p)|^2 \mathrm{d}p - \kappa \int_0^\infty \int_0^\infty \overline{f(p)}\, k_l^C(p,q)\, f(q)\, \mathrm{d}q\, \mathrm{d}p \qquad \text{(C.14)}
\end{aligned}$$

with maximal form domain denoted by $\mathfrak{Q}(b_{j,l}(\kappa))$ and $\mathfrak{Q}(c_l(\kappa))$, cf. [17]. In the above expression, the integral kernel $k_{j,l}^B$ depends, in contrast to the massless case, on both j and l and is given by

$$\begin{aligned}
&k_{j,l}^B(p,q) \\
&\quad := \frac{1}{\pi}\left[\phi_0(p) Q_l\left(\tfrac{1}{2}\left(\tfrac{p}{q}+\tfrac{q}{p}\right)\right)\phi_0(q) + \phi_1(p) Q_{2j-l}\left(\tfrac{1}{2}\left(\tfrac{p}{q}+\tfrac{q}{p}\right)\right)\phi_1(q)\right].
\end{aligned}$$

The form (C.13) defines a self-adjoint semi-bounded operator $b_{j,l}(\kappa)$ if and only if $\kappa \le \kappa_j^B$ (Evans et al. [17]). A trivially modified argument shows that (C.14) defines a self-adjoint semi-bounded operator $c_l(\kappa)$ if and only if $\kappa \le \kappa_l^C$. In fact, the semiboundedness of the massive cases and the massless cases are equivalent, since the differences of the massive and massless forms are bounded (Tix [75, Thm. 1]). (One even knows that $b_{j,l}(\kappa)+1$ is non-negative (Tix [76]).)

D Critical Chandrasekhar operator on a finite domain

Lieb and Yau [46] have shown that the critical Chandrasekhar operator $|\mathbf{p}| - \kappa^C|\mathbf{x}|^{-1}$ when restricted to a ball has only discrete spectrum with eigenvalues accumulating at infinity at the rate predicted by the semiclassical result for $|\mathbf{p}|$ alone. This is remarkable since the semiclassical phase-space volume corresponding to $|\mathbf{p}| - \kappa|\mathbf{x}|^{-1}$ is infinite.

We aim at proving an analogous result for the Chandrasekhar operator restricted to a ball and restricted to the subspace of fixed angular momentum. In the proof of Theorem 2.1 it is essential to handle coupling constants which are larger than κ^C, all the way up to and including κ_1^C.

In order to define the above operator we consider for $R > 0$ and $l \in \mathbb{N}$ the Hilbert space

$$\mathfrak{F}_l(R) := \left\{ f \in L^2(0,\infty) \,\middle|\, \left(\mathcal{F}_l^{-1} f\right)(r) = 0 \text{ for all } r \ge R \right\},$$

where $\mathcal{F}_l$ denotes the Fourier-Bessel transformation, cf. (A.5). The quadratic form given by $\langle f, c_l^{(0)}(\kappa) f\rangle$ with domain $\mathfrak{F}_l(R) \cap \mathfrak{Q}(c_l^{(0)}(\kappa))$ defines for all $\kappa \le \kappa_l^C$ a self-adjoint, non-negative operator in $\mathfrak{F}_l(R)$ which we will denote by $c_l^{(0)}(\kappa, R)$.

Lemma D.1. *Let $l \in \mathbb{N}$. There is a constant such that for all $R > 0$, $\mu > 0$, and $\kappa \leq \kappa_l^C$*

$$\operatorname{tr}\left(c_l^{(0)}(\kappa, R) - \mu\right)_- \leq \text{const } \mu^2 R. \tag{D.1}$$

We have not tried to track the l-dependence of the constant, since the cases $l = 0, 1$ will be enough for our purpose.

Proof. For a proof of (D.1) we basically follow the argument in [46]. The starting point is the following reduction to a simpler variational problem involving only functions. Namely, for any non-negative function $h : \mathbb{R}_+ \to \mathbb{R}_+$, let

$$t(p) := \frac{\kappa_l^C}{\pi h(p)} \int_0^\infty Q_l\big(\tfrac{1}{2}(\tfrac{p}{q} + \tfrac{q}{p})\big) h(q)\, \mathrm{d}q\,.$$

Then

$$-\operatorname{tr}\left(c_l(\kappa_l^C, R) - \mu\right)_- \geq \inf\left\{\int_0^\infty \sigma(p)\,(p - \mu - t(p))\,\mathrm{d}p \,\middle|\, 0 \leq \sigma \leq M_l\right\} \tag{D.2}$$

where $M_l := R \sup_{r>0}(2/\pi) r^2 j_l^2(r)$. The proof of (D.2) is analogous to the one of [46, Eq. (7.8)]. We merely replace the Fourier transformation in $\mathbb{R}^3$ by the Fourier-Bessel transformation $\mathcal{F}_l$ in $\mathbb{R}_+$.

From now on we assume that $l \geq 1$ and comment on the necessary changes in case $l = 0$ at the end. We choose h of the form

$$h(p) = \begin{cases} p^{-1} - (A/2)p^{-2} & \text{if } p > A, \\ (2A)^{-1} & \text{if } p \leq A. \end{cases}$$

Below we shall show that the constant A can be picked in such a way that for some $\delta > 0$

$$p - \mu - t(p) \geq \begin{cases} 0 & \text{if } p \geq \delta^{-1}A, \\ -\text{const } A^{-1}\mu^2 & \text{if } p < \delta^{-1}A. \end{cases} \tag{D.3}$$

In view of (D.2) this will prove the result, since then

$$\begin{aligned} &\inf\left\{\int_0^\infty \sigma(p)\,(p - \mu - t(p))\,\mathrm{d}p \,\middle|\, 0 \leq \sigma \leq M_l\right\} \\ &\geq -\text{ const } A^{-1}\mu^2 M_l \int_0^{\delta^{-1}A} \mathrm{d}p = -\text{const } \delta^{-1}\mu^2 M_l. \end{aligned}$$

To prove (D.3) we recall that $\int_0^\infty Q_l\big(\frac{1}{2}(t + \frac{1}{t})\big)\,\frac{dt}{t} = \pi(\kappa_l^C)^{-1}$, cf. (C.8), and hence by a straightforward calculation

$$\begin{aligned} p - t(p) &= p\,\frac{\kappa_l^C}{\pi} \int_0^\infty Q_l\big(\tfrac{1}{2}(t + \tfrac{1}{t})\big) \left(\tfrac{1}{t} - \tfrac{h(tp)}{h(p)}\right) \mathrm{d}t \\ &= p\,\frac{\kappa_l^C}{\pi} \begin{cases} \frac{A/2p}{1 - A/2p}\,(F(1) - F(A/p)) & \text{if } p \geq A, \\ (-F(1) + F(p/A)) & \text{if } p < A. \end{cases} \end{aligned}$$

Here for $0 < s \leq 1$ we have set

$$F(s) := \int_0^s Q_l\big(\tfrac{1}{2}(t+\tfrac{1}{t})\big)\,\big(\tfrac{1}{t}-\tfrac{1}{s}\big)^2\,\mathrm{d}t\,.$$

Since $Q_l(\tau) \leq Q_1(\tau)$, which vanishes like a constant times τ^{-2} as $\tau \to \infty$, one has $F(s) \to 0$ as $s \to 0$. Choosing $\delta \in (0,1)$ such that $F(s) \leq \frac{1}{2}F(1)$ for all $0 < s \leq \delta$, we have shown that for all $p \geq \delta^{-1}A$ one has

$$p - t(p) \geq \frac{A\;F(1)}{8(1-A/2p)} \geq A\;\frac{F(1)}{8}\,.$$

For $A \leq p < \delta^{-1}A$ we use the monotonicity, $dF/ds \geq 0$, to bound

$$p - t(p) \geq 0.$$

Finally, for $0 \leq p < A$ we drop the term $F(p/A) \geq 0$ to obtain

$$p - t(p) \geq -\frac{p}{2}F(1) \geq -A\frac{F(1)}{2}.$$

Choosing $A := 8\mu/F(1)$ yields the claimed inequality (D.3).
In case $l = 0$, the function h can be chosen as before. However, the corresponding expressions $F(1) - F(s)$ should be interpreted as a single integral, and estimated with slightly more care. □

Corollary D.2. *Let $l \in \mathbb{N}$. Then there is a constant such that for all $0 < \kappa \leq \kappa_l^C$, all $\mu > 0$ and all functions χ on $\mathbb{R}_+$ with $\chi = 0$ on $[R,\infty)$ for some $R > 0$ one has*

$$N_l(0, \chi\left(|\mathbf{p}| - \kappa|\mathbf{x}|^{-1} - \mu\right)\chi) \leq \mathrm{const}\;\mu R.$$

Proof. The variational principle implies that

$$N_l(0, \chi\left(|\mathbf{p}| - \kappa|\mathbf{x}|^{-1} - \mu\right)\chi) \leq N(\mu, c_l^{(0)}(\kappa, R)).$$

Indeed, if $\mathcal{V}_l$ is the negative spectral subspace of $\chi\left(|\mathbf{p}| - \kappa|\mathbf{x}|^{-1} - \mu\right)\chi$ with fixed l, then any $f \in \mathcal{F}_l\chi\mathcal{V}_l \subset \mathfrak{F}_l(R)$ satisfies $\langle f, (c_l^{(0)}(\kappa,R) - \mu)\,f\rangle < 0$. Hence, in order to prove the assertion, it suffices to show that

$$N(\mu, c_l^{(0)}(\kappa, R)) \leq \mathrm{const}\;\mu R.$$

For a proof, we note that the elementary inequality $\chi_{(-\infty,\mu)}(E) \leq \frac{(E-\lambda)_-}{\lambda-\mu}$, valid for any $\mu < \lambda$, together with Lemma D.1 implies that

$$N(\mu, c_l^{(0)}(\kappa,R)) \leq (\lambda-\mu)^{-1}\,\mathrm{tr}(c_l^{(0)}(\kappa,R) - \lambda)_- \leq \mathrm{const}\;(\lambda-\mu)^{-1}\lambda^2 R.$$

The proof is completed by optimizing over λ. □

E The trial density matrix

In this section we define the density matrices d^S and d^B that we use to bound the Schrödinger energy, respectively the Brown-Ravenhall energy, from above. Both density matrices are split into two parts corresponding to low and high angular momenta

$$d^S := d^S_< + d_>, \qquad d^B := d^B_< + d_>.$$

Low angular momenta correspond to orbits whose perinucleon is close to the nucleus, while high angular momenta ensure that the orbits are never close to the nucleus. We will cut between these two at $L := [Z^{1/12}]$.

E.1 Low angular momenta

In the vicinity of the nucleus the nuclear attraction dominates the interaction with the other electrons. This motivates to choose the orbitals as the ones of the Bohr atom, i.e., as the eigenfunctions of the unscreened operator with nuclear charge Z. The corresponding density matrices $d^\#_<$ are of the form

$$d^\#_< = \sum_{l=0}^{L-1} d^\#_j, \qquad d^\#_l = \sum_{j=l\pm 1/2,\ j\geq 1/2} d^\#_{j,l}$$

and

$$d^\#_{j,l} = \sum_{m=-j}^{j} \sum_{n=1}^{K-l} |\psi^\#_{j,l,m,n}\rangle\langle\psi^\#_{j,l,m,n}|.$$

Here $K = [\mathrm{const}\ Z^{1/3}]$ with some positive constant, i.e., on the order of the last occupied shell of the Bohr atom. We now turn to the definition of the orbitals $\psi^\#_{j,l,m,n}$ for which we consider the cases $\# = B, S$ separately.

In the Brown-Ravenhall case we choose $\psi^B_{j,l,m,n}$ such that its Fourier transform is

$$\hat\psi^B_{j,l,m,n}(\mathbf{p}) = p^{-1} f^B_{j,l,n}(p)\Omega_{j,l,m}(\omega_{\mathbf{p}}),$$

where $f^B_{j,l,n}$ is the n-th eigenfunction of the operator $V_c\, b_{j,l}(Z/c)\, V_c^*$ in $L^2(\mathbb{R}_+)$. Here the unitary scaling operator V_c is defined by $(V_c f)(p) := c^{-1/2} f(p/c)$ and we recall that the operator $b_{j,l}(\kappa)$ was defined in Subsection C.2. The operators $V_c\, b_{j,l}(Z/c)\, V_c^*$ appear as the angular momentum reductions of $B_c[Z|\mathbf{x}|^{-1}]$. Indeed, by (C.11) and scaling one has

$$\langle\psi, B_c[Z|\mathbf{x}|^{-1}]\,\psi\rangle = c^2 \sum_{(j,l,m)\in\mathcal{I}} \langle\hat\psi_{j,m,l}, V_c\, b_{j,l}(Z/c)\, V_c^*\, \hat\psi_{j,m,l}\rangle.$$

In the Schrödinger case we choose

$$\psi^S_{j,l,m,n}(\mathbf{x}) = r^{-1} f^S_{l,n}(r)\Omega_{j,l,m}(\omega_{\mathbf{x}}),$$

where $f^S_{l,n}$ is the n-th eigenfunction of $-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{l(l+1)}{2r^2} - \frac{Z}{r}$ in $L^2(\mathbb{R}_+)$ with Dirichlet boundary conditions.

E.2 High angular momenta

For large angular momenta, the electrons are sufficiently far from the center moving – classically speaking – slowly. This motivates to pick non-relativistic orbitals in both in the relativistic and non-relativistic case. Moreover, for large quantum numbers the correspondence principle would predict quasi-classical behavior (in the quantum sense) as well. This motivates the following choice which we take – with slight modifications – from [55]:

$$d_> := \sum_{l\geq L} d_l,\quad d_l := \sum_{j=l\pm 1/2}\sum_{m=-j}^{j}\sum_{n\in\mathbb{Z}} w_{n,l}|\varphi_{n,l}\Omega_{j,l,m}\rangle\langle\varphi_{n,l}\Omega_{j,l,m}|. \tag{E.1}$$

We repeat at this point the construction of the Macke orbitals $\varphi_{n,l}$ and their weights $w_{n,l}$. We will also present a new estimate not directly given in that paper.

The semi-classical mean-field in which the electrons move is the Thomas-Fermi potential ϕ_{TF} (see (4.3)). According to Hellmann [36] the semi-classical electron density for fixed angular momentum is

$$\sigma_l^H(r) := \frac{2(2l+1)}{\pi}\sqrt{2\left[n_Z\phi_{\mathrm{TF}}(r) - \frac{(l+\frac12)^2}{2r^2}\right]_+}, \tag{E.2}$$

where we added the factor $n_Z = (1-aZ^{-1/2})^{2/3}$ for normalization purposes with some fixed positive a and where we replaced the self-generated field of the sum of the radial densities σ_l by the Thomas-Fermi potential. We will write ρ_l^H for the functions σ_l^H when $a=0$, i.e., no normalization factor occurs. In passing we note that the densities ρ_l^H are the minimizers of the Hellmann functional with external potential given by the Thomas-Fermi density and no other interaction between the electrons (see [61]).

The functions σ_l^H vanish for large l and we define

$$k' := \min\{l\in\mathbb{N}\,|\,\sigma_l^H\equiv 0\}.$$

By scaling, k' is of the order $Z^{1/3}$. Moreover, since the function $r\mapsto\phi_{\mathrm{TF}}(r)r^2$ has exactly one maximum, the support of σ_l^H is an interval $[r_1(l), r_2(l)]$.

We cannot use the density σ_l^H directly in defining semi-classical orbitals, since the derivative of its square root is not square integrable. Thus we pick two points,

$$x_1(l) := r_1(l) + T(l+\tfrac12)Z^{-1},\ x_2(l) := r_2(l) - SZ^{-2/3} \tag{E.3}$$

for some positive S and $T\in(0,4)$, and set

$$\rho_l(r) := \begin{cases} 2(2l+1)\alpha^2 r^{2l+2}, & r\in[0,x_1(l)],\\ \sigma_l^H(r), & r\in[x_1(l),x_2(l)],\\ 2(2l+1)\beta^2\exp(-2^{3/2}Z^{2/3}r), & r\in[x_2(l),\infty). \end{cases} \tag{E.4}$$

The constants α and β are chosen such that ρ_l is continuous. We suppress their dependence on l in the notation.
Next, we define for $l < k'$ and $n \in \mathbb{Z}$ the Macke orbitals

$$\varphi_{n,l}(r) := \frac{\sqrt{\zeta_l'(r)}}{r} e^{\mathrm{i}\pi k_{n,l}\zeta_l(r)} \tag{E.5}$$

where $\zeta_l : [0,\infty) \to [0,1)$ is the Macke transform

$$\zeta_l(r) := \frac{\int_0^r \rho_l(t)\mathrm{d}t}{\int_0^\infty \rho_l(t)\mathrm{d}t}. \tag{E.6}$$

For $l \geq k'$ we set $\varphi_{n,l} :\equiv 0$. The integral

$$N_{j,l,m} := \frac{1}{2(2l+1)} \int_0^\infty \rho_l(r)\mathrm{d}r,$$

which is independent of j and m, will represent the number of electrons in the angular momentum channel (j,l,m). Moreover, we set $\varepsilon_l := N_{j,l,m} - [N_{j,l,m}]$. If $[N_{j,l,m}]$ is odd, we pick $k_{n,l} = 2n$, otherwise $k_{n,l} = 2n-1$. The weights are chosen as

$$w_{n,l} := \begin{cases} 1 & |k_{n,l}| \leq [N_{j,l,m}] - 1 \\ \varepsilon_l/2 & |k_{n,l}| = [N_{j,l,m}] + 1 \\ 0 & \text{otherwise} \end{cases} \tag{E.7}$$

which guarantees that $\sum_{n\in\mathbb{Z}} w_{n,l} = N_{j,l,m}$.
Strictly speaking, our trial density matrix differs from the one used in [55], since we label the orbitals by the modulus of total angular momentum, by the third component of total angular momentum, and by the orbital angular momentum. This, however, is merely a minor rearrangement of terms.
We also adapt to atomic units used in this paper which changes the value of the Thomas-Fermi constant and gives a factor 1/2 in front of all three kinetic energy terms in the Hellmann-Weizsäcker functional.

E.3 Energy estimates for high angular momenta

For the convenience of the reader, we gather from [55] (based on the construction in [60]) two estimates on the order of the average kinetic and potential energy of the Schrödinger operator associated with the semi-classical density matrix $d_>$,

$$\mathrm{tr}(\mathbf{p}^2 d_>) = \mathcal{O}(Z^{7/3}), \qquad \mathrm{tr}(|\mathbf{x}|^{-1} d_>) = \mathcal{O}(Z^{4/3}). \tag{E.8}$$

We also need a more detailed estimate on the kinetic energy.

Lemma E.1. *Let $L = [Z^{1/12}]$. Then for large Z,*

$$\sum_{l=L}^\infty l^{-2}\, \mathrm{tr}(\mathbf{p}^2 d_l) = \mathcal{O}(Z^2/L). \tag{E.9}$$

Proof. The definition of d_l implies (cf. [55, (2.3)]) that for angular momenta $l < k'$ one has

$$\mathrm{tr}(\mathbf{p}^2 d_l) = \int_0^\infty \left[\sqrt{\rho_l}\ '^2 + \frac{\alpha_l}{3}\rho_l^3 + \frac{l(l+1)}{r^2}\rho_l \right] \mathrm{d}r + F_l \tag{E.10}$$

where we set

$$F_l := \frac{\alpha_l}{3} \left(\frac{-1+6\varepsilon_l - 3\varepsilon_l^2}{N_{j,l,m}^2} + \frac{2\varepsilon_l^3 - 6\varepsilon_l^2 + 4\varepsilon_l}{N_{j,l,m}^3} \right) \int_0^\infty \rho_l^3 \mathrm{d}r,$$
$$\alpha_l := \frac{\pi^2}{4(2l+1)^2},$$

and emphasize that α_l should not be confused with α from (E.4). According to [55, Proposition 3.6]

$$\sum_{l=L}^\infty l^{-2} F_l \le \sum_{l=L}^\infty F_l \le \text{const } Z^{5/3}$$

where $L = [Z^{1/12}]$. The first term on the right-hand side of (E.10) is estimated according to

$$\int_0^\infty \left[\sqrt{\rho_l}\ '^2 + \frac{\alpha_l}{3}\rho_l^3 + \frac{l(l+1)}{r^2}\rho_l \right] \mathrm{d}r$$
$$\le \int_0^\infty \left[\frac{\alpha_l}{3}\rho_l^H(r)^3 + \frac{(l+\frac{1}{2})^2}{r^2}\rho_l^H(r) \right] \mathrm{d}r + G_l + H_l + I_l. \tag{E.11}$$

with

$$G_l := \int_0^{x_1(l)} \left[\sqrt{\rho_l}\ '^2 + \frac{\alpha_l}{3}\rho_l^3 + \frac{\left(l+\frac{1}{2}\right)^2}{r^2}\rho_l \right] \mathrm{d}r \le \text{const } Z^2 \left(l + \frac{1}{2}\right)^{-3/2},$$
$$H_l := \int_{x_2(l)}^\infty \left[\sqrt{\rho_l}\ '^2 + \frac{\alpha_l}{3}\rho_l^3 + \frac{\left(l+\frac{1}{2}\right)^2}{r^2}\rho_l \right] \mathrm{d}r \le \text{const } Z^{7/6} \left(l + \frac{1}{2}\right)$$

where the inequalities were obtained by integration as in [55, (3.4)]. Inequality [55, (3.9)] for the gradient term in the middle region reads

$$I_l := \int_{x_1(l)}^{x_2(l)} \sqrt{\rho_l}\ '^2 \mathrm{d}r \le \text{const } \left(l + \frac{1}{2}\right)$$
$$\times \left[Z^2 \left(l+\frac{1}{2}\right)^{-3} + Z + Z^2 \left(l + \frac{1}{2}\right)^{-5/2} + \frac{Z^{5/3}\left(l+\frac{1}{2}\right)^{-1/2}}{\min\{l+\frac{1}{2}, Z^{1/4}\}} \right].$$

This implies

$$\begin{aligned}
&\sum_{l=L}^{\infty} l^{-2}\, G_l \leq \text{const}\ Z^2 \sum_{l=L}^{\infty} l^{-7/2} \leq \text{const}\ Z^2 L^{-5/2},\\
&\sum_{l=L}^{\infty} l^{-2}\, H_l \leq \sum_{l=L}^{k'} l^{-2}\, H_l \leq \text{const}\ Z^{7/6}\log k' \leq \text{const}\ Z^{7/6}\log Z\\
&\sum_{l=L}^{\infty} l^{-2}\, I_l \leq \text{const}\ \left[Z^2L^{-3} + Z\log Z + Z^2L^{-5/2} + Z^{5/3}L^{-3/2}\right]\\
&\qquad\qquad \leq \text{const}\ Z^{43/24}.
\end{aligned}$$

It thus remains to estimate the sum of the first terms on the right-hand side of (E.11). We begin with the first summand,

$$\begin{aligned}
&\sum_{l=L}^{\infty}\frac{1}{l^2}\int_0^\infty \frac{\alpha_l}{3}\rho_l^H(r)^3 \mathrm{d}r \leq \text{const}\ \sum_{l=L}^{\infty}\frac{1}{l}\int_0^\infty (Z/r - l^2/r^2)_+^{3/2}\mathrm{d}r\\
=&\text{const}\ Z^2 \sum_{l\geq L}\frac{1}{l^2}\int_0^\infty r^{-3/2}(1-r^{-1})_+^{3/2}\mathrm{d}r = \mathcal{O}(Z^2/L)
\end{aligned}$$

where we used that the Thomas-Fermi potential is bounded from above by Z/r. This leaves the second summand,

$$\begin{aligned}
&\sum_{l=L}^{\infty}\frac{1}{l^2}\int_0^\infty \frac{\left(l+\frac{1}{2}\right)^2}{r^2}\rho_l^H(r)\mathrm{d}r \leq \text{const}\ \sum_{l=L}^{\infty} l\int_0^\infty r^{-2}(Z/r - l^2/r^2)_+^{1/2}\mathrm{d}r\\
=&\ \text{const}\ Z^2 \sum_{l=L}^{\infty}\frac{1}{l^2}\int_0^\infty r^{-5/2}(1-r^{-1})_+^{1/2}\mathrm{d}r = \mathcal{O}(Z^2/L),
\end{aligned}$$

which completes the proof of Lemma E.1. □

References

[1] Volker Bach. A proof of Scott's conjecture for ions. *Rep. Math. Phys.*, 28(2):213–248, October 1989.

[2] Volker Bach. Error bound for the Hartree-Fock energy of atoms and molecules. *Comm. Math. Phys.*, 147:527–548, 1992.

[3] A. A. Balinsky and W. D. Evans. On the virial theorem for the relativistic operator of Brown and Ravenhall, and the absence of embedded eigenvalues. *Lett. Math. Phys.*, 44(3):233–248, 1998.

[4] Pedro Balodis. A proof of Scott's correction for matter. *Comm. Math. Phys.*, 249(1):79–132, 2004.

[5] Hans A. Bethe and Edwin E. Salpeter. Quantum mechanics of one- and two-electron atoms. In S. Flügge, editor, *Handbuch der Physik, XXXV*, pages 88–436. Springer, Berlin, 1 edition, 1957.

[6] G. E. Brown and D. G. Ravenhall. On the interaction of two electrons. *Proc. Roy. Soc. London Ser. A.*, 208:552–559, 1951.

[7] Roch Cassanas and Heinz Siedentop. The ground-state energy of heavy atoms according to Brown and Ravenhall: absence of relativistic effects in leading order. *J. Phys. A*, 39(33):10405–10414, 2006.

[8] Charles G. Darwin. The wave equation of the electron. *Proc. Roy. Soc. (London)*, 1928.

[9] Ingrid Daubechies. An uncertainty principle for Fermions with generalized kinetic energy. *Comm. Math. Phys.*, 90:511–520, September 1983.

[10] Paul A. M. Dirac. Note on exchange phenomena in the Thomas-Fermi atom. *Proc. Cambridge Philos. Soc.*, 26:376–385, 1931.

[11] Jean Dolbeault, Maria J. Esteban, and Eric Séré. On the eigenvalues of operators with gaps. Application to Dirac operators. *J. Funct. Anal.*, 174(1):208–226, 2000.

[12] Berthold-Georg Englert. *Semiclassical Theory of Atoms.* Number 300 in Lecture Notes in Physics. Springer-Verlag, Berlin, 1 edition, 1988.

[13] Berthold-Georg Englert and Julian Schwinger. Statistical atom: Handling the strongly bound electrons. *Phys. Rev. A*, 29(5):2331–2338, 1984.

[14] Berthold-Georg Englert and Julian Schwinger. Statistical atom: Some quantum improvements. *Phys. Rev. A*, 29(5):2339–2352, 1984.

[15] Berthold-Georg Englert and Julian Schwinger. Atomic-binding-energy oscillations. *Phys. Rev. A*, 32:47–63, 1985.

[16] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi. *Tables of integral transforms. Vol. II.* McGraw-Hill Book Company, Inc., New York-Toronto-London, 1954. Based, in part, on notes left by Harry Bateman.

[17] William Desmond Evans, Peter Perry, and Heinz Siedentop. The spectrum of relativistic one-electron atoms according to Bethe and Salpeter. *Comm. Math. Phys.*, 178(3):733–746, July 1996.

[18] C. Fefferman and L. Seco. Eigenfunctions and eigenvalues of ordinary differential operators. *Adv. Math.*, 95(2):145–305, October 1992.

[19] C. Fefferman and L. Seco. The density of a one-dimensional potential. *Adv. Math.*, 107(2):187–364, September 1994.

[20] C. Fefferman and L. Seco. The eigenvalue sum of a one-dimensional potential. *Adv. Math.*, 108(2):263–335, October 1994.

[21] C. Fefferman and L. Seco. On the Dirac and Schwinger corrections to the ground-state energy of an atom. *Adv. Math.*, 107(1):1–188, August 1994.

[22] C. Fefferman and L. Seco. The density in a three-dimensional radial potential. *Adv. Math.*, 111(1):88–161, March 1995.

[23] C. L. Fefferman and L. A. Seco. An upper bound for the number of electrons in a large ion. *Proc. Nat. Acad. Sci. USA*, 86:3464–3465, 1989.

[24] C. L. Fefferman and L. A. Seco. Asymptotic neutrality of large ions. *Comm. Math. Phys.*, 128:109–130, 1990.

[25] C. L. Fefferman and L. A. Seco. On the energy of a large atom. *Bull. AMS*, 23(2):525–530, October 1990.

[26] Charles L. Fefferman and Luis A. Seco. Aperiodicity of the Hamiltonian flow in the Thomas-Fermi potential. *Revista Mathemática Iberoamericana*, 9(3):409–551, 1993.

[27] E. Fermi. Un metodo statistico per la determinazione di alcune proprietá dell'atomo. *Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, Classe di Scienze Fisiche, Matematiche e Naturali*, 6(12):602–607, 1927.

[28] E. Fermi. Eine statistische Begründung zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendungen auf die Theorie des periodischen Systems der Elemente. *Z. Phys.*, 48:73–79, 1928.

[29] Rupert L. Frank. A simple proof of Hardy-Lieb-Thirring inequalities. *Comm. Math. Phys.*, 290(2):789–800, 2009.

[30] Rupert L. Frank, Elliott H. Lieb, and Robert Seiringer. Hardy-Lieb-Thirring inequalities for fractional Schrödinger operators. *J. Amer. Math. Soc.*, 21(4):925–950, 2008.

[31] Rupert L. Frank, Heinz Siedentop, and Simone Warzel. The ground state energy of heavy atoms: relativistic lowering of the leading energy correction. *Comm. Math. Phys.*, 278(2):549–566, 2008.

[32] Walter Gordon. Die Energieniveaus des Wasserstoffatoms nach der Diracschen Quantentheorie. *Z. Phys.*, 48:11–14, 1928.

[33] Walter Greiner. *Relativistic Quantum Mechanics*, volume 3 of *Theoretical Physics – Text and Excercise Books*. Springer, Berlin, 1 edition, 1990.

[34] Marcel Griesemer, Roger T. Lewis, and Heinz Siedentop. A minimax principle for eigenvalues in spectral gaps: Dirac operators with Coulomb potential. *Doc. Math.*, 4:275–283, 1999.

[35] Marcel Griesemer and Heinz Siedentop. A minimax principle for the eigenvalues in spectral gaps. *J. London Math. Soc. (2)*, 60(2):490–500, 1999.

[36] Heinrich Hellmann. Ein kombiniertes Näherungsverfahren zur Energieberechnung im Vielelektronenproblem. II. *Acta Physicochim. U.S.S.R.*, 4:225–244, 1936.

[37] Webster Hughes. *An Atomic Energy Lower Bound that Gives Scott's Correction*. PhD thesis, Princeton, Department of Mathematics, 1986.

[38] Webster Hughes. An atomic lower bound that agrees with Scott's correction. *Adv. in Math.*, 79:213–270, 1990.

[39] V. Ja. Ivrii and I. M. Sigal. Asymptotics of the ground state energies of large Coulomb systems. *Annals of Math.*, 138(2):243–335, 1993.

[40] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1 edition, 1966.

[41] W. Lenz. Über die Anwendbarkeit der statistischen Methode auf Ionengitter. *Z. Phys.*, 77:713–721, 1932.

[42] Elliott H. Lieb. Variational principle for many-fermion systems. *Phys. Rev. Lett.*, 46(7):457–459, 1981.

[43] Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.

[44] Elliott H. Lieb and Barry Simon. The Thomas-Fermi theory of atoms, molecules and solids. *Adv. Math.*, 23:22–116, 1977.

[45] Elliott H. Lieb and Walter E. Thirring. Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In Elliott H. Lieb, Barry Simon, and Arthur S. Wightman, editors, *Studies in Mathematical Physics: Essays in Honor of Valentine Bargmann*. Princeton University Press, Princeton, 1976.

[46] Elliott H. Lieb and Horng-Tzer Yau. The stability and instability of relativistic matter. *Comm. Math. Phys.*, 118:177–213, 1988.

[47] Paul Mancas, A. M. Klaus Müller, and Heinz Siedentop. The optimal size of the exchange hole and reduction to one-particle Hamiltonians. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 111(1):49–53, February 2004.

[48] Albert Messiah. *Mécanique Quantique*, volume 1. Dunod, Paris, 2 edition, 1969.

[49] Isaac Newton. *Philosophiae naturalis principia mathematica. Vol. I.* Harvard University Press, Cambridge, Mass., 1972. Reprinting of the third edition (1726) with variant readings, Assembled and edited by Alexandre Koyré and I. Bernard Cohen with the assistance of Anne Whitman.

[50] Thomas Østergaard Sørensen. *Towards a Relativistic Scott Correction.* PhD thesis, Aarhus Universitet, http://www.imf.au.dk/cgi-bin/viewers/viewpublications.cgi?id=79, October 1998.

[51] Thomas Østergaard Sørensen. The large-Z behavior of pseudorelativistic atoms. *J. Math. Phys.*, 46(5):052307, 24, 2005.

[52] Julian Schwinger. Thomas-Fermi model: The leading correction. *Phys. Rev. A*, 22(5):1827–1832, 1980.

[53] Julian Schwinger. Thomas-Fermi model: The second correction. *Phys. Rev. A*, 24(5):2353–2361, 1981.

[54] J. M. C. Scott. The binding energy of the Thomas-Fermi atom. *Phil. Mag.*, 43:859–867, 1952.

[55] Heinz Siedentop and Rudi Weikard. On the leading energy correction for the statistical model of the atom: Interacting case. *Comm. Math. Phys.*, 112:471–490, 1987.

[56] Heinz Siedentop and Rudi Weikard. Upper bound on the ground state energy of atoms that proves Scott's conjecture. *Phys. Lett. A*, 120:341–342, 1987.

[57] Heinz Siedentop and Rudi Weikard. On the leading energy correction of the statistical atom: Lower bound. *Europhysics Letters*, 6:189–192, 1988.

[58] Heinz Siedentop and Rudi Weikard. On the leading correction of the Thomas-Fermi model: Lower bound – with an appendix by A. M. K. Müller. *Invent. Math.*, 97:159–193, 1989.

[59] Heinz Siedentop and Rudi Weikard. A new phase space localization technique with application to the sum of negative eigenvalues of Schrödinger operators. *Annales Scientifiques de l'École Normale Supérieure*, 24(2):215–225, 1991.

[60] Heinz K. H. Siedentop. On the relation between the Hellmann energy functional and the ground state energy of an N-Fermion system. *Z. Phys. A*, 302:213–218, 1981.

[61] Heinz K. H. Siedentop and Rudi Weikard. On the leading energy correction for the statistical model of the atom: Non-interacting case. *Abh. Braunschweig. Wiss. Ges.*, 38:145–158, 1986.

[62] B. Simon. Fifteen problems in mathematical physics. In *Perspectives in Mathematics*. Birkhäuser, 1984.

[63] Barry Simon. *Functional Integration and Quantum Physics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979.

[64] Jan Philip Solovej. The relativistic Scott correction. In Maria J. Esteban, Claude Le Bris, and Gustavo Scuseria, editors, *Mathematical and Numerical Aspects of Quantum Chemistry Problems*, volume 47/2006 of *Oberwolfach Report*, pages 52–53. Mathematisches Forschungsinstitut Oberwolfach, European Mathematical Society, September 2006.

[65] Jan Philip Solovej, Thomas Østergaard Sørensen, and Wolfgang L. Spitzer. The relativistic Scott correction for atoms and molecules, *Commun. Pure Appl. Math.*, 63:39–118, 2010.

[66] Jan Philip Solovej and Wolfgang L. Spitzer. A new coherent states approach to semiclassics which gives Scott's correction. *Comm. Math. Phys.*, 241(2-3):383–420, 2003.

[67] Irene A. Stegun. Legendre functions. In Milton Abramowitz and Irene A. Stegun, editors, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, chapter 8, pages 331–353. Dover Publications, New York, 1965.

[68] Elias M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[69] J. Sucher. Foundations of the relativistic theory of many-electron atoms. *Phys. Rev. A*, 22(2):348–362, August 1980.

[70] J. Sucher. Foundations of the relativistic theory of many-electron bound states. *International Journal of Quantum Chemistry*, 25:3–21, 1984.

[71] J. Sucher. Relativistic many-electron Hamiltonians. *Phys. Scripta*, 36:271–281, 1987.

[72] Walter Thirring. *Lehrbuch der Mathematischen Physik 3: Quantenmechanik von Atomen und Molekülen.* Springer-Verlag, Wien, New York, 1 edition, 1979.

[73] L. H. Thomas. The calculation of atomic fields. *Proc. Camb. Phil. Soc.*, 23:542–548, 1927.

[74] C. Tix. Lower bound for the ground state energy of the no-pair Hamiltonian. *Phys. Lett. B*, 405(3-4):293–296, 1997.

[75] C. Tix. Self-adjointness and spectral properties of a pseudo-relativistic Hamiltonian due to Brown and Ravenhall. *Preprint*, mp-arc: 97-441, 1997.

[76] C. Tix. Strict positivity of a relativistic Hamiltonian due to Brown and Ravenhall. *Bull. London Math. Soc.*, 30(3):283–290, 1998.

[77] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis; An Introduction to the General Theory of Infinite Processes and of Analytic Functions, with an Account of the Principal Transcendental Functions.* Cambridge University Press, Cambridge, 4 edition, 1927.

Rupert L. Frank
Dept. of Mathematics,
Princeton University,
Princeton, NJ 08544-1000,
USA
rlfrank@math.princeton.edu

Heinz Siedentop
Mathematisches Institut
LMU München
Theresienstraße 39
80333 München, Germany
h.s@lmu.de

Simone Warzel
TU München
Zentrum Mathematik - M7
Boltzmannstraße 3
85747 Garching, Germany
warzel@ma.tum.de

Function Fields of One Variable over PAC Fields

Moshe Jarden and Florian Pop

Received: July 8, 2009

Communicated by Peter Schneider

Abstract. We give evidence for a conjecture of Serre and a conjecture of Bogomolov.

2000 Mathematics Subject Classification: 12E30, 20G15
Keywords and Phrases: Function fields of one variable, PAC fields, Conjecture II of Serre,

Conjecture II of Serre considers a field F of characteristic p with $\mathrm{cd}(\mathrm{Gal}(F)) \leq 2$ such that either $p = 0$ or $p > 0$ and $[F : F^p] \leq p$ and predicts that $H^1(\mathrm{Gal}(F), G) = 1$ (i.e. each principal homogeneous G-spaces has an F-rational point) for each simply connected semi-simple linear algebraic group G [Ser97, p. 139].
As Serre notes, the hypothesis of the conjecture holds in the case where F is a field of transcendence degree 1 over a perfect field K with $\mathrm{cd}(\mathrm{Gal}(K)) \leq 1$. Indeed, in this case $\mathrm{cd}(\mathrm{Gal}(F)) \leq 2$ [Ser97, p. 83, Prop. 11] and $[F : F^p] \leq p$ if $p > 0$ (by the theory of p-bases [FrJ08, Lemma 2.7.2]). We prove the conjecture for F in the special case, where K is PAC of characteristic 0 that contains all roots of unity.
One of the main ingredients of the proof is the projectivity of $\mathrm{Gal}(K(x)_{\mathrm{ab}})$ (where x is transcendental over K and $K(x)_{\mathrm{ab}}$ is the maximal Abelian extension of $K(x)$). We also use the same ingredient to establish an analog to the wellknown open problem of Shafarevich that $\mathrm{Gal}(\mathbb{Q}_{\mathrm{ab}})$ is free. Under the assumption that K is PAC and contains all roots of unity we prove that $\mathrm{Gal}(K(x)_{\mathrm{ab}})$ is not only projective but even free. This proves a stronger version of a conjecture of Bogomolov for a function field of one variable F over a PAC field that contains all roots of unity [Pos05, Conjecture 1.1].

1. The Projectivity of $\mathrm{Gal}(K(x)_{\mathrm{ab}})$

We denote the separable (resp. algebraic) closure of a field K by K_s (resp. $\tilde{K}$) and its absolute Galois group by $\mathrm{Gal}(K)$. The field K is said to be PAC if every absolutely irreducible variety defined over K has a K-rational point. The proof of the projectivity result applies a local-global principle for Brauer groups to reduce the statement to Henselian fields.

For a prime number p and an Abelian group A, we say that A is p'-divisible, if for each $a \in A$ and every positive integer n with $p \nmid n$ there exists $b \in A$ such that $a = nb$. Note that if $p = 0$, then "p'-divisible" is the same as "divisible".

Lemma 1.1: *Let p be 0 or a prime number, B a torsion free Abelian group, and A is a p'-divisible subgroup of B of finite index. Then B is also p'-divisible.*

Proof: First suppose $p = 0$ and let $m = (B : A)$. Then, for each $b \in B$ and a positive integer n there exists $a \in A$ such that $mb = mna$. Since B is torsion free, $m = na$. Thus, B is divisible.

Now suppose p is a prime number, let $mp^k = (B : A)$, with $p \nmid m$ and $k \geq 0$, and consider $b \in B$. Then $mp^k b \in A$. Hence, for each positive integer n with $p \nmid n$ there exists $a \in A$ with $mp^k b = mna$. Thus, $p^k b = na$. Since $p \nmid n$, there exist $x, y \in \mathbb{Z}$ such that $xp^k + yn = 1$. It follows from $xp^k b = xna$ that $b = n(xa + yb)$, as claimed. □

Corollary 1.2: *Let L/K be an algebraic field extension, v a valuation of L, and $p = 0$ or p is a prime number. Suppose that $v(K^\times)$ is p'-divisible. Then $v(L^\times)$ is p'-divisible.*

Proof: Let $x \in L^\times$ and n a positive integer with $p \nmid n$. Then $v(K(x)^\times)$ is a torsion free Abelian group and $v(K^\times)$ is a subgroup of index at most $[K(x) : K]$. Since $v(K^\times)$ is p'-divisible, Lemma 1.1 gives $y \in K(x)^\times$ such that $v(x) = nv(y)$. It follows that $v(L^\times)$ is p'-divisible. □

Given a Henselian valued field (M, v) we use v also for its unique extension to M_s. We use a bar to denote the residue with respect to v of objects associated with M, let O_M be the valuation ring of M, and let $\Gamma_M = v(M^\times)$ be the value group of M.

We write $\mathrm{cd}_l(K)$ and $\mathrm{cd}(K)$ for the lth cohomological dimension and the cohomological dimension of $\mathrm{Gal}(K)$ and note that $\mathrm{cd}(K) \leq 1$ if and only if $\mathrm{Gal}(K)$ is projective [Ser97, p. 58, Cor. 2].

Lemma 1.3: *Let (M, v) be a Henselian valued field. Suppose $p = \mathrm{char}(M) = \mathrm{char}(\bar{M})$, $\mathrm{Gal}(\bar{M})$ is projective, and Γ_M is p'-divisible. Then $\mathrm{Gal}(M)$ is projective.*

Proof: We denote the inertia field of M by M_u. It is determined by its absolute Galois group: $\mathrm{Gal}(M_u) = \{\sigma \in \mathrm{Gal}(M) \mid v(\sigma x - x) > 0 \text{ for all } x \in M_s \text{ with } v(x) \geq 0\}$. The map $\sigma \mapsto \bar{\sigma}$ of $\mathrm{Gal}(M)$ into $\mathrm{Gal}(\bar{M})$ such that $\bar{\sigma}\bar{x} = \overline{\sigma x}$ for each $x \in O_M$ is a well defined epimorphism [Efr06, Thm. 16.1.1] whose kernel is $\mathrm{Gal}(M_u)$. It therefore defines an isomorphism

$$\mathrm{Gal}(M_u/M) \cong \mathrm{Gal}(\bar{M}). \tag{1}$$

CLAIM A: *$\bar{M}_u$ is separably closed.* Let $g \in \bar{M}[X]$ be a monic irreducible separable polynomial of degree $n \geq 1$. Then there exists a monic polynomial $f \in O_{M_u}[X]$ of degree n such that $\bar{f} = g$. We observe that f is also irreducible and separable. Moreover, if $f(X) = \prod_{i=1}^n (X - x_i)$ with $x_1, \ldots, x_n \in M_s$, then $g(X) = \prod_{i=1}^n (X - \bar{x}_i)$. Given $1 \leq i, j \leq n$ there exists $\sigma \in \mathrm{Gal}(M_u)$ such that $\sigma x_i = x_j$. By definition, $\bar{x}_j = \overline{\sigma x_i} = \bar{\sigma}\bar{x}_i = \bar{x}_i$. Since g is separable, $i = j$, so $n = 1$. We conclude that $\bar{M}_u$ is separably closed.

CLAIM B: *Each l-Sylow group of $\mathrm{Gal}(M_u)$ with $l \neq p$ is trivial.* Indeed, let L be the fixed field of an l-Sylow group of $\mathrm{Gal}(M_u)$ in M_s. If $l = 2$, then $\zeta_l = -1 \in L$. If $l \neq 2$, then $[L(\zeta_l) : L] | l - 1$ and $[L(\zeta_l) : L]$ is a power of l, so $\zeta_l \in L$.

Assume that $\mathrm{Gal}(L) \neq 1$. By the the theory of finite l-groups, L has a cyclic extension L' of degree l. By the preceding paragraph and Kummer theory, there exists $a \in L^\times$ such that $L' = L(\sqrt[l]{a})$. By Corollary 1.2, there exists $b \in L^\times$ such that $lv(b) = v(a)$. Then $c = \frac{a}{b^l}$ satisfies $v(c) = 0$. By Claim A, $\bar{L}$ is separably closed. Therefore, $\bar{c}$ has an lth root in $\bar{L}$. By Hensel's lemma, c has an lth root in L. It follows that a has an l-root in L. This contradiction implies that $L = M_s$, as claimed.

Having proved Claim B, we consider again a prime number $l \neq p$ and let G_l be an l-Sylow subgroup of $\mathrm{Gal}(M)$. By the Claim, $G_l \cap \mathrm{Gal}(M_u) = 1$, hence the map res: $\mathrm{Gal}(M) \to \mathrm{Gal}(M_u/M)$ maps G_l isomorphically onto an l-Sylow subgroup of $\mathrm{Gal}(M_u/M)$. By (1), G_l is isomorphic to an l-Sylow subgroup of $\mathrm{Gal}(\bar{M})$. Since the latter group is projective, so is G_l, i.e. $\mathrm{cd}_l(G) \leq 1$ [Ser97, p. 58, Cor. 2].

Finally, if $p \neq 0$, then $\mathrm{cd}_p(M) \leq 1$ [Ser97, p. 75, Prop. 3], because then $\mathrm{char}(M) = p$. It follows that $\mathrm{cd}(M) \leq 1$ [Ser97, p. 58, Cor. 2]. □

LEMMA 1.4: *Let F be an extension of a PAC field K of transcendence degree 1 and characteristic p. Suppose $v(F^\times)$ is p'-divisible for each valuation v of F/K. Then $\mathrm{Gal}(F)$ is projective.*

Proof: Let K_{ins} be the maximal purely inseparable algebraic extension of K and set $F' = FK_{\mathrm{ins}}$. Then K_{ins} is PAC [FrJ08, Cor. 11.2.5], $\mathrm{trans.deg}(F'/K_{\mathrm{ins}}) = 1$, and $v((F')^\times)$ is p'-divisible for every valuation v of F' (by Corollary 1.2). Moreover, $\mathrm{Gal}(F') = \mathrm{Gal}(F)$. Thus, we may replace K by K_{ins} and F by F', if necessary, to assume that K is perfect.

Let $V(F/K)$ be a system of representatives of the equivalence classes of valuations of F that are trivial on K. For each $v \in V(F/K)$ we choose a Henselian closure F_v of F at v. By [Efr01, Thm. 3.4], there is an injection of Brauer groups,

$$\mathrm{Br}(F) \to \prod_{v \in V(F/K)} \mathrm{Br}(F_v). \tag{2}$$

For each $v \in V(F/K)$ we have, $v(F_v^\times) = v(F^\times)$ is p'-divisible. Also, the residue field $\bar{F}_v$ is an algebraic extension of K. Since K is PAC, a theorem of Ax says that $\mathrm{Gal}(K)$ is projective [FrJ08, Thm. 11.6.2], hence $\mathrm{Gal}(\bar{F}_v)$ is projective [FrJ08, Prop. 22.4.7]. Finally, $\mathrm{char}(F_v) = \mathrm{char}(\bar{F}_v)$. Therefore, by Lemma 1.3, $\mathrm{Gal}(F_v)$ is projective, hence $\mathrm{Br}(F_v) = 0$ [Ser97, p. 78, Prop. 5]. It follows from the injectivity of (2) that $\mathrm{Br}(F) = 0$.

If F_1 is a finite separable extension of F, $v_1 \in V(F_1/K)$, and $v = v_1|_F$, then $v(F^\times)$ is p'-divisible. Hence, by Corollary 1.2, $v_1((F_1)^\times)$ is p'-divisible. It follows from the preceding paragraph that $\mathrm{Br}(F_1) = 0$. Consequently, by [Ser97, p. 78, Prop. 5], $\mathrm{cd}(\mathrm{Gal}(F)) \le 1$. □

LEMMA 1.5: *Let p be either 0 or a prime number and let Γ be an additive subgroup of $\mathbb{Q}$. Suppose $\frac{1}{n} \in \Gamma$ for each positive integer n with $p \nmid n$. Then Γ is p'-divisible.*

Proof: We consider $\gamma \in \Gamma$. If $p = 0$, we write $\gamma = \frac{a}{b}$, with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Given $n \in \mathbb{N}$, we have $\frac{\gamma}{n} = a \cdot \frac{1}{nb} \in \Gamma$.

If $p > 0$, we write $\gamma = \frac{a}{bp^k}$, where $a \in \mathbb{Z}$, $b \in \mathbb{N}$, $k \in \mathbb{Z}$, and $p \nmid a, b$. Let $n \in \mathbb{N}$ with $p \nmid n$. If $k \le 0$, then $\frac{\gamma}{n} = ap^{-k} \cdot \frac{1}{nb} \in \Gamma$. If $k > 0$, we may choose $x, y \in \mathbb{Z}$ such that $xp^k + ynb = 1$. Then $\frac{\gamma}{n} = \frac{a}{nbp^k} = \frac{axp^k + aynb}{nbp^k} = ax \cdot \frac{1}{nb} + by \cdot \frac{a}{bp^k} \in \Gamma$, as claimed. □

PROPOSITION 1.6: *Let K be a PAC field that contains all roots of unity and let E be an extension of K of transcendence degree 1. Then $\mathrm{Gal}(E_{\mathrm{ab}})$ is projective.*

Proof: First we consider the case where $E = K(x)$, where x is transcendental over K, and set $F = E_{\mathrm{ab}}$. In the notation of Lemma 1.4 we consider a valuation $v \in V(F/K)$ normalized in such a way that $v(E^\times) = \mathbb{Z}$. Then $v(F^\times) \le \mathbb{Q}$. On the other hand, let $p = \mathrm{char}(K)$ and consider a positive integer n with $p \nmid n$. Let $e \in E$ with $v(e) = 1$. Then $e^{1/n} \in F$ (because K contains a root of 1 of order n). Therefore, $\frac{1}{n} = v(e^{1/n}) \in v(F^\times)$. By Lemma 1.5, $v(F^\times)$ is p'-divisible. We conclude from Lemma 1.4 that $\mathrm{Gal}(F)$ is projective.

In the general case we choose $x \in E$ transcendental over K. By the preceding paragraph, $\mathrm{Gal}(K(x)_{\mathrm{ab}})$ is projective. Since taking purely inseparable extensions of a field does not change its absolute Galois group, $\mathrm{Gal}(K(x)_{\mathrm{ab,ins}})$ is projective. Now note that $\mathrm{Gal}(E_{\mathrm{ab,ins}})$ as a subgroup of $\mathrm{Gal}(K(x)_{\mathrm{ab,ins}})$ is also projective. Hence, $\mathrm{Gal}(E_{\mathrm{ab}})$ is projective. □

Remark 1.7: Proposition 1.6 is false if K does not contain all roots of unity. Indeed, the authors will elsewhere provide an example of a prime number l and a PAC field K of characteristic 0 that contains all roots of unity of order n with $l \nmid n$ but not ζ_l such that $\mathrm{Gal}(K(x)_{\mathrm{ab}})$ is not projective. □

2. SERRE AND SHAFAREVICH

We refer to a simply connected semi-simple linear algebraic group G as a SIMPLY CONNECTED GROUP. In this case $H^1(\mathrm{Gal}(K), G)$ will be also denoted by $H^1(K, G)$. Since each element of $H^1(K, G)$ is represented by a principal homogeneous space V of G and V is an absolutely irreducible variety defined over

K, V has a K-rational point if K is PAC. Hence, V is equivalent to G [LaT58, Prop. 4]. Thus, $H^1(K,G) = 1$.

The proof of Serre's Conjecture II in our case is based on the following consequence of a theorem of Colliot-Thélène, Gille, and Parimala:

Proposition 2.1: *Let F be a field and G a simply connected group defined over F. Suppose F is a C_2-field of characteristic 0, $\mathrm{cd}(F) \le 2$, and $\mathrm{cd}(F_{\mathrm{ab}}) \le 1$. Then $H^1(F,G) = 1$.*

Proof: Let F' be a finite extension of F. Since F is C_2, [CGP04, Thm. 1.1(vi)] implies that if the exponent e of a central simple algebra A over F' is a power of 2 or a power of 3, then e is equal to the index of A.

Since $\mathrm{cd}(F) \le 2$ and $\mathrm{cd}(F_{\mathrm{ab}}) \le 1$, [CGP04, Thm. 1.2(v)] implies that $H^1(F,G) = 1$. □

Remark 2.2: By Merkuriev-Suslin, the assumption that F is a C_2-field implies that $\mathrm{cd}(F) \le 2$ [Ser97, end of page 88]. However, we will be able to prove both properties of F directly in the application we have in mind. □

The following result establishes the first condition on F.

Lemma 2.3: *Let F be an extension of transcendence degree 1 over a perfect PAC field K. Suppose either $\mathrm{char}(K) > 0$ and K contains all roots of unity or $\mathrm{char}(K) = 0$. Then $\mathrm{cd}(F) \le 2$ and F is a C_2-field.*

Proof: By Ax, $\mathrm{cd}(K) \le 1$ [FrJ08, Thm. 11.6.2]. Hence, by [Ser97, p. 83, Prop. 11], $\mathrm{cd}(F) \le 2$.

A conjecture of Ax from 1968 says that every perfect PAC field K is C_1 [FrJ08, Problem 21.2.5]. The conjecture holds if K contains an algebraically closed field [FrJ08, Lemma 21.3.6(a)]. In particular, if $p = \mathrm{char}(K) > 0$ and K contains all roots of unity, then $\tilde{\mathbb{F}}_p \subseteq K$, so K is C_1. If $\mathrm{char}(K) = 0$, K is C_1, by [Kol07, Thm. 1]. It follows that in each case, F is C_2 [FrJ08, Prop. 21.2.12]. □

Theorem 2.4: *Let F be an extension of transcendence degree 1 of a PAC field K of characteristic 0. Suppose K contains all roots of unity. Then F satisfies Serre's conjecture II. That is, $H^1(F,G) = 1$ for each simply connected group G defined over F.*

Proof: By Lemma 2.3, $\mathrm{cd}(F) \le 2$ and F is a C_2-field. By Proposition 1.6, $\mathrm{cd}(F_{\mathrm{ab}}) \le 1$. It follows from Proposition 2.1 that $H^1(F,G) = 1$ for each simply connected group G. □

Remark 2.5: All of the ingredients of the proof of Theorem 2.4 except possibly Proposition 2.1 work also when $\mathrm{char}(K) > 0$. □

The proof of the freeness of $\mathrm{Gal}(K(x)_{\mathrm{ab}})$ applies the notion of "quasi-freeness" due to Harbater and Stevenson. To this end recall that a finite split embedding problem $\mathcal{E}$ for a profinite group G is a pair (φ: $G \to A$, α: $B \to A$), where A, B are finite groups, φ, α are epimorphisms, and α has a group theoretic section. A solution of $\mathcal{E}$ is an epimorphism γ: $G \to B$ such that

$\alpha \circ \gamma = \varphi$. We say that G is QUASI-FREE if its rank m is infinite and every finite split embedding problem for G has m distinct solutions.

THEOREM 2.6: *Let F be a function field of one variable over a PAC field K of cardinality m containing all roots of unity and let x be a variable. Then* $\mathrm{Gal}(F_{\mathrm{ab}})$ *is isomorphic to the free profinite group of rank m.*

Proof: Since K is PAC, K is AMPLE, that is every absolutely irreducible curve defined over K with a K-rational simple point has infinitely many K-rational points. By [HaS05, Cor. 4.4], $\mathrm{Gal}(F)$ is quasi-free of rank $m = \mathrm{card}(K)$. Hence, by [Har09, Thm. 2.4], $\mathrm{Gal}(F_{\mathrm{ab}})$ is also quasi-free of rank m. Since by Proposition 1.6, $\mathrm{Gal}(F_{\mathrm{ab}})$ is projective, it follows from a result of Chatzidakis and Melnikov [FrJ08, Lemma 25.1.8] that $\mathrm{Gal}(F_{\mathrm{ab}})$ is free of rank m. □

Acknowledgment: The authors thank Jean-Louis Colliot-Thélène for stimulating discussions, in particular for pointing out Proposition 2.1 to them. □

REFERENCES

[CGP04] J.-L. Colliot-Thélène, P. Gille, and R. Parimala, *Arithmetic of linear algebraic groups over two-dimensional geometric fields,* Duke Mathematical Journal 121 (2004), 285–341.

[Efr01] I. Efrat, *A Hasse principle for function fields over PAC fields,* Israel Journal of Mathematics 122 (2001), 43–60.

[Efr06] I. Efrat, *Valuations, Orderings, and Milnor K-Theory,* Mathematical surveys and monographs 124, American Mathematical Society, Providence, 2006.

[FrJ08] M. D. Fried and M. Jarden, *Field Arithmetic, Third Edition, revised by Moshe Jarden,* Ergebnisse der Mathematik (3) 11, Springer, Heidelberg, 2008.

[Har09] D. Harbater, *On function fields with free absolute Galois groups,* Journal für die reine und andgewandte Mathematik 632 (2009), 85–103.

[HaS05] D. Harbater and K. Stevenson, *Local Galois theory in dimension two,* Advances in Mathematics 198 (2005), 623–653.

[Kol07] J. Kollár, *A conjecture of Ax and degenerations of Fano varieties,* Israel Journal of Mathematics 162 (2007), 235–251.

[LaT58] S. Lang and J. Tate, *Principal homogeneous spaces over abelian varieties,* American Journal of Mathematics 80 (1958), 659–684.

[Pos05] L. Positselski, *Koszul property and Bogomolov's conjecture,* International Mathematics Research Notices 31 (2005), 1901–1936.

[Ser97] J.-P. Serre, *Galois Cohomology,* Springer, Berlin, 1997.

Moshe Jarden
School of Mathematics
Tel Aviv University
Ramat Aviv, Tel Aviv 69978, Israel
jarden@post.tau.ac.il

Florian Pop
Department of Mathematics
University of Pennsylvania
Philadelphia, Pennsylvania 19104, USA
pop@math.upenn.edu

Geometric Description of the Connecting Homomorphism for Witt Groups

Paul Balmer[1] and Baptiste Calmès[2]

Received: December 10, 2008
Revised: October 23, 2009

Communicated by Ulf Rehmann

Abstract. We give a geometric setup in which the connecting homomorphism in the localization long exact sequence for Witt groups decomposes as the pull-back to the exceptional fiber of a suitable blow-up followed by a push-forward.

2000 Mathematics Subject Classification: Primary 19G12; Secondary 11E81.
Keywords and Phrases: Witt group, localization sequence, connecting homomorphism, blow-up, push-forward, non-oriented theory.

Contents

[1]Research supported by SNSF grant PP002-112579 and NSF grant 0654397
[2]Research supported by EPSRC grant EP/E01786X/1

1 Introduction

Witt groups form a very interesting cohomology theory in algebraic geometry. (For a survey, see [5].) Unlike the better known K-theory and Chow theory, Witt theory is not *oriented* in the sense of Levine-Morel [17] or Panin [22], as already visible on the non-standard projective bundle theorem, see Arason [2] and Walter [26]. Another way of expressing this is that push-forwards do not exist in sufficient generality for Witt groups. This "non-orientability" can make computations unexpectedly tricky. Indeed, the Witt groups of such elementary schemes as Grassmann varieties will appear for the first time in the companion article [6], whereas the corresponding computations for oriented cohomologies have been achieved more than 35 years ago in [16], using the well-known cellular decomposition of Grassmann varieties. See also [21] for general statements on cellular varieties.
In oriented theories, there is a very useful computational technique, recalled in Theorem 1.3 below, which allows inductive computations for families of cellular varieties. Our paper originates in an attempt to extend this result to the non-oriented setting of Witt theory. Roughly speaking, such an extension is possible "half of the time". In the remaining "half", some specific ideas must come in and reflect the truly non-oriented behavior of Witt groups. To explain this rough statement, let us fix the setup, which will remain valid for the entire paper.

Setup 1.1. We denote by $\mathcal{S}ch$ the category of separated connected noetherian $\mathbb{Z}[\frac{1}{2}]$-scheme. Let $X, Z \in \mathcal{S}ch$ be schemes and let $\iota : Z \hookrightarrow X$ be a regular closed immersion of codimension $c \geq 2$. Let $Bl = Bl_Z X$ be the blow-up of X along Z and E the exceptional fiber. Let $U = X - Z \cong Bl - E$ be the unaltered open complement. We have a commutative diagram

$$\begin{tikzcd} Z \arrow[r, hook, "\iota"] & X & U \arrow[l, hook', "\upsilon"'] \arrow[dl, hook', "\tilde{\upsilon}"] \\ E \arrow[u, "\tilde{\pi}"] \arrow[r, hook, "\tilde{\iota}"'] & Bl \arrow[u, "\pi"'] & \end{tikzcd} \tag{1}$$

with the usual morphisms.

Consider now a cohomology theory with supports, say H^*

$$\cdots \xrightarrow{\partial} H^*_Z(X) \longrightarrow H^*(X) \xrightarrow{\upsilon^*} H^*(U) \xrightarrow{\partial} H^{*+1}_Z(X) \longrightarrow \cdots \tag{2}$$

In this paper we shall focus on the case of Witt groups $H^* = \mathrm{W}^*$ but we take inspiration from H^* being an oriented cohomology theory. Ideally, we would like conditions for the vanishing of the connecting homomorphism $\partial = 0$ in the above localization long exact sequence. Even better would be conditions for the restriction υ^* to be split surjective. When H^* is an oriented theory, there is a well-known hypothesis under which such a splitting actually exists, namely:

Hypothesis 1.2. Assume that there exists an auxiliary morphism $\tilde{\alpha} : Bl \to Y$

$$\begin{tikzcd} Z \arrow[r, hook, "\iota"] & X & U \arrow[l, hook', "\upsilon"'] \arrow[dl, "\tilde{\upsilon}"'] \arrow[d, dotted, "\alpha"] \\ E \arrow[u, "\tilde{\pi}"] \arrow[r, hook, "\tilde{\iota}"'] & Bl \arrow[u, "\pi"] \arrow[r, dotted, "\tilde{\alpha}"'] & Y \end{tikzcd} \tag{3}$$

such that $\alpha := \tilde{\alpha} \circ \tilde{\upsilon} : U \to Y$ is an $\mathbb{A}^*$-*bundle*, i.e. every point of Y has a Zariski neighborhood over which α is isomorphic to a trivial $\mathbb{A}^r$-bundle, for some $r \geq 0$. See Ex. 1.5 for an explicit example with X, Y and Z being Grassmann varieties.

Theorem 1.3 (The oriented technique). *Under Setup 1.1 and Hypothesis 1.2, assume X, Y and Z regular. Assume the cohomology theory H^* is homotopy invariant for regular schemes and* oriented*, in that it admits push-forwards along proper morphisms satisfying flat base-change. Then, the restriction $\upsilon^* : H^*(X) \to H^*(U)$ is split surjective with explicit section $\pi_* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1}$, where $\pi_* : H^*(Bl) \to H^*(X)$ is the push-forward. Hence the connecting homomorphism $\partial : H^*(U) \to H_Z^{*+1}(X)$ vanishes and the above localization long exact sequence* (2) *reduces to split short exact sequences $0 \to H_Z^*(X) \to H^*(X) \to H^*(U) \to 0$.*

Proof. By homotopy invariance, we have $\alpha^* : H^*(Y) \xrightarrow{\sim} H^*(U)$. By base-change, $\upsilon^* \circ \pi_* = \tilde{\upsilon}^*$ and since $\tilde{\upsilon}^* \circ \tilde{\alpha}^* = \alpha^*$, we have $\upsilon^* \circ \pi_* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1} = \mathrm{id}$. □

The dichotomy between the cases where the above technique extends to Witt groups and the cases where is does not, comes from the duality. To understand this, recall that one can consider Witt groups $\mathrm{W}^*(X, L)$ with duality twisted by a line bundle L on the scheme X. Actually only the class of the twist L in $\mathrm{Pic}(X)/2$ really matters since we have square-periodicity isomorphisms for all $M \in \mathrm{Pic}(X)$

$$\mathrm{W}^*(X, L) \cong \mathrm{W}^*(X, L \otimes M^{\otimes 2}) . \tag{4}$$

Here is a condensed form of our Theorem 2.3 and Main Theorem 2.6 below:

Theorem 1.4. *Under Hypothesis 1.2, assume X, Y and Z regular. Let $L \in \mathrm{Pic}(X)$. Then there exists an integer $\lambda(L) \in \mathbb{Z}$ (defined by* (8) *below) such that:*

(A) If $\lambda(L) \equiv c - 1 \mod 2$ then the restriction $\upsilon^* : \mathrm{W}^*(X, L) \to \mathrm{W}^*(U, L_{|_U})$ is split surjective with a section given by the composition $\pi_* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1}$. Hence the connecting homomorphism $\mathrm{W}^*(U, L_{|_U}) \xrightarrow{\partial} \mathrm{W}_Z^{*+1}(X, L)$ vanishes and the localization long exact sequence reduces to split short exact sequences

$$0 \longrightarrow \mathrm{W}_Z^*(X, L) \longrightarrow \mathrm{W}^*(X, L) \longrightarrow \mathrm{W}^*(U, L_{|_U}) \longrightarrow 0 .$$

(B) If $\lambda(L) \equiv c \mod 2$ then the connecting homomorphism ∂ is equal to a composition of pull-backs and push-forwards: $\partial = \iota_* \circ \tilde{\pi}_* \circ \tilde{\iota}^* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1}$.

This statement requires some explanations. First of all, note that we have used push-forwards for Witt groups, along $\pi : Bl \to X$ in (A) and along $\tilde{\pi} : E \to Z$ and $\iota : Z \to X$ in (B). To explain this, recall that the push-forward in Witt theory is only conditionally defined. Indeed, given a proper morphism $f : X' \to X$ between (connected) regular schemes and given a line bundle $L \in \mathrm{Pic}(X)$, the push-forward homomorphism does not map $\mathrm{W}^*(X', f^*L)$ into $\mathrm{W}^*(X, L)$, as one could naively expect, but the second author and Hornbostel [8] showed that Grothendieck-Verdier duality yields a twist by the relative canonical line bundle $\omega_f \in \mathrm{Pic}(X')$:

$$\mathrm{W}^{i+\dim(f)}\big(\, X' \,,\, \omega_f \otimes f^*L \,\big) \xrightarrow{f_*} \mathrm{W}^i(\, X \,,\, L \,)\,. \tag{5}$$

Also note the shift by the relative dimension, $\dim(f) := \dim X' - \dim X$, which is not problematic, since we can always replace $i \in \mathbb{Z}$ by $i - \dim(f)$.

More trickily, if you are given a line bundle $M \in \mathrm{Pic}(X')$ and if you need a push-forward $\mathrm{W}^*(X', M) \to \mathrm{W}^{*-\dim(f)}(X, ?)$ along $f : X' \to X$, you first need to check that M is of the form $\omega_f \otimes f^*L$ for some $L \in \mathrm{Pic}(X)$, at least module squares. *Otherwise, you simply do not know how to push-forward.* This is precisely the source of the dichotomy of Theorem 1.4, as explained in Proposition 2.1 below.

At the end of the day, it is only possible to transpose to Witt groups the oriented technique of Theorem 1.3 when the push-forward π_* exists for Witt groups. But actually, the remarkable part of Theorem 1.4 is Case (B), that is our Main Theorem 2.6 below, which gives a description of the connecting homomorphism ∂ when we cannot prove it zero by the oriented method. This is the part where the non-oriented behavior really appears. See more in Remark 2.7. Main Theorem 2.6 is especially striking since the original definition of the connecting homomorphism given in [3, § 4] does not have such a geometric flavor of pull-backs and push-forwards but rather involves abstract techniques of triangulated categories, like symmetric cones, and the like. Our new geometric description is also remarkably simple to use in applications, see [6]. Here is the example in question.

Example 1.5. Let k be a field of characteristic not 2. (We describe flag varieties over k by giving their k-points, as is customary.) Let $1 \le d \le n$. Fix a codimension one subspace k^{n-1} of k^n. Let $X = \mathrm{Gr}_d(n)$ be the Grassmann variety of d-dimensional subspaces $V_d \subset k^n$ and let $Z \subset X$ be the closed subvariety of those subspaces V_d contained in k^{n-1}. The open complement $U = X - Z$ consists of those $V_d \not\subset k^{n-1}$. For such $V_d \in U$, the subspace $V_d \cap k^{n-1} \subset k^{n-1}$ has dimension $d-1$. This construction defines an $\mathbb{A}^{n-d}$-bundle $\alpha : U \to Y := \mathrm{Gr}_{d-1}(n-1)$, mapping V_d to $V_d \cap k^{n-1}$. This situation relates the Grassmann variety $X = \mathrm{Gr}_d(n)$ to the smaller ones $Z = \mathrm{Gr}_d(n-1)$

and $Y = \mathrm{Gr}_{d-1}(n-1)$. Diagram (1) here becomes

$$\begin{tikzcd}
\mathrm{Gr}_d(n-1) \arrow[r, hook, "\iota"] & \mathrm{Gr}_d(n) & U \arrow[l, hook', "\upsilon"'] \arrow[d, "\alpha"] \arrow[dl, hook', "\tilde{\upsilon}"'] \\
E \arrow[u, "\tilde{\pi}"] \arrow[r, hook, "\tilde{\iota}"'] & Bl \arrow[u, "\pi"] \arrow[r, "\tilde{\alpha}"'] & \mathrm{Gr}_{d-1}(n-1)\,.
\end{tikzcd}$$

The blow-up Bl is the variety of pairs of subspaces $V_{d-1} \subset V_d$ in k^n, such that $V_{d-1} \subset k^{n-1}$. The morphisms $\pi : Bl \to X$ and $\tilde{\alpha} : Bl \to Y$ forget V_{d-1} and V_d respectively. The morphism $\tilde{\upsilon}$ maps $V_d \not\subset k^{n-1}$ to the pair $(V_d \cap k^{n-1}) \subset V_d$. Applying Theorem 1.3 to this situation, Laksov [16] computes the Chow groups of Grassmann varieties by induction. For Witt groups though, there are cases where the restriction $\mathrm{W}^*(X, L) \to \mathrm{W}^*(U, L_{|U})$ is not surjective (see [6, Cor. 6.7]). Nevertheless, thank to our geometric description of the connecting homomorphism, we have obtained a complete description of the Witt groups of Grassmann varieties, for all shifts and all twists, to appear in [6]. In addition to the present techniques, our computations involve other ideas, specific to Grassmann varieties, like Schubert cells and desingularisations thereof, plus some combinatorial bookkeeping by means of special Young diagrams. Including all this here would misleadingly hide the simplicity and generality of the present paper. We therefore chose to publish the computation of the Witt groups of Grassmann varieties separately in [6].

The paper is organized as follows. Section 2 is dedicated to the detailed explanation of the above dichotomy and the proof of the above Case (A), see Theorem 2.3. We also explain Case (B) in our Main Theorem 2.6 but its proof is deferred to Section 5. The whole Section 2 is written, as above, under the assumption that all schemes are regular. This assumption simplifies the statements but can be removed at the price of introducing dualizing complexes and coherent Witt groups, which provide the natural framework over non-regular schemes. This generalization is the purpose of Section 3. There, we even drop the auxiliary Hypothesis 1.2, i.e. the dotted part of Diagram (3). Indeed, our Main Lemma 3.5 gives a very general description of the connecting homomorphism applied to a Witt class over U, if that class comes from the blow-up Bl via restriction $\tilde{\upsilon}^*$. The proof of Main Lemma 3.5 occupies Section 4. Finally, Hypothesis 1.2 re-enters the game in Section 5, where we prove our Main Theorem 2.6 as a corollary of a non-regular generalization given in Theorem 5.1. For the convenience of the reader, we gathered in Appendix A the needed results about Picard groups, canonical bundles and dualizing complexes, which are sometimes difficult to find in the literature. The conscientious reader might want to start with that appendix.

2 The regular case

We keep notation as in Setup 1.1 and we assume all schemes to be regular. This section can also be considered as an expanded introduction.

As explained after Theorem 1.4 above, we have to decide when the push-forward along $\pi : Bl \to X$ and along $\tilde{\pi} : E \to Z$ exist. By (5), we need to determine the canonical line bundles $\omega_\pi \in \mathrm{Pic}(Bl)$ and $\omega_{\tilde{\pi}} \in \mathrm{Pic}(E)$. This is classical and is recalled in Appendix A. First of all, Proposition A.6 gives

$$\mathrm{Pic}\left(\begin{tikzcd} Z \arrow[r, hook, "\iota"] & X & U \arrow[l, hook', "\upsilon"'] \arrow[dl, "\tilde{\upsilon}"] \\ E \arrow[u, "\tilde{\pi}"] \arrow[r, hook, "\tilde{\iota}"'] & Bl \arrow[u, "\pi"'] & \end{tikzcd}\right) \cong \begin{tikzcd} \mathrm{Pic}(Z) \arrow[d, "\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)"'] & \mathrm{Pic}(X) \arrow[l, "\iota^*"'] \arrow[r, equal] \arrow[d, "\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)"'] & \mathrm{Pic}(X)\,. \\ \mathrm{Pic}(Z)\oplus\mathbb{Z} & \mathrm{Pic}(X)\oplus\mathbb{Z} \arrow[l, "\left(\begin{smallmatrix}\iota^* & 0\\0 & 1\end{smallmatrix}\right)"'] \arrow[ur, "(1\ 0)"'] & \end{tikzcd}$$

The $\mathbb{Z}$ summands in $\mathrm{Pic}(Bl)$ and $\mathrm{Pic}(E)$ are generated by $\mathcal{O}(E) = \mathcal{O}_{Bl}(-1)$ and $\mathcal{O}(E)_{|E} = \mathcal{O}_E(-1)$ respectively. Then Proposition A.11 gives the wanted

$$\begin{aligned} \omega_\pi &= (0, c-1) \quad \text{in} \quad \mathrm{Pic}(X)\oplus\mathbb{Z} \cong \mathrm{Pic}(Bl) \quad \text{and} \\ \omega_{\tilde{\pi}} &= (-\omega_\iota, c) \quad \text{in} \quad \mathrm{Pic}(Z)\oplus\mathbb{Z} \cong \mathrm{Pic}(E)\,. \end{aligned} \tag{6}$$

So, statistically, picking a line bundle $M \in \mathrm{Pic}(Bl)$ at random, there is a 50% chance of being able to push-forward $\mathrm{W}^*(Bl, M) \to \mathrm{W}^*(X, L)$ along π for some suitable line bundle $L \in \mathrm{Pic}(X)$. To justify this, observe that

$$\mathrm{coker}\big(\,\mathrm{Pic}(X) \xrightarrow{\pi^*} \mathrm{Pic}(Bl)\,\big)/2 \ \cong\ \mathbb{Z}/2$$

and tensoring by ω_π is a bijection, so half of the elements of $\mathrm{Pic}(Bl)/2$ are of the form $\omega_\pi \otimes \pi^*(L)$. The same probability of 50% applies to the push forward along $\tilde{\pi} : E \to Z$ but interestingly in complementary cases, as we summarize now.

Proposition 2.1. *With the notation of 1.1, assume X and Z regular. Recall that $c = \mathrm{codim}_X(Z)$. Let $M \in \mathrm{Pic}(Bl)$. Let $L \in \mathrm{Pic}(X)$ and $\ell \in \mathbb{Z}$ be such that $M = (L, \ell)$ in $\mathrm{Pic}(Bl) = \mathrm{Pic}(X)\oplus\mathbb{Z}$, that is, $M = \pi^*L \otimes \mathcal{O}(E)^{\otimes \ell}$.*

(A) If $\ell \equiv c-1 \mod 2$, we can push-forward along $\pi : Bl \to X$, as follows:

$$\mathrm{W}^*(Bl, M) \cong \mathrm{W}^*(Bl, \omega_\pi \otimes \pi^* L) \xrightarrow{\pi_*} \mathrm{W}^*(X, L)\,.$$

(B) If $\ell \equiv c \mod 2$, we can push-forward along $\tilde{\pi} : E \to Z$, as follows:

$$\mathrm{W}^*(E, M_{|E}) \cong \mathrm{W}^*\big(E, \omega_{\tilde{\pi}} \otimes \tilde{\pi}^*(\omega_\iota \otimes L_{|Z})\big) \xrightarrow{\tilde{\pi}_*} \mathrm{W}^{*-c+1}(Z, \omega_\iota \otimes L_{|Z}).$$

In each case, the isomorphism $\cong$ comes from square-periodicity in the twist (4) *and the subsequent homomorphism is the push-forward* (5).

Proof. We only have to check the congruences in $\mathrm{Pic}/2$. By (6), when $\ell \equiv c-1 \mod 2$, we have $[\omega_\pi \otimes \pi^* L] = [(L, \ell)] = [M]$ in $\mathrm{Pic}(Bl)/2$. When $\ell \equiv c \mod 2$, we have $[\omega_{\tilde{\pi}} \otimes \tilde{\pi}^*(\omega_\iota \otimes L_{|Z})] = [(L_{|Z}, \ell)] = [M_{|E}]$ in $\mathrm{Pic}(E)/2$. To apply (5), note that $\dim(\pi) = 0$ since π is birational and $\dim(\tilde{\pi}) = c-1$ since $E = \mathbb{P}_Z(C_{Z/X})$ is the projective bundle of the rank-c conormal bundle $C_{Z/X}$ over Z. $\square$

So far, we have only used Setup 1.1. Now add Hypothesis 1.2 with Y regular.

Remark 2.2. Since Picard groups of regular schemes are homotopy invariant, the $\mathbb{A}^*$-bundle $\alpha : U \to Y$ yields an isomorphism $\alpha^* : \mathrm{Pic}(Y) \xrightarrow{\sim} \mathrm{Pic}(U)$. Let us identify $\mathrm{Pic}(Y)$ with $\mathrm{Pic}(U)$, and hence with $\mathrm{Pic}(X)$ as we did above since $c = \mathrm{codim}_X(Z) \geq 2$. We also have $\mathcal{O}(E)_{|_U} \simeq \mathcal{O}_U$. Putting all this together, the right-hand part of Diagram (3) yields the following on Picard groups:

$$\mathrm{Pic}\left(\begin{tikzcd} X \arrow[d, leftarrow, "\pi"'] & U \arrow[l, hook', "\upsilon"'] \arrow[d, "\alpha"] \arrow[dl, "\tilde\upsilon"'] \\ Bl \arrow[r, "\tilde\alpha"'] & Y \end{tikzcd}\right) \cong \begin{tikzcd} \mathrm{Pic}(X) \arrow[r, equal] \arrow[d, "\binom{1}{0}"'] & \mathrm{Pic}(X) \arrow[d, equal] \\ \mathrm{Pic}(X)\oplus\mathbb{Z} \arrow[ur, "(1\ 0)"] & \mathrm{Pic}(X)\,. \arrow[l, "\binom{1}{\lambda}"'] \end{tikzcd}$$

Note that the lower right map $\mathrm{Pic}(X) \cong \mathrm{Pic}(Y) \xrightarrow{\tilde\alpha^*} \mathrm{Pic}(Bl) \cong \mathrm{Pic}(X)\oplus\mathbb{Z}$ must be of the form $\binom{1}{\lambda}$ by commutativity (i.e. since $(1\ 0)\cdot\binom{1}{\lambda} = 1$) but there is no reason for its second component $\lambda : \mathrm{Pic}(X) \to \mathbb{Z}$ to vanish. This is indeed a key observation. In other words, we have two homomorphisms from $\mathrm{Pic}(X)$ to $\mathrm{Pic}(Bl)$, the direct one π^* and the circumvolant one $\tilde\alpha^* \circ (\alpha^*)^{-1} \circ \upsilon^*$ going via U and Y

$$\begin{tikzcd} \mathrm{Pic}(X) \arrow[r, "\upsilon^*", "\simeq"'] \arrow[d, "\pi^*"'] \arrow[dr, phantom, "\neq"] & \mathrm{Pic}(U) \arrow[d, "\simeq"', "(\alpha^*)^{-1}"] \\ \mathrm{Pic}(Bl) & \mathrm{Pic}(Y) \arrow[l, "\tilde\alpha^*"] \end{tikzcd} \tag{7}$$

and they do *not* coincide in general. The difference is measured by λ, which depends on the choice of Y and on the choice of $\tilde\alpha : Bl \to Y$, in Hypothesis 1.2. So, for every $L \in \mathrm{Pic}(X)$, the integer $\lambda(L) \in \mathbb{Z}$ is defined by the equation

$$\tilde\alpha^*\,(\alpha^*)^{-1}\,\upsilon^*(L) = \pi^*(L) \otimes \mathcal{O}(E)^{\otimes\lambda(L)} \tag{8}$$

in $\mathrm{Pic}(Bl)$. Under the isomorphism $\mathrm{Pic}(Bl) \cong \mathrm{Pic}(X)\oplus\mathbb{Z}$, the above equation can be reformulated as $\tilde\alpha^*\,(\alpha^*)^{-1}\,\upsilon^*(L) = \big(L, \lambda(L)\big)$.

Theorem 2.3 (Partial analogue of Theorem 1.3). *With the notation of 1.1, assume Hypothesis 1.2 and assume X, Y, Z regular. Recall that $c = \mathrm{codim}_X(Z)$. Let $L \in \mathrm{Pic}(X)$ and consider the integer $\lambda(L) \in \mathbb{Z}$ defined in (8) above.*
If $\lambda(L) \equiv c-1 \mod 2$ then the restriction $\upsilon^* : \mathrm{W}^*(X, L) \to \mathrm{W}^*(U, L_{|_U})$ is split surjective, with an explicit section given by the composition $\pi_* \circ \tilde\alpha^* \circ (\alpha^*)^{-1}$

$$\begin{tikzcd} \mathrm{W}^*(X,L) & & \mathrm{W}^*(U, L_{|_U}) \arrow[d, "\simeq"', "(\alpha^*)^{-1}"] \\ \mathrm{W}^*(Bl\,,\,\omega_\pi \otimes \pi^*L) \cong \mathrm{W}^*\big(Bl\,,\,\tilde\alpha^*\,(\alpha^*)^{-1} L_{|_U}\big) \arrow[u, "\pi_*"] & & \mathrm{W}^*\big(Y\,,\,(\alpha^*)^{-1}\,L_{|_U}\big) \arrow[ll, "\tilde\alpha^*"] \end{tikzcd}$$

Proof. The whole point is that π_* can be applied after $\tilde{\alpha}^* \circ (\alpha^*)^{-1}$, that is, on $\mathrm{W}^*\big(Bl\,,\,\tilde{\alpha}^*\,(\alpha^*)^{-1}\,\upsilon^*(L)\big)$. This holds by Proposition 2.1 (A) applied to

$$M := \tilde{\alpha}^*\,(\alpha^*)^{-1}\,\upsilon^*(L) \overset{(8)}{=} (L, \lambda(L)) \in \mathrm{Pic}(X) \oplus \mathbb{Z} = \mathrm{Pic}(Bl)\,. \tag{9}$$

The assumption $\lambda(L) \equiv c-1 \bmod 2$ expresses the hypothesis of Proposition 2.1 (A). Checking that we indeed have a section goes as in the oriented case, see Thm. 1.3 :

$$\upsilon^* \circ \pi_* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1} = \tilde{\upsilon}^* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1} = \alpha^* \circ (\alpha^*)^{-1} = \mathrm{id}\,.$$

The first equality uses base-change [8, Thm. 6.9] on the left-hand cartesian square :

$$\begin{array}{ccccc} X \overset{\upsilon}{\hookleftarrow} U & & L & & L_{|U} \\ {\scriptstyle\pi}\uparrow \quad\quad \uparrow{\scriptstyle\mathrm{id}} & & & & \\ Bl \overset{\tilde{\upsilon}}{\hookleftarrow} U & & \omega_\pi \otimes \pi^* L & & L_{|U} \end{array}$$

with respect to the right-hand line bundles. Note that $(\omega_\pi)_{|U} = \mathcal{O}_U$ by (6). $\square$

Remark 2.4. In the above proof, see (9), we do not apply Proposition 2.1 to M being π^*L, as one could first expect; see Remark 2.2. Consequently, our condition on L, namely $\lambda(L) \equiv c-1 \bmod 2$, does not only depend on the codimension c of Z in X but also involves (hidden in the definition of λ) the particular choice of the auxiliary scheme Y and of the morphism $\tilde{\alpha} : Bl \to Y$ of Hypothesis 1.2.

Remark 2.5. The legitimate question is now to decide what to do in the remaining case, that is, when $\lambda(L) \equiv c \bmod 2$. As announced, this is the central goal of our paper (Thm. 2.6 below). So, let $L \in \mathrm{Pic}(X)$ be a twist such that push-forward along $\pi : Bl \to X$ cannot be applied to define a section to the restriction $\mathrm{W}^*(X, L) \to \mathrm{W}^*(U, L_{|U})$ as above. Actually, we can find examples of such line bundles for which this restriction is simply not surjective (see Ex. 1.5). The natural problem then becomes to compute the possibly non-zero connecting homomorphism $\partial : \mathrm{W}^*(U, L_{|U}) \to \mathrm{W}_Z^{*+1}(X, L)$. Although not absolutely necessary, it actually simplifies the formulation of Theorem 2.6 to use *dévissage* from [9, §6], i.e. the fact that push-forward along a regular closed immersion is an isomorphism

$$\iota_* : \mathrm{W}^{*-c}(Z, \omega_\iota \otimes L_{|Z}) \overset{\sim}{\to} \mathrm{W}_Z^*(X, L)\,. \tag{10}$$

Using this isomorphism, we can replace the Witt groups with supports by Witt groups of Z in the localization long exact sequence, and obtain a long exact sequence

$$\begin{tikzcd} \cdots \ \mathrm{W}^*(X,L) \arrow[r, "\upsilon^*"] & \mathrm{W}^*(U, L_{|U}) \arrow[r, "\partial'"] \arrow[dr, "\partial"'] & \mathrm{W}^{*+1-c}(Z, \omega_\iota \otimes L_{|Z}) \arrow[r] \arrow[d, "\iota_*", "\simeq"'] & \mathrm{W}^{*+1}(X,L) \ \cdots \\ & & \mathrm{W}_Z^{*+1}(X,L) \arrow[ur] & \end{tikzcd} \tag{11}$$

We now want to describe ∂' when $\lambda(L) \equiv c \bmod 2$ (otherwise $\partial' = 0$ by Thm. 2.3).
By the complete dichotomy of Proposition 2.1, we know that when the push-forward $\pi_* : \mathrm{W}^*(Bl, M) \to \mathrm{W}^*(X, ?)$ does not exist, here for $M = \tilde{\alpha}^* \, (\alpha^*)^{-1} \, \upsilon^*(L)$ by (9), then the following composition $\tilde{\pi}_* \circ \tilde{\iota}^*$ exists and starts from the very group where π_* cannot be defined and arrives in the very group where ∂' itself arrives:

$$\mathrm{W}^*(Bl, M) \xrightarrow{\tilde{\iota}^*} \mathrm{W}^*(E, M_{|E}) \xrightarrow{\tilde{\pi}_*} \mathrm{W}^{*-c+1}(Z, \omega_\iota \otimes L_{|Z}) .$$

Hence, in a moment of exaltation, if we blindly apply this observation at the precise point where the oriented technique fails for Witt groups, we see that when we cannot define a section to restriction by the formula $\pi_* \circ \big(\tilde{\alpha}^* \circ (\alpha^*)^{-1}\big)$ we can instead define a mysterious homomorphism $(\tilde{\pi}_* \circ \tilde{\iota}^*) \circ \big(\tilde{\alpha}^* \circ (\alpha^*)^{-1}\big)$.

Theorem 2.6 (Main Theorem in regular case). *With the notation of 1.1, assume Hypothesis 1.2 and assume X, Y, Z regular. Let $L \in \mathrm{Pic}(X)$ and recall the integer $\lambda(L) \in \mathbb{Z}$ defined by* (8)*.*
If $\lambda(L) \equiv c \bmod 2$ then the composition $\tilde{\pi}_* \circ \tilde{\iota}^* \circ \tilde{\alpha}^* \circ (\alpha^*)^{-1}$ is equal to the connecting homomorphism ∂' of (11)*, that is, the following diagram commutes:*

$$\begin{tikzcd}
\mathrm{W}^{*+1-c}(Z, \omega_\iota \otimes L_{|Z}) & & \mathrm{W}^*(U, L_{|U}) \arrow[ll, "\partial'"'] \arrow[d, "\simeq"', "(\alpha^*)^{-1}"] \\
\mathrm{W}^*\big(E\,,\, \omega_{\tilde{\pi}} \otimes \tilde{\pi}^*(\omega_\iota \otimes L_{|Z})\big) \cong \mathrm{W}^*\big(E\,,\, \tilde{\iota}^*\, \tilde{\alpha}^*\, (\alpha^*)^{-1} L_{|U}\big) \arrow[u, "\tilde{\pi}_*"] & & \mathrm{W}^*\big(Y, (\alpha^*)^{-1}\, L_{|U}\big) \arrow[ll, "\tilde{\iota}^*\, \tilde{\alpha}^*"]
\end{tikzcd}$$

This statement implies Thm. 1.4 (B) since $\partial = \iota_* \, \partial'$ by (11). Its proof will be given after generalization to the non-regular setting, at the end of Section 5.

Remark 2.7. Let us stress the peculiar combination of Theorem 2.3 and Theorem 2.6. Start with a Witt class w_U over the open $U \subset X$, for the duality twisted by some $L \in \mathrm{Pic}(U) = \mathrm{Pic}(X)$, and try to extend w_U to a Witt class w_X over X:

$$\begin{tikzcd}
\partial'(w_U) & w_X \arrow[r, mapsto, dotted, "\upsilon^*"] & w_U \arrow[d, mapsto, "(\alpha^*)^{-1}"] \\
w_E \arrow[u, mapsto, dotted, "\tilde{\pi}_*"] \arrow[ur, phantom, "\boxed{\leftarrow 50\% \mid 50\% \rightarrow}"] & w_{Bl} \arrow[l, mapsto, "\tilde{\iota}^*"] \arrow[u, mapsto, dotted, "\pi_*"] & w_Y \arrow[l, mapsto, "\tilde{\alpha}^*"]
\end{tikzcd}$$

Then, *either* we can apply the same construction as for oriented theories, i.e. push-forward the class $w_{Bl} := \tilde{\alpha}^* \circ (\alpha^*)^{-1}(w_U)$ from Bl to X along π, constructing in this way an extension $w_X := \pi_*(w_{Bl})$ as wanted, *or* this last push-forward π_* is forbidden on w_{Bl} because of the twist, in which case the Witt class w_U might simply not belong to the image of restriction υ^*. The latter means that w_U might have a non-zero boundary $\partial'(w_U)$ over Z, which then deserves to be computed. The little miracle precisely is that in order to compute this $\partial'(w_U)$, it suffices to resume the above process where it failed, i.e. with w_{Bl}, and, since we cannot push it forward along π, we can consider the bifurcation of Proposition 2.1 and restrict this class w_{Bl} to the exceptional fiber E, say $w_E := \tilde{\iota}^* w_{Bl}$, and then push it forward along $\tilde{\pi}$. Of course, this does not construct an extension of w_U anymore, since this new class $\tilde{\pi}_*(w_E)$ lives over Z, not over X. Indeed, there is no reason a priori for this new class to give anything sensible at all. Our Main Theorem is that this construction in fact gives a formula for the boundary $\partial'(w_U)$.

Bottom line: *Essentially the same geometric recipe of pull-back and push-forward either splits the restriction or constructs the connecting homomorphism. In particular, the connecting homomorphism is explicitly described in both cases.*

3 The non-regular case

In Section 2, we restricted our attention to the regular case in order to grasp the main ideas. However, most results can be stated in the greater generality of separated and noetherian $\mathbb{Z}[\frac{1}{2}]$-schemes admitting a dualizing complex. The goal of this section is to provide the relevant background and to extend Theorem 1.4 to this non-regular setting, see Main Lemma 3.5.

Remark 3.1. The *coherent Witt groups* $\tilde{\mathrm{W}}^*(X, K_X)$ of a scheme $X \in \mathcal{S}ch$ (see 1.1) are defined using the derived category $\mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(X)$ of complexes of $\mathcal{O}_X$-modules whose cohomology is coherent and bounded. Since X is noetherian and separated, $\mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(X)$ is equivalent to its subcategory $\mathrm{D}^{\mathrm{b}}(\mathrm{coh}(X))$ of bounded complexes of coherent $\mathcal{O}_X$-modules; see for instance [8, Prop. A.4]. The duality is defined using the derived functor $\mathrm{RHom}(-, K_X)$ where $K_X \in \mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(X)$ is a *dualizing complex* (see [19, § 3] or [8, § 2]), meaning that the functor $\mathrm{RHom}(-, K_X)$ defines a duality on $\mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(X)$. For example, a scheme is Gorenstein if and only if $\mathcal{O}_X$ itself is an injectively bounded dualizing complex and, in that case, all other dualizing complexes are shifted line bundles (see Lemma A.7). Regular schemes are Gorenstein, and for them, coherent Witt groups coincide with the usual "locally free" Witt groups $\mathrm{W}^*(X, L)$ (i.e. the ones defined using bounded complexes of locally free sheaves instead of coherent ones). For any line bundle L, we still have a square-periodicity isomorphism

$$\tilde{\mathrm{W}}(X, K_X) \cong \tilde{\mathrm{W}}(X, K_X \otimes L^{\otimes 2}) \tag{12}$$

given by the multiplication by the class in $\mathrm{W}^0(X, L^{\otimes 2})$ of the canonical form $L \to L^\vee \otimes L^{\otimes 2}$, using the pairing between locally free and coherent Witt groups. For any closed embedding $Z \hookrightarrow X$ with open complement $\upsilon : U \hookrightarrow X$, the restriction $K_U := \upsilon^* K_X$ is a dualizing complex [19, Thm. 3.12] and the general triangulated framework of [3] gives a localization long exact sequence

$$\cdots \xrightarrow{\partial} \tilde{\mathrm{W}}^*_Z(X, K_X) \longrightarrow \tilde{\mathrm{W}}^*(X, K_X) \longrightarrow \tilde{\mathrm{W}}^*(U, K_U) \xrightarrow{\partial} \tilde{\mathrm{W}}^{*+1}_Z(X, K_X) \longrightarrow \cdots \tag{13}$$

As for K-theory, no such sequence holds in general for singular schemes and locally free Witt groups.

Remark 3.2. For coherent Witt groups, the push-forward along a proper morphism $f : X' \to X$ takes the following very round form: If K_X is a dualizing complex on X then $f^! K_X$ is a dualizing complex on X' ([8, Prop. 3.9]) and the functor $\mathrm{R}\, f_* : \mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(X') \to \mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(X)$ induces a *push-forward* ([8, Thm. 4.4])

$$f_* : \tilde{\mathrm{W}}^i(X', f^! K_X) \to \tilde{\mathrm{W}}^i(X, K_X)\,. \tag{14}$$

Recall that $f^! : \mathrm{D}_{\mathrm{Qcoh}}(X) \to \mathrm{D}_{\mathrm{Qcoh}}(X')$ is the right adjoint of $\mathrm{R}\, f_*$. If we twist the chosen dualizing complex K_X by a line bundle $L \in \mathrm{Pic}(X)$, this is transported to X' via the following formula (see [8, Thm. 3.7])

$$f^!(K_X \otimes L) \simeq f^!(K_X) \otimes f^* L\,. \tag{15}$$

In the regular case, push-forward maps are also described in Nenashev [20].

Remark 3.3. Let us also recall from [8, Thm. 4.1] that the pull-back

$$f^* : \tilde{\mathrm{W}}^i(X, K_X) \to \tilde{\mathrm{W}}^i(X', \mathrm{L}f^* K_X)$$

along a finite Tor-dimension morphism $f : X' \to X$ is defined if $\mathrm{L}f^*(K_X)$ is a dualizing complex (this is not automatically true). Together with the push-forward, this pull-back satisfies the usual flat base-change formula (see [8, Thm. 5.5]).
A regular immersion $f : X' \hookrightarrow X$ has finite Tor-dimension since it is even perfect (see [1, p. 250]). Moreover, in that case, $\mathrm{L}f^*$ is the same as $f^!$ up to a twist and a shift (see Proposition A.8), hence it preserves dualizing complexes.

Proposition 3.4. *In Setup 1.1, let K_X be a dualizing complex on X. Let $L \in \mathrm{Pic}(X)$ and $\ell \in \mathbb{Z}$. Then $K = \pi^!(K_X) \otimes \pi^* L \otimes \mathcal{O}(E)^{\otimes \ell}$ is a dualizing complex on Bl and any dualizing complex has this form, for some $L \in \mathrm{Pic}(X)$ and $\ell \in \mathbb{Z}$. Moreover, the dichotomy of Proposition 2.1 here becomes:*

(A) If $\ell \equiv 0 \mod 2$, we can push-forward along $\pi : Bl \to X$, as follows:

$$\tilde{\mathrm{W}}^*(Bl, K) \cong \tilde{\mathrm{W}}^*\big(Bl, \pi^!(K_X \otimes L)\big) \xrightarrow{\ \pi_*\ } \tilde{\mathrm{W}}^*\big(X, K_X \otimes L\big)\,.$$

(B) If $\ell \equiv 1 \mod 2$, we can push-forward along $\tilde{\pi} : E \to Z$, as follows:

$$\tilde{\mathrm{W}}^*(E, \mathrm{L}\tilde{\iota}^* K) \cong \tilde{\mathrm{W}}^{*+1}\big(E, \tilde{\pi}^! \iota^!(K_X \otimes L)\big)\big) \xrightarrow{\ \tilde{\pi}_*\ } \tilde{\mathrm{W}}^{*+1}\big(Z, \iota^!(K_X \otimes L)\big).$$

As before, in both cases, the first isomorphism $\cong$ comes from square-periodicity (12) and the second morphism is push-forward (14).

Proof. The complex $K_{Bl} := \pi^! K_X$ is a dualizing complex on Bl by Remark 3.2. By Lemma A.7 and Proposition A.6 (i), all dualizing complexes on Bl are of the form $K = \pi^!(K_X) \otimes \pi^* L \otimes \mathcal{O}(E)^{\otimes \ell}$, for unique $L \in \mathrm{Pic}(X)$ and $\ell \in \mathbb{Z}$. We only need to check the relevant parity for applying (12). Case (A) follows easily from (15) by definition of K and parity of ℓ. In (B), we need to compare $\mathrm{L}\tilde\iota^* K$ and $\tilde\pi^! \iota^! (K_X \otimes L)[1]$. By Proposition A.11 (iv), we know that $\tilde\iota^!(-) \cong \tilde\iota^* \mathcal{O}(E)[-1] \otimes \mathrm{L}\tilde\iota^*(-)$. We apply this and (15) in the second equality below, the first one using simply that $\iota\tilde\pi = \pi\tilde\iota$:

$$\tilde\pi^! \iota^! (K_X \otimes L)[1] \cong \tilde\iota^! \pi^! (K_X \otimes L)[1] \cong \tilde\iota^* \mathcal{O}(E)[-1] \otimes \mathrm{L}\tilde\iota^* \big(\pi^!(K_X) \otimes \pi^* L\big)[1] \cong$$
$$\cong \tilde\iota^* \mathcal{O}(E) \otimes \mathrm{L}\tilde\iota^* (K \otimes \mathcal{O}(E)^{\otimes -\ell}) \cong \tilde\iota^* \mathcal{O}(E)^{\otimes (1-\ell)} \otimes \mathrm{L}\tilde\iota^* K.$$

Since $1-\ell$ is even, $\tilde\iota^* \mathcal{O}(E)^{\otimes(1-\ell)}$ is a square, as desired. □

We now want to give the key technical result of the paper, which is an analogue of Theorem 1.4 in the non-regular setting. The idea is to describe the connecting homomorphism on Witt classes over U which admit an extension to the blow-up Bl. The key fact is the existence of an additional twist on Bl, namely the twist by $\mathcal{O}(E)$, which disappears on U (see A.1) and hence allows Case (B) below.

Main Lemma 3.5. *In Setup 1.1, assume that X has a dualizing complex K_X and let $K_U = \upsilon^*(K_X)$ and $K_{Bl} = \pi^!(K_X)$; see Remarks 3.1 and 3.2. Let $i \in \mathbb{Z}$.*

(A) The following composition vanishes:

$$\tilde{\mathrm{W}}^i(Bl, K_{Bl}) \xrightarrow{\tilde\upsilon^*} \tilde{\mathrm{W}}^i(U, K_U) \xrightarrow{\partial} \tilde{\mathrm{W}}^{i+1}_Z(X, K_X)\,.$$

(B) The following composition (well-defined since $\tilde\upsilon^* \mathcal{O}(E) \simeq \mathcal{O}_U$)

$$\tilde{\mathrm{W}}^i\big(Bl, K_{Bl} \otimes \mathcal{O}(E)\big) \xrightarrow{\tilde\upsilon^*} \tilde{\mathrm{W}}^i\big(U, K_U \otimes \tilde\upsilon^* \mathcal{O}(E)\big) \cong \tilde{\mathrm{W}}^i(U, K_U) \xrightarrow{\partial} \tilde{\mathrm{W}}^{i+1}_Z(X, K_X)$$

coincides with the composition

$$\begin{array}{ccccc} \tilde{\mathrm{W}}^i(Bl, \mathcal{O}(E) \otimes K_{Bl}) & & & & \tilde{\mathrm{W}}^{i+1}_Z(X, K_X) \\ \big\downarrow{\scriptstyle \tilde\iota^*} & & & & \big\uparrow{\scriptstyle \iota_*} \\ \tilde{\mathrm{W}}^i\big(E, \mathrm{L}\tilde\iota^*(\mathcal{O}(E) \otimes K_{Bl})\big) & \cong & \tilde{\mathrm{W}}^{i+1}\big(E, \tilde\pi^! \iota^! K_X\big) & \xrightarrow{\tilde\pi_*} & \tilde{\mathrm{W}}^{i+1}\big(Z, \iota^! K_X\big) \end{array}$$

where the latter isomorphism $\cong$ is induced by the composition

$$\mathrm{L}\tilde\iota^*(\mathcal{O}(E) \otimes K_{Bl}) \cong \tilde\iota^*(\mathcal{O}(E)) \otimes \mathrm{L}\tilde\iota^*(K_{Bl}) \cong \tilde\iota^! K_{Bl}[1] \cong \tilde\pi^! \iota^! K_X[1]\,. \quad (16)$$

The proof of this result occupies Section 4. Here are just a couple of comments on the statement. Let us first of all explain the announced sequence of isomorphisms (16). The first one holds since $\mathrm{L}\tilde\iota^*$ is a tensor functor and since $\mathcal{O}(E)$ is a line bundle (hence is flat). The second one holds by Proposition A.9 (v). The last one follows by definition of K_{Bl} and the fact that $\iota\,\tilde\pi = \pi\,\tilde\iota$. Finally, note that we use the pull-back $\tilde\iota^*$ on coherent Witt groups as recalled in Remark 3.3.

4 The main argument

Surprisingly enough for a problem involving the blow-up $Bl = Bl_Z(X)$ of X along Z, see (1), the case where $\operatorname{codim}_X(Z) = 1$ is also interesting, even though, of course, in that case $Bl = X$ and $E = Z$. In fact, this case is crucial for the proof of Main Lemma 3.5 and this is why we deal with it first. In the "general" proof where $\operatorname{codim}_X(Z)$ is arbitrary, we will apply the case of codimension one to $\tilde{\iota} : E \hookrightarrow Bl$. Therefore, we use the following notation to discuss codimension one.

Notation 4.1. Let $B \in \mathcal{S}ch$ be a scheme with a dualizing complex K_B and $\tilde{\iota} : E \hookrightarrow B$ be a prime divisor, that is, a regular closed immersion of codimension one, of a subscheme $E \in \mathcal{S}ch$. Let $\mathcal{O}(E)$ be the line bundle on B associated to E (see Definition A.1). Let $\tilde{\upsilon} : U \hookrightarrow B$ be the open immersion of the open complement

$$E \xhookrightarrow{\tilde{\iota}} B \xhookleftarrow{\tilde{\upsilon}} U$$

$U = B - E$ and let K_U be the dualizing complex $\tilde{\upsilon}^*(K_B)$ on U.

Lemma 4.2 (Main Lemma in codimension one). *With Notation 4.1, let $i \in \mathbb{Z}$. Then:*

(A) The composition

$$\tilde{\mathrm{W}}^i(B, K_B) \xrightarrow{\tilde{\upsilon}^*} \tilde{\mathrm{W}}^i(U, K_U) \xrightarrow{\partial} \tilde{\mathrm{W}}^{i+1}_E(B, K_B)$$

is zero.

(B) The composition

$$\tilde{\mathrm{W}}^i\big(B, K_B \otimes \mathcal{O}(E)\big) \xrightarrow{\tilde{\upsilon}^*} \tilde{\mathrm{W}}^i\big(U, K_U \otimes \tilde{\upsilon}^*\mathcal{O}(E)\big) \cong \tilde{\mathrm{W}}^i(U, K_U) \xrightarrow{\partial} \tilde{\mathrm{W}}^{i+1}_E(B, K_B)$$

coincides with the composition

$$\begin{tikzcd}
\tilde{\mathrm{W}}^i(B, \mathcal{O}(E) \otimes K_B) \arrow[d, "\tilde{\iota}^*"'] & & \tilde{\mathrm{W}}^{i+1}_E(B, K_B) \\
\tilde{\mathrm{W}}^i\big(E, \mathrm{L}\tilde{\iota}^*(\mathcal{O}(E) \otimes K_B)\big) \arrow[r, phantom, "\cong"] & \tilde{\mathrm{W}}^i\big(E, \tilde{\iota}^! K_B[1]\big) \arrow[r, phantom, "\cong"] & \tilde{\mathrm{W}}^{i+1}\big(E, \tilde{\iota}^! K_B\big) \arrow[u, "\tilde{\iota}_*"']
\end{tikzcd}$$

where the first isomorphism $\cong$ is induced by the following isomorphism

$$\mathrm{L}\tilde{\iota}^*(\mathcal{O}(E) \otimes K_B) \cong \tilde{\iota}^*(\mathcal{O}(E)) \otimes \mathrm{L}\tilde{\iota}^*(K_B) \cong \tilde{\iota}^!(K_B)[1]. \tag{17}$$

Proof. Case (A) is simple: The composition of two consecutive morphisms in the localization long exact sequence (13) is zero. Case (B) is the nontrivial one. The isomorphisms (17) are the same as in (16).

At this stage, we upload the definition of the connecting homomorphism for Witt groups $\partial : \tilde{\mathrm{W}}^i(U, K_U) \to \tilde{\mathrm{W}}^{i+1}_E(B, K_B)$, which goes as follows: Take a non-degenerate symmetric space (P, ϕ) over U for the i^{th}-shifted duality with values in K_U; there exists a possibly degenerate symmetric pair (Q, ψ) over B for the same duality (with values in K_B) which restricts to (P, ϕ) over U; compute its symmetric cone $d(Q, \psi)$, which is essentially the cone of ψ equipped with a natural metabolic form; see [3, § 4] or [4, Def. 2.3] for instance; for any choice of such a pair (Q, ψ), the boundary $\partial(P, \phi) \in \tilde{\mathrm{W}}^{i+1}_E(B, K_B)$ is the Witt class of $d(Q, \psi)$.

There is nothing really specific to dualizing complexes here. The above construction is a purely triangular one, as long as one uses the *same* duality for the ambient scheme B, for the open $U \subset B$ and for the Witt group of B with supports in the closed complement E. The subtlety of statement (B) is that we start with a twisted duality on the scheme B which is not the duality used for ∂, but which agrees with it on U by the first isomorphism $\cong$ in statement (B).

Now, take an element in $\tilde{\mathrm{W}}^i(B, \mathcal{O}(E) \otimes K_B)$. It is the Witt-equivalence class of a symmetric space (P, ϕ) over B with respect to the i^{th}-shifted duality with values in $\mathcal{O}(E) \otimes K_B$. The claim of the statement is that, modulo the above identifications of dualizing complexes, we should have

$$\partial(\tilde{v}^*(P, \phi)) = \tilde{\iota}_*(\tilde{\iota}^*(P, \phi)) \tag{18}$$

in $\tilde{\mathrm{W}}^{i+1}_E(B, K_B)$. By the above discussion, in order to compute $\partial(\tilde{v}^*(P, \phi))$, we need to find a symmetric pair (Q, ψ) over B, for the duality given by K_B, and such that $\tilde{v}^*(Q, \psi) = \tilde{v}^*(P, \phi)$. Note that we cannot take for (Q, ψ) the pair (P, ϕ) itself because (P, ϕ) is symmetric for the twisted duality $\mathcal{O}(E) \otimes K_B$ on B. Nevertheless, it is easy to "correct" (P, ϕ) as follows.

As in Definition A.1, we have a canonical homomorphism of line bundles:

$$\sigma_E : \mathcal{O}(E)^\vee \to \mathcal{O}_B \,.$$

The pair $(\mathcal{O}(E)^\vee, \sigma_E)$ is symmetric in the derived category $\mathrm{D^b}(\mathrm{VB}(B))$ of vector bundles over B, with respect to the 0^{th}-shifted duality twisted by $\mathcal{O}(E)^\vee$, because the target of σ_E is the dual of its source: $(\mathcal{O}(E)^\vee)^\vee[0] \otimes \mathcal{O}(E)^\vee \cong \mathcal{O}_B$. Let us define the wanted symmetric pair (Q, ψ) in $\mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}(B)$ for the i^{th}-shifted duality with values in K_B as the following product:

$$(Q, \psi) := (\mathcal{O}(E)^\vee, \sigma_E) \otimes (P, \phi) \,.$$

Note that we tensor a complex of vector bundles with a coherent one to get a coherent one, following the formalism of [4, § 4] where such external products are denoted by $\star$. We claim that the restriction of (Q, ψ) to U is nothing but $\tilde{v}^*(P, \phi)$. This is easy to check since $\mathcal{O}(E)_{|U} = \mathcal{O}_U$ via σ_E (see A.1), which means $(\mathcal{O}(E)^\vee, \sigma_E)_{|U} = 1_U$. So, by the construction of the connecting homomorphism ∂ recalled at the beginning of the proof, we know that $\partial(\tilde{v}^*(P, \phi))$ can be computed as $d(Q, \psi)$. This reads:

$$\partial(\tilde{v}^*(P, \phi)) = d\big((\mathcal{O}(E)^\vee, \sigma_E) \otimes (P, \phi)\big).$$

Now, we use that (P,ϕ) is non-degenerate and that therefore (see [4, Rem. 5.4] if necessary) we can take (P,ϕ) out of the above symmetric cone $d(...)$, i.e.

$$\partial(\tilde{v}^*(P,\phi)) = d(\mathcal{O}(E)^\vee, \sigma_E) \otimes (P,\phi)\,. \tag{19}$$

Let us compute the symmetric cone $d(\mathcal{O}(E)^\vee, \sigma_E) =: (C,\chi)$. Note that this only involves vector bundles. We define C to be the cone of σ_E and we equip it with a symmetric form $\chi : C \overset{\sim}{\to} C^\vee[1] \otimes \mathcal{O}(E)^\vee$ for the duality used for $(\mathcal{O}(E),\sigma_E)$ but shifted by one, that is, for the 1^{st} shifted duality with values in $\mathcal{O}(E)^\vee$. One checks that (C,χ) is given by the following explicit formula:

$$\begin{tikzcd}
C = \arrow[d, "\chi"'] & \big(\cdots \arrow[r] & 0 \arrow[r] \arrow[d] & 0 \arrow[r] \arrow[d] & \mathcal{O}(E)^\vee \arrow[r, "\sigma_E"] \arrow[d, "-1"'] & \mathcal{O}_B \arrow[r] \arrow[d, "1"] & 0 \arrow[r] \arrow[d] & 0 \arrow[r] \arrow[d] & \cdots\big) \\
C^\vee[1]\otimes\mathcal{O}(E)^\vee = & \big(\cdots \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & \mathcal{O}(E)^\vee \arrow[r, "-(\sigma_E)^\vee"'] & \mathcal{O}_B \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & \cdots\big)
\end{tikzcd} \tag{20}$$

where the complexes have $\mathcal{O}_B$ in degree zero. Now, observe that the complex C is a resolution of $\tilde{\iota}_*(\mathcal{O}_E)$ over B, by Definition A.1, that is, $C \simeq \tilde{\iota}_*(\mathcal{O}_E)$ in the derived category of B. Moreover, by Propositions A.8 and A.9 (ii), we have $\tilde{\iota}^!(\mathcal{O}(E)^\vee[1]) = \omega_{\tilde{\iota}}[-1] \otimes \tilde{\iota}^*(\mathcal{O}(E)^\vee[1]) \cong \mathcal{O}_E$. Using this, one checks the conceptually obvious fact that χ is also the push-forward along the perfect morphism $\tilde{\iota}$ of the unit form on $\mathcal{O}_E$. See Remark 4.3 below for more details. This means that we have an isometry in $\mathrm{D}^{\mathrm{b}}_E(\mathrm{VB}(B))$

$$d\big(\mathcal{O}(E)^\vee, \sigma_E\big) = \tilde{\iota}_*(1_E)$$

of symmetric spaces with respect to the 1^{st} shifted duality with values in $\mathcal{O}(E)^\vee$. Plugging this last equality in (19), and using the projection formula (see Remark 4.3) we obtain

$$\partial(\tilde{v}^*(P,\phi)) = \tilde{\iota}_*(1_E)\otimes(P,\phi) = \tilde{\iota}_*\big(1_E \otimes \tilde{\iota}^*(P,\phi)\big) = \tilde{\iota}_*\big(\tilde{\iota}^*(P,\phi)\big)\,.$$

This is the claimed equality (18). $\square$

Remark 4.3. In the above proof, we use the "conceptually obvious fact" that the push-forward of the unit form on $\mathcal{O}_E$ is indeed the χ of (20). This fact is obvious to the expert but we cannot provide a direct reference for this exact statement. However, if the reader does not want to do this lengthy verification directly, the computation of [8, § 7.2] can be applied essentially verbatim. The main difference is that here, we are considering a push-forward of locally free instead of coherent Witt groups along a regular embedding. Such a push-forward is constructed using the same tensor formalism as the proper push-forwards for coherent Witt groups considered in loc. cit. along morphisms that are proper, perfect and Gorenstein, which is true of a regular embedding. In loc. cit. there is an assumption that the schemes are Gorenstein, ensuring that

the line bundles are dualizing complexes. But here, the dualizing objects for our category of complexes of locally free sheaves are line bundles anyway and the extra Gorenstein assumption is irrelevant.
Moreover, the projection formula used in the above proof is established in complete generality for non necessarily regular schemes by the same method as in [8, § 5.7] using the pairing between the locally free derived category and the coherent one to the coherent one. More precisely, this pairing is just a restriction of the quasi-coherent pairing $\mathrm{D}_{\mathrm{Qcoh}} \times \mathrm{D}_{\mathrm{Qcoh}} \xrightarrow{\otimes} \mathrm{D}_{\mathrm{Qcoh}}$ of loc. cit. to these subcategories. By the general tensor formalism of [10], for any morphism $f : X \to Y$ as above, for any object A (resp. B) in the quasi-coherent derived category of X (resp. Y), we obtain a projection morphism in $\mathrm{D}_{\mathrm{Qcoh}}(Y)$

$$\mathrm{R}\, f_*(A) \otimes B \longrightarrow \mathrm{R}\, f_*(A \otimes \mathrm{L} f^*(B)),$$

see [10, Prop. 4.2.5]. It is an isomorphism by [8, Thm. 3.7]. We actually only use it for A a complex of locally free sheaves and B a complex with coherent and bounded cohomology. The projection formula is implied by [10, Thm. 5.5.1].

Proof of Main Lemma 3.5. Case (A) follows from the codimension one case and the compatibility of push-forwards with connecting homomorphisms (here along the identity of U). Case (B) follows from the outer commutativity of the following diagram:

$$\begin{tikzcd}
\tilde{\mathrm{W}}^i(U, K_U) \arrow[r, "\partial"] \arrow[d, equal] & \tilde{\mathrm{W}}^{i+1}_Z(X, K_X) & \tilde{\mathrm{W}}^{i+1}(Z, \iota^! K_X) \arrow[l, "\iota_*"'] \\
\tilde{\mathrm{W}}^i(U, K_U) \arrow[r, "\partial"] & \tilde{\mathrm{W}}^{i+1}_E(Bl, K_{Bl}) \arrow[u, "\pi_*"'] & \tilde{\mathrm{W}}^{i+1}(E, \tilde{\pi}^! \iota^! K_X) \arrow[l, "\tilde{\iota}_*"'] \arrow[u, "\tilde{\pi}_*"'] \\
\tilde{\mathrm{W}}^i(Bl, \mathcal{O}(E) \otimes K_{Bl}) \arrow[r, "\tilde{\iota}^*"] \arrow[u, "\tilde{v}^*"] & \tilde{\mathrm{W}}^i\big(Bl, \mathrm{L}\tilde{\iota}^*(\mathcal{O}(E) \otimes K_{Bl})\big) \arrow[r, phantom, "\cong"] & \tilde{\mathrm{W}}^{i+1}(E, \tilde{\iota}^! K_{Bl}) \arrow[u, "\cong"']
\end{tikzcd} \tag{21}$$

We shall now verify the inner commutativity of this diagram. The upper left square of (21) commutes by compatibility of push-forward with connecting homomorphisms. The upper right square of (21) simply commutes by functoriality of push-forward applied to $\iota \circ \tilde{\pi} = \pi \circ \tilde{\iota}$. Most interestingly, the lower part of (21) commutes by Lemma 4.2 applied to the codimension one inclusion $\tilde{\iota} : E \hookrightarrow Bl$. ∎

5 The Main Theorem in the non-regular case

Without regularity assumptions, we have shown in Main Lemma 3.5 how to compute the connecting homomorphism $\partial : \tilde{\mathrm{W}}^*(U, K_U) \to \tilde{\mathrm{W}}^{*+1}_Z(X, K_X)$ on those Witt classes over U which come from $Bl = Bl_Z(X)$ by restriction $\tilde{v}^*$. The whole point of adding Hypothesis 1.2 is precisely to split $\tilde{v}^*$, that is, to construct for each Witt class on U an extension on Bl. In the regular case, this follows from homotopy invariance of Picard groups and Witt groups. In the non-regular setting, things are a little more complicated. Let us give the statement and comment on the hypotheses afterwards (see Remark 5.2).

Main Theorem 5.1. *In Setup 1.1, assume that X has a dualizing complex K_X and equip U with the restricted complex $K_U = \upsilon^*(K_X)$. Assume Hypothesis 1.2 and further make the following hypotheses:*

(a) There exists a dualizing complex K_Y on Y such that $\alpha^* K_Y = K_U$.

(b) The $\mathbb{A}^*$-bundle α induces an isomorphism $\tilde{\mathrm{W}}^*(Y, K_Y) \xrightarrow{\sim} \tilde{\mathrm{W}}^*(U, K_U)$.

(c) The morphism $\tilde{\alpha}$ is of finite Tor dimension and $\mathrm{L}\tilde{\alpha}^*(K_Y)$ is dualizing.

(d) Sequence (25) is exact: $\mathbb{Z} \to \mathrm{Pic}(Bl) \to \mathrm{Pic}(U)$. (See Proposition A.3.)

Then $\mathrm{L}\tilde{\alpha}^*(K_Y) \simeq \pi^! K_X \otimes \mathcal{O}(E)^{\otimes n}$ for some $n \in \mathbb{Z}$, and the following holds true:

(A) If n can be chosen even, the composition $\pi_* \tilde{\alpha}^* (\alpha^*)^{-1}$ is a section of υ^*.

(B) If n can be chosen odd, the composition $\iota_* \tilde{\pi}_* \tilde{\iota}^* \tilde{\alpha}^* (\alpha^*)^{-1}$ coincides with the connecting homomorphism $\partial : \tilde{\mathrm{W}}^*(U, K_U) \to \tilde{\mathrm{W}}^{*+1}_E(X, K_X)$.

Proof. By (c) and Remark 3.2 respectively, both $\mathrm{L}\tilde{\alpha}^*(K_Y)$ and $\pi^! K_X$ are dualizing complexes on Bl. By Lemma A.7 (i), they differ by a shifted line bundle: $\mathrm{L}\tilde{\alpha}^*(K_Y) \simeq \pi^! K_X \otimes L[m]$ with $L \in \mathrm{Pic}(Bl)$ and $m \in \mathbb{Z}$. Restricting to U, we get

$$K_U \otimes \tilde{\upsilon}^* L[m] \simeq \tilde{\upsilon}^* \pi^! K_X \otimes \tilde{\upsilon}^* L[m] \simeq \tilde{\upsilon}^* (\pi^! K_X \otimes L[m]) \simeq \tilde{\upsilon}^* \, \mathrm{L}\tilde{\alpha}^*(K_Y) \simeq \alpha^* K_Y \simeq K_U$$

where the first equality holds by flat base-change ([8, Thm. 5.5]). Thus, $\tilde{\upsilon}^* L[m]$ is the trivial line bundle on U by Lemma A.7 (ii). So $m = 0$ and, by (d), $L \simeq \mathcal{O}(E)^{\otimes n}$ for some $n \in \mathbb{Z}$. This gives $\mathrm{L}\tilde{\alpha}^*(K_Y) \simeq \pi^! K_X \otimes \mathcal{O}(E)^{\otimes n}$ as claimed.

We now consider coherent Witt groups. By (c) and Remark 3.3, $\tilde{\alpha}$ induces a morphism $\tilde{\alpha}^* : \tilde{\mathrm{W}}^*(Y, K_Y) \to \tilde{\mathrm{W}}(Bl, \mathrm{L}\tilde{\alpha}^* K_Y)$. By Lemma A.12, the flat morphism α induces a homomorphism $\alpha^* : \tilde{\mathrm{W}}^*(Y, K_Y) \to \tilde{\mathrm{W}}^*(U, K_U)$ which is assumed to be an isomorphism in (b). So, we can use $(\alpha^*)^{-1}$. When n is even, we have

$$\upsilon^* \pi_* \tilde{\alpha}^* (\alpha^*)^{-1} = \tilde{\upsilon}^* \tilde{\alpha}^* (\alpha^*)^{-1} = \alpha^* (\alpha^*)^{-1} = \mathrm{id}$$

where the first equality holds by flat base-change ([8, Thm. 5.5]). This proves (A). On the other hand, when n is odd, we have

$$\iota_* \tilde{\pi}_* \tilde{\iota}^* \tilde{\alpha}^* (\alpha^*)^{-1} = \partial \, \tilde{\upsilon}^* \tilde{\alpha}^* (\alpha^*)^{-1} = \partial \, \alpha^* (\alpha^*)^{-1} = \partial$$

where the first equality holds by Main Lemma 3.5 (B). $\square$

Remark 5.2. Hypothesis (a) in Theorem 5.1 is always true when Y admits a dualizing complex and homotopy invariance holds over Y for the Picard group (e.g. Y regular). Homotopy invariance for coherent Witt groups should hold in general but only appears in the literature when Y is Gorenstein, see Gille [12]. This means that Hypothesis (b) is a mild one. Hypothesis (d) is discussed in Proposition A.3.

Remark 5.3. In Theorem 5.1, the equation $\mathrm{L}\tilde{\alpha}^*(K_Y) \simeq \pi^! K_X \otimes \mathcal{O}(E)^{\otimes n}$, for $n \in \mathbb{Z}$, should be considered as a non-regular analogue of Equation (8). In Remark 2.2, we discussed the compatibility of the various lines bundles on the schemes X, U, Y and Bl. Here, we need to control the relationship between dualizing complexes instead and we do so by restricting to U and by using the exact sequence (25). Alternatively, one can remove Hypothesis (d) and directly assume the relation $\mathrm{L}\tilde{\alpha}^*(K_Y) \simeq \pi^!(K_X) \otimes \mathcal{O}(E)^{\otimes n}$ for some $n \in \mathbb{Z}$. This might hold in some particular examples even if (25) is not exact.

For the convenience of the reader, we include the proofs of the following facts.

Lemma 5.4. *If X is Gorenstein, then Z and Bl are Gorenstein. If X is regular, Bl is regular.*

Proof. By Prop. A.8, $\pi^!(\mathcal{O}_X)$ is the line bundle ω_π. Since π is proper, $\pi^!$ preserves injectively bounded dualizing complexes and ω_π is dualizing and since it is a line bundle, Bl is Gorenstein. The same proof holds for Z, since $\iota^!(\mathcal{O}_X)$ is ω_ι (shifted) which is also a line bundle by Prop. A.11. For regularity, see [18, Thm. 1.19]. □

Proof of Theorem 2.6. Note that all the assumptions of Theorem 5.1 are fulfilled in the regular case, that is, in the setting of Section 2. Indeed, if X and Y are regular, Bl and U are regular, and the dualizing complexes on X, Y, Bl and U are simply shifted line bundles. The morphism $\alpha^* : \mathrm{Pic}(Y) \to \mathrm{Pic}(U)$ is then an isomorphism (homotopy invariance) and $\tilde{\alpha}$ is automatically of finite Tor dimension, as any morphism to a regular scheme. Finally, the sequence on Picard groups is exact by Proposition A.3.
Let $K_X = L$ be the chosen line bundle on X. Then set $L_U := K_U = \upsilon^* L$ and choose $L_Y = K_Y$ to be the unique line bundle (up to isomorphism) such that $\alpha^* L_Y = L_U$. By (8), we have $\tilde{\alpha}^* L_Y = \pi^* L \otimes \mathcal{O}(E)^{\otimes \lambda(L)} = \pi^! L \otimes \mathcal{O}(E)^{\otimes(\lambda(L)-c+1)}$, where the last equality holds since $\pi^! L = \mathcal{O}(E)^{\otimes(c-1)} \otimes \pi^* L$ by Proposition A.11 (vi). In other words, we have proved that $\tilde{\alpha}^* K_Y = \pi^! K_X \otimes \mathcal{O}(E)^{\otimes(\lambda(L)-c+1)}$. In Theorem 5.1, we can then take $n = \lambda(L) - c + 1$ and the parity condition becomes $\lambda(L) \equiv c-1 \mod 2$ for Case (A) and $\lambda(L) \equiv c \mod 2$ for Case (B). So, Case (A) is the trivial one and corresponds to Theorem 2.3. Case (B) exactly gives Theorem 2.6 up to the identifications of line bundles explained in Appendix A. □

A Line bundles and dualizing complexes

We use Hartshorne [15] or Liu [18] as general references for algebraic geometry. We still denote by $\mathcal{S}ch$ the category of noetherian separated connected schemes (we do not need "over $\mathbb{Z}[\frac{1}{2}]$" in this appendix).

Definition A.1. Let $\tilde{\iota} : E \hookrightarrow B$ be a regular closed immersion of codimension one, with $B \in Sch$. Consider the ideal $I_E \subset \mathcal{O}_B$ defining E

$$0 \longrightarrow I_E \xrightarrow{\sigma_E} \mathcal{O}_B \longrightarrow \tilde{\iota}_* \mathcal{O}_E \longrightarrow 0\ . \tag{22}$$

By assumption, I_E is an invertible ideal, i.e. a line bundle. The *line bundle associated to* E is defined as its dual $\mathcal{O}(E) := (I_E)^\vee$, see [15, II.6.18]. We thus have by construction a global section $\sigma_E : \mathcal{O}(E)^\vee \to \mathcal{O}_B$, which vanishes exactly on E. This gives an explicit trivialization of $\mathcal{O}(E)$ outside E. On the other hand, the restriction of $\mathcal{O}(E)$ to E is the normal bundle $\mathcal{O}(E)_{|_E} \cong N_{E/B}$.

Example A.2. Let $Bl = Bl_Z(X)$ be the blow-up of X along a regular closed immersion $Z \hookrightarrow X$ as in Setup 1.1. Let $I = I_Z \subset \mathcal{O}_X$ be the sheaf of ideals defining Z. By construction of the blow-up, we have $Bl = \mathrm{Proj}(\mathcal{S})$ where $\mathcal{S}$ is the sheaf of graded $\mathcal{O}_X$-algebras

$$\mathcal{S} := \mathcal{O}_X \oplus I \oplus I^2 \oplus I^3 \oplus \cdots$$

Similarly, $E = \mathrm{Proj}(S/\mathcal{J})$ where $\mathcal{J} := I \cdot \mathcal{S} \subset \mathcal{S}$ is the sheaf of homogeneous ideals

$$\mathcal{J} = I \oplus I^2 \oplus I^3 \oplus I^4 \oplus \cdots$$

So, $E = \mathbb{P}_Z(C_{Z/X})$ is a projective bundle over Z associated to the vector bundle $C_{Z/X} = I/I^2$ which is the conormal bundle of Z in X. Associating $\mathcal{O}_{Bl}$-sheaves to graded $\mathcal{S}$-modules, the obvious exact sequence $0 \to \mathcal{J} \to \mathcal{S} \to \mathcal{S}/\mathcal{J} \to 0$ yields

$$0 \longrightarrow \widetilde{\mathcal{J}} \xrightarrow{\sigma_E} \mathcal{O}_{Bl} \longrightarrow \tilde{\iota}_* \mathcal{O}_E \longrightarrow 0\ . \tag{23}$$

Compare (22). This means that here $I_E = \widetilde{\mathcal{J}}$. But now, $\mathcal{J}$ is obviously $\mathcal{S}(1)$ truncated in non-negative degrees. Since two graded $\mathcal{S}$-modules which coincide above some degree have the same associated sheaves, we have $I_E = \widetilde{\mathcal{J}} = \widetilde{\mathcal{S}(1)} = \mathcal{O}_{Bl}(1)$. Consequently, $\mathcal{O}(E) = (I_E)^\vee = \mathcal{O}_{Bl}(-1)$. In particular, we get

$$\mathcal{O}(E)_{|_E} = \mathcal{O}_{Bl}(-1)_{|_E} = \mathcal{O}_E(-1). \tag{24}$$

Proposition A.3 (Picard group in codimension one). *Let $B \in \mathcal{S}ch$ be a scheme and $\tilde{\iota} : E \hookrightarrow B$ be a regular closed immersion of codimension one of an irreducible subscheme $E \in \mathcal{S}ch$ with open complement $\tilde{\upsilon} : U \hookrightarrow B$. We then have a complex*

$$\mathbb{Z} \longrightarrow \mathrm{Pic}(B) \xrightarrow{\tilde{\upsilon}^*} \mathrm{Pic}(U) \tag{25}$$

where the first map sends 1 *to the line bundle* $\mathcal{O}(E)$ *associated to* E. *This complex is exact if* B *is normal, and* $\tilde{\upsilon}^*$ *is surjective when* B *is furthermore regular. It is also exact when* B *is the blow-up of a normal scheme* X *along a regular embedding.*

Proof. (25) is a complex since $\mathcal{O}(E)$ is trivial on U. When B is normal, $\mathrm{Pic}(B)$ injects in the group $\mathrm{Cl}(B)$ of Weil divisors (see [18, 7.1.19 and 7.2.14 (c)]), for which the same sequence holds by [15, Prop. II.6.5]. Exactness of (25) then follows by diagram chase. The surjectivity of $\tilde{\upsilon}^*$ when B is regular follows from [15, Prop. II.6.7 (c)]. When B is the blow-up of X along Z, we can assume that $\mathrm{codim}_X(Z) \geq 2$ by the previous point. Then, the result again follows by diagram chase, using that $\mathrm{Pic}(B) = \mathrm{Pic}(X) \oplus \mathbb{Z}$, as proved in Proposition A.6 (i) below. $\square$

Remark A.4. Note that the blow-up of a normal scheme along a regular closed embedding isn't necessarily normal if the subscheme is not reduced. For example, take $X = \mathbb{A}^2 = \mathrm{Spec}(k[x,y])$ and Z defined by the equations $x^2 = y^2 = 0$. Then, Bl is the subscheme of $\mathbb{A}^2 \times \mathbb{P}^1$ defined by the equations $x^2 v = y^2 u$ where $[u:v]$ are homogeneous coordinates for $\mathbb{P}^1$ and it is easy to check that the whole exceptional fiber is singular. Thus Bl is not normal (not even regular in codimension one).

Proposition A.5 (Picard group of a projective bundle). *Let* $X \in Sch$ *be a (connected) scheme and* $\mathcal{F}$ *a vector bundle over* X. *We consider the projective bundle* $\mathbb{P}_X(\mathcal{F})$ *associated to* $\mathcal{F}$. *Its Picard group is* $\mathrm{Pic}(X) \oplus \mathbb{Z}$ *where* $\mathbb{Z}$ *is generated by* $\mathcal{O}(-1)$ *and* $\mathrm{Pic}(X)$ *comes from the pull-back from* X.

Proof. Surjectivity of $\mathrm{Pic}(X) \oplus \mathbb{Z} \to \mathrm{Pic}(\mathbb{P}_X(\mathcal{F}))$ is a formal consequence of Quillen's formula [23, Prop. 4.3] for the K-theory of a projective bundle. Indeed, the determinant map $K_0 \to \mathrm{Pic}$ is surjective with an obvious set theoretic section and can easily be computed on each component of Quillen's formula. Injectivity is obtained by pulling back to the fiber of a point for the $\mathbb{Z}$ component, and by the projection formula for the remaining $\mathrm{Pic}(X)$ component. $\square$

Proposition A.6 (Picard group of a blow-up). *Under Setup 1.1, we have:*

(i) *The Picard group of* $Bl = Bl_Z(X)$ *is isomorphic to* $\mathrm{Pic}(X) \oplus \mathbb{Z}$ *where the direct summand* $\mathrm{Pic}(X)$ *comes from the pull-back* π^* *and* $\mathbb{Z}$ *is generated by the class of the exceptional divisor* $\mathcal{O}(E) = \mathcal{O}_{Bl}(-1)$.

(ii) *If* X *is normal, the map* $\upsilon^* : \mathrm{Pic}(X) \to \mathrm{Pic}(U)$ *is injective. If* X *is regular it is an isomorphism.*

(iii) *The exceptional fiber* E *is the projective bundle* $\mathbb{P}(C_{Z/X})$ *over* Z *and its Picard group is therefore* $\mathrm{Pic}(Z) \oplus \mathbb{Z}$ *where* $\mathbb{Z}$ *is generated by* $\mathcal{O}_E(-1)$.

(iv) *The pull-back* $\tilde{\iota}^* : \mathrm{Pic}(Bl) \to \mathrm{Pic}(E)$ *maps* $[\mathcal{O}(E)] \in \mathrm{Pic}(Bl)$ *to* $[\mathcal{O}_E(-1)]$.

Under these identifications, Diagram (1) *induces the following pull-back maps on Picard groups:*

$$\begin{tikzcd}
\operatorname{Pic}(Z) \arrow[d, "\binom{\mathrm{id}}{0}"'] & \operatorname{Pic}(X) \arrow[l, "\iota^*"'] \arrow[r, "\upsilon^*"] \arrow[d, "\binom{\mathrm{id}}{0}"] & \operatorname{Pic}(U)\,. \\
\operatorname{Pic}(Z)\oplus\mathbb{Z} & \operatorname{Pic}(X)\oplus\mathbb{Z} \arrow[l, "\left(\begin{smallmatrix}\iota^* & 0\\ 0 & \mathrm{id}\end{smallmatrix}\right)"'] \arrow[ur, "(\upsilon^*\ 0)"'] &
\end{tikzcd}$$

Proof. By Example A.2, we get (iv) and we can deduce (iii) from Proposition A.5. To prove (ii), use that for X normal (resp. regular) $\operatorname{Pic}(X)$ injects into (resp. is isomorphic to) the group $\operatorname{Cl}(X)$ of Weil divisors classes (see [18, 7.1.19 and 7.2.14 (c), resp. 7.2.16]), and that $\operatorname{Cl}(X)=\operatorname{Cl}(U)$ since $\operatorname{codim}_X(Z)\geq 2$. Finally, for (i), consider the commutative diagram

$$\begin{tikzcd}
\mathrm{D}_{\mathrm{perf}}(Z) \arrow[d, "\mathrm{L}\tilde{\pi}^*"'] & \mathrm{D}_{\mathrm{perf}}(X) \arrow[l, "\mathrm{L}\iota^*"'] \arrow[d, "\mathrm{L}\pi^*"] \\
\mathrm{D}_{\mathrm{perf}}(E) & \mathrm{D}_{\mathrm{perf}}(Bl) \arrow[l, "\mathrm{L}\tilde{\iota}^*"']
\end{tikzcd} \tag{26}$$

of induced functors on the derived categories of perfect complexes. We will use:

Fact 1: The tensor triangulated functors $\mathrm{L}\pi^*$ and $\mathrm{L}\tilde{\pi}^*$ are fully faithful with left inverse $\mathrm{R}\,\pi_*$ and $\mathrm{R}\,\tilde{\pi}_*$ respectively, see Thomason [24, Lemme 2.3].

Fact 2: If $M\in\mathrm{D}_{\mathrm{perf}}(Bl)$ is such that $\mathrm{L}\tilde{\iota}^*(M)\simeq\mathrm{L}\tilde{\pi}^*(N)$ for some $N\in\mathrm{D}_{\mathrm{perf}}(Z)$, then $M\simeq\mathrm{L}\pi^*(L)$ for some $L\in\mathrm{D}_{\mathrm{perf}}(X)$, which must then be $\mathrm{R}\,\pi_*(M)$ by Fact 1. This follows from [11, Prop. 1.5]. (In their notation, our assumption implies that M is zero in all successive quotients $\mathrm{D}^{i+1}_{\mathrm{perf}}(Bl)/\,\mathrm{D}^{i}_{\mathrm{perf}}(Bl)$ hence belongs to $\mathrm{D}^{0}_{\mathrm{perf}}(Bl)$.)

Hence $\operatorname{Pic}(X)\oplus\mathbb{Z}\to\operatorname{Pic}(Bl)$ is injective: If L is a line bundle on X and $n\in\mathbb{Z}$ are such that $\mathrm{L}\pi^*(L)\otimes\mathcal{O}_{Bl}(n)$ is trivial then we get $n=0$ by restricting to E and applying (iii), and we get $L\simeq\mathrm{R}\,\pi_*\,\mathrm{L}\pi^*L\simeq\mathrm{R}\,\pi_*\mathcal{O}_{Bl}\simeq\mathrm{R}\,\pi_*\,\mathrm{L}\pi^*\mathcal{O}_X\simeq\mathcal{O}_X$ by Fact 1. So, let us check surjectivity of $\operatorname{Pic}(X)\oplus\mathbb{Z}\to\operatorname{Pic}(Bl)$. Let M be a line bundle on Bl. Using (iii) again and twisting with $\mathcal{O}_{Bl}(n)$ if necessary, we can assume that $\mathrm{L}\tilde{\iota}^*(M)$ is isomorphic to $\mathrm{L}\tilde{\pi}^*N=\tilde{\pi}^*N$ for some line bundle N on Z. By Fact 2, there exists $L\in\mathrm{D}_{\mathrm{perf}}(X)$ such that $\mathrm{L}\pi^*(L)\simeq M$. It now suffices to check that this $L\in\mathrm{D}_{\mathrm{perf}}(X)$ is a line bundle. The natural (evaluation) map $L^\vee\otimes L\to\mathcal{O}_X$ is an isomorphism, since it is so after applying the fully faithful tensor functor $\mathrm{L}\pi^*:\mathrm{D}_{\mathrm{perf}}(X)\to\mathrm{D}_{\mathrm{perf}}(Bl)$. So $L\in\mathrm{D}_{\mathrm{perf}}(X)$ is an invertible object, hence it is the m^{th} suspension of a line bundle for $m\in\mathbb{Z}$, see [7, Prop. 6.4]. Using (26), one checks by restricting to Z that $m=0$, i.e. L is a line bundle. $\square$

* * *

We now discuss dualizing complexes and relative canonical bundles. First of all, we mention the essential uniqueness of dualizing complexes on a scheme.

Lemma A.7. *Let $X \in \mathcal{S}ch$ be a scheme admitting a dualizing complex K_X. Then:*

(i) For any line bundle L and any integer i, the complex $K_X \otimes L[i]$ is also a dualizing complex and any dualizing complex on X is of this form.

(ii) If $K_X \otimes L[i] \simeq K_X$ in the derived category of X, for some line bundle L and some integer i, then $L \simeq \mathcal{O}_X$ and $i = 0$.

In other words, the set of isomorphism classes of dualizing complexes on X is a principal homogeneous space under the action of $\mathrm{Pic}(X) \oplus \mathbb{Z}$.

Proof. For (i), see [19, Lemma 3.9]. Let us prove (ii). We have the isomorphisms

$$\begin{aligned}\mathcal{O}_X \xrightarrow{\sim} \mathrm{RHom}(K_X, K_X) \simeq \mathrm{RHom}(K_X, K_X \otimes L[n])\\ \simeq \mathrm{RHom}(K_X, K_X) \otimes L[n] \xleftarrow{\sim} L[n]\end{aligned}$$

in the coherent derived category. The first and last ones hold by [19, Prop. 3.6]. We thus obtain an isomorphism $\mathcal{O}_X \simeq L[n]$ in the derived category of perfect complexes (it is a full subcategory of the coherent one). This forces $n = 0$ and the existence of an honest isomorphism of sheaves $\mathcal{O}_X \simeq L$, see [7, Prop. 6.4] if necessary. ☐

We now use the notion of local complete intersection (l.c.i.) morphism, that is, a morphism which is locally a regular embedding followed by a smooth morphism, see [18, § 6.3.2]. The advantage of such morphisms $f : X' \to X$ is that $f^!$ is just $\mathrm{L}f^*$ twisted by a line bundle ω_f and shifted by the relative dimension $\dim(f)$.

Proposition A.8. *Let $f : X' \to X$ be an l.c.i. morphism with $X, X' \in \mathcal{S}ch$. Assume that f is proper. Then $f^!(\mathcal{O}_X)$ is a shifted line bundle $\omega_f[\dim(f)]$ and there exists a natural isomorphism $f^!(\mathcal{O}_X) \otimes \mathrm{L}f^*(-) \xrightarrow{\sim} f^!(-)$. In particular, $f^!$ preserves the subcategory $\mathrm{D}_{\mathrm{perf}}$ of $\mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}$.*

Proof. There is always a natural morphism $f^!(\mathcal{O}_X) \otimes \mathrm{L}f^*(-) \to f^!(-)$. One shows that it is an isomorphism and that $f^!(\mathcal{O}_X)$ is a line bundle directly from the definition, since both these facts can be checked locally, are stable by composition and are true for (closed) regular immersions and smooth morphisms by Hartshorne [14, Ch. III]. The subcategory $\mathrm{D}_{\mathrm{perf}}$ is then preserved since both $\mathrm{L}f^*$ and tensoring by a line bundle preserve it. ☐

The above proposition reduces the description of $f^!$ to that of the line bundle ω_f.

PROPOSITION A.9. *In the following cases, we have concrete descriptions of* ω_f.

(i) *When* $f : X' \to X$ *is smooth and proper,* $\omega_f \simeq \det(\Omega^1_{X'/X})$ *is the determinant of the sheaf of differentials. In particular, when* f *is the projection of a projective bundle* $\mathbb{P}(\mathcal{F})$ *to its base, where* $\mathcal{F}$ *is a vector bundle of rank* r*, then* $\omega_f \simeq f^*(\det\mathcal{F}) \otimes \mathcal{O}_{\mathbb{P}(\mathcal{F})}(-r)$.

(ii) *When* $f : X' \hookrightarrow X$ *is a regular closed immersion,* $\omega_f \simeq \det(N_{X'/X})$ *is the determinant of the normal bundle. In particular when* $f : E \hookrightarrow B$ *is the inclusion of a prime divisor (Def. A.1), we have* $\omega_f \simeq \mathcal{O}(E)_{|_E}$.

Proof. See [25, Prop. 1 and Thm. 3]. See alternatively [18, § 6.4.2]. □

REMARK A.10. All morphisms along which we consider push-forward in this article are l.c.i. It might not be obvious for $\pi : Bl \to X$ but this follows from [1, VII 1.8 p. 424] (it is locally of the form mentioned there). So, ω_π is also a line bundle. Let us now describe the relative canonical line bundles in terms of $\omega_\iota = \det(N_{Z/X})$.

PROPOSITION A.11. *With the notation of Setup 1.1, we have*

(i) $\omega_{\tilde\iota} = \mathcal{O}(E)_{|_E} = \mathcal{O}_E(-1)$

(ii) $\omega_{\tilde\pi} = \tilde\pi^*\omega_\iota^\vee \otimes \mathcal{O}(E)^{\otimes c}_{|_E} = \tilde\pi^*\omega_\iota^\vee \otimes \mathcal{O}_E(-c)$

(iii) $\omega_\pi = \mathcal{O}(E)^{\otimes(c-1)} = \mathcal{O}_{Bl}(1-c)$.

By Proposition A.8, it implies that we have

(iv) $\tilde\iota^!(-) = \mathcal{O}(E)_{|_E} \otimes \mathrm{L}\tilde\iota^*(-)[-1] = \mathcal{O}_E(-1) \otimes \mathrm{L}\tilde\iota^*(-)[-1]$

(v) $\tilde\pi^!(-) = \tilde\pi^*\omega_\iota^\vee \otimes \mathcal{O}(E)^{\otimes c}_{|_E} \otimes \mathrm{L}\tilde\pi^*(-)[c-1] = \tilde\pi^*\omega_\iota^\vee \otimes \mathcal{O}_E(-c) \otimes \mathrm{L}\tilde\pi^*(-)[c-1]$

(vi) $\pi^!(-) = \mathcal{O}(E)^{\otimes(c-1)} \otimes \mathrm{L}\pi^*(-) = \mathcal{O}_{Bl}(1-c) \otimes \mathrm{L}\pi^*(-)$.

Proof. Points (i) and (ii) follow from Proposition A.9 (ii) and (i), respectively. They imply (iv) and (v). To prove point (iii) let us first observe that the exact sequence (22) gives rise to an exact triangle

$$\mathcal{O}_{Bl}(l+1) \to \mathcal{O}_{Bl}(l) \to \mathrm{R}\,\tilde\iota_*(\mathcal{O}_E(l)) \to \mathcal{O}_{Bl}(l+1)[1]$$

in $\mathrm{D}_{\mathrm{perf}}(Bl)$ for any $l \in \mathbb{Z}$. Applying $\mathrm{R}\,\pi_*$ to this triangle and using that

$$\mathrm{R}\,\pi_*\,\mathrm{R}\,\tilde\iota_*\mathcal{O}_E(l) = \mathrm{R}\,\iota_*\,\mathrm{R}\,\tilde\pi_*\mathcal{O}_E(l) = 0 \quad \text{for} \quad -c < l < 0$$

(by [13, 2.1.15]), we obtain by induction that $\mathrm{R}\,\pi_*\mathcal{O}_{Bl}(l) = \mathrm{R}\,\pi_*\mathcal{O}_{Bl} = \mathcal{O}_X$ for $-c < l \leq 0$. In particular $\mathrm{R}\,\pi_*\mathcal{O}_{Bl}(1-c) = \mathcal{O}_X$. We now use the filtration of [11, Prop. 1.5]. Let us show that $\pi^!(\mathcal{O}_X) \otimes \mathcal{O}_{Bl}(c-1)$ is in $\mathrm{D}^0_{\mathrm{perf}}(Bl)$. By loc. cit. it suffices to show that $\mathrm{L}\tilde{\iota}^*(\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_{Bl}(c-1))$ is in $\mathrm{D}^0_{\mathrm{perf}}(E)$. It follows from the sequence of isomorphisms

$$\mathrm{L}\tilde{\iota}^*(\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_{Bl}(c-1))\simeq\mathrm{L}\tilde{\iota}^*\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_E(c-1)\overset{\text{(iv)}}{\simeq}\tilde{\iota}^!\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_E(c)[1]$$

$$\simeq\tilde{\pi}^!\iota^!(\mathcal{O}_X)\otimes\mathcal{O}_E(c)[1]\overset{\text{A.8}}{\simeq}\tilde{\pi}^!(\omega_\iota)\otimes\mathcal{O}_E(c)[-c+1]\overset{\text{(v)}}{\simeq}\mathcal{O}_E\,.$$

Since $L:=\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_{Bl}(c-1)$ is in $\mathrm{D}^0_{\mathrm{perf}}(Bl)$, it is of the form $\mathrm{L}\pi^*M$ for $M=\mathrm{R}\,\pi_*L$ (by Fact 1 in the proof of Prop. A.6) which we compute by duality:

$$\mathrm{R}\,\pi_*(\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_{Bl}(c-1))\simeq\mathrm{R}\,\pi_*\mathrm{RHom}(\mathcal{O}_{Bl},\pi^!(\mathcal{O}_X)\otimes\mathcal{O}_{Bl}(c-1))\simeq$$

$$\simeq\mathrm{R}\,\pi_*\mathrm{RHom}(\mathcal{O}_{Bl}(1-c),\pi^!(\mathcal{O}_X))\overset{(\dagger)}{\simeq}\mathrm{RHom}(\mathrm{R}\,\pi_*\mathcal{O}_{Bl}(1-c),\mathcal{O}_X)\overset{(\star)}{\simeq}$$

$$\simeq\mathrm{RHom}(\mathcal{O}_X,\mathcal{O}_X)\simeq\mathcal{O}_X$$

where $(\star)$ is by the computation at the beginning of the proof, $(\dagger)$ is the duality isomorphism and all other isomorphisms are obtained as consequences of the monoidal structure on the $\mathrm{D}^{\mathrm{b}}_{\mathrm{coh}}$ involved (see [10] and [8] for details). Hence, $\pi^!(O_X)\simeq\mathcal{O}_{Bl}(1-c)$ as announced. This proves (iii) and thus (vi). ☐

Finally, we also use dualizing complexes in the context of an $\mathbb{A}^*$-bundle $U\to Y$, i.e. a morphism that is locally of the form $\mathbb{A}^n_Y\to Y$ (and is in particular flat).

Lemma A.12. *Let $\alpha:U\to Y$ be an $\mathbb{A}^*$-bundle. Assume that Y admits a dualizing complex K_Y. Then $\mathrm{L}\alpha^*(K_Y)=\alpha^*(K_Y)$ is a dualizing complex on U.*

Proof. This can be checked locally, so we can assume α decomposes as $\alpha=f\circ u$ for an open immersion $u:\mathbb{A}^n_Y\hookrightarrow\mathbb{P}^n_Y$ followed by the structural projection $f:\mathbb{P}^n_Y\to Y$. Note that α, u and f are all flat. We have by Propositions A.8 and A.9 (i)

$$u^*f^!K_Y[n]=u^*(\mathcal{O}(-n-1)\otimes f^*K_Y)\simeq u^*f^*K_Y=\alpha^*K_Y$$

where the second equality comes from the triviality of $\mathcal{O}(-n-1)$ on $\mathbb{A}^n_Y$. Now $u^*f^!K_Y[n]$ is dualizing because proper morphisms, open immersions and shifting preserve dualizing complexes. ☐

Acknowledgments : We thank Marc Levine for precious discussions on Theorem 1.3 and Example 1.5, Jean Fasel for a florilegium of relative bundles and Burt Totaro for several discussions on algebraic geometry.

References

[1] *Théorie des intersections et théorème de Riemann-Roch*, Séminaire de Géométrie Algébrique du Bois-Marie 1966–1967 (SGA 6), dirigé par P. Berthelot, A. Grothendieck et L. Illusie. Lecture Notes in Mathematics, Vol. 225. Springer-Verlag, Berlin, 1971.

[2] J. Arason, *Der Wittring projektiver Räume*, Math. Ann. 253 (1980), no. 3, 205–212.

[3] P. Balmer, *Triangular Witt Groups Part 1: The 12-Term Localization Exact Sequence*, K-Theory 4 (2000), no. 19, 311–363.

[4] ______, *Products of degenerate quadratic forms*, Compos. Math. 141 (2005), no. 6, 1374–1404.

[5] ______, *Witt groups*, Handbook of K-theory. Vol. 2, Springer, Berlin, 2005, pp. 539–576.

[6] P. Balmer and B. Calmès, *Witt groups of Grassmann varieties*, preprint, 2008.

[7] P. Balmer and G. Favi, *Gluing techniques in triangular geometry*, Q J Math 58 (2007), no. 4, 415–441.

[8] B. Calmès and J. Hornbostel, *Push-forwards for Witt groups of schemes*, arXiv:0806.0571, 2008, to appear in Comment. Math. Helv.

[9] ______, *Witt motives, transfers and dévissage*, preprint, available at `http://www.math.uiuc.edu/K-theory/0786/`, 2006.

[10] B. Calmes and J. Hornbostel, *Tensor-triangulated categories and dualities*, Theory Appl. Categ. 22 (2009), No. 6, 136–198.

[11] G. Cortiñas, C. Haesemeyer, M. Schlichting, and C. Weibel, *Cyclic homology, cdh-cohomology and negative K-theory*, Ann. of Math. 167 (2008), 549–573.

[12] S. Gille, *Homotopy invariance of coherent Witt groups*, Math. Z. 244 (2003), no. 2, 211–233.

[13] A. Grothendieck, *Éléments de géométrie algébrique. III. Étude cohomologique des faisceaux cohérents. I*, Inst. Hautes Études Sci. Publ. Math. (1961), no. 11, 167.

[14] R. Hartshorne, *Residues and duality*, Lecture notes of a seminar on the work of A. Grothendieck, given at Harvard 1963/64. With an appendix by P. Deligne. Lecture Notes in Mathematics, No. 20, Springer-Verlag, Berlin, 1966.

[15] ______, *Algebraic geometry*, Springer-Verlag, New York, 1977, Graduate Texts in Mathematics, No. 52.

[16] D. Laksov, *Algebraic Cycles on Grassmann Varieties*, Adv. in Math. 9 (1972), no. 3, 267–295.

[17] M. Levine and F. Morel, *Algebraic cobordism*, Springer Monographs in Mathematics, Springer, Berlin, 2007.

[18] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics, vol. 6, Oxford University Press, Oxford, 2002.

[19] A. Neeman, *Derived categories and Grothendieck duality*, preprint 791 of `http://www.crm.es/Publications/Preprints08.htm`.

[20] A. Nenashev, *Projective push-forwards in the Witt theory of algebraic varieties*, Adv. Math. 220 (2009), no. 6, 1923–1944.

[21] A. Nenashev and K. Zainoulline, *Oriented cohomology and motivic decompositions of relative cellular spaces*, J. Pure Appl. Algebra 205 (2006), no. 2, 323–340.

[22] I. Panin, *Riemann-Roch theorems for oriented cohomology*, Axiomatic, enriched and motivic homotopy theory, NATO Sci. Ser. II Math. Phys. Chem., vol. 131, Kluwer Acad. Publ., Dordrecht, 2004, pp. 261–333.

[23] D. Quillen, *Higher algebraic K-theory: I*, Algebraic K-theory, Lecture Notes in Math., no. 341, Springer-Verlag, 1973, pp. 83–147.

[24] R. W. Thomason, *Les K-groupes d'un schéma éclaté et une formule d'intersection excédentaire*, Invent. Math. 112 (1993), no. 1, 195–215.

[25] J.-L. Verdier, *Base change for twisted inverse image of coherent sheaves*, Algebraic Geometry (Internat. Colloq., TIFR, Bombay, 1968), Oxford Univ. Press, London, 1969, pp. 393–408.

[26] C. Walter, *Grothendieck-Witt groups of projective bundles*, preprint, 2003.

Paul Balmer
Department of Mathematics
UCLA
Los Angeles, CA 90095-1555
USA
balmer@math.ucla.edu

Baptiste Calmès
LML
Fac. des Sciences J. Perrin
Université d'Artois
62307 Lens
France
calmes@math.jussieu.fr

Motivic Landweber Exactness

Niko Naumann, Markus Spitzweck, Paul Arne Østvær

Received: November 2, 2008

Communicated by Stefan Schwede

Abstract. We prove a motivic Landweber exact functor theorem. The main result shows the assignment given by a Landweber-type formula involving the MGL-homology of a motivic spectrum defines a homology theory on the motivic stable homotopy category which is representable by a Tate spectrum. Using a universal coefficient spectral sequence we deduce formulas for operations of certain motivic Landweber exact spectra including homotopy algebraic K-theory. Finally we employ a Chern character between motivic spectra in order to compute rational algebraic cobordism groups over fields in terms of rational motivic cohomology groups and the Lazard ring.

2000 Mathematics Subject Classification: 55N22, 55P42, 14A20, 14F42, 19E08.
Keywords and Phrases:

Contents

1 Introduction

The Landweber exact functor theorem combined with Brown representability provides an almost unreasonably efficient toolkit for constructing homotopy types out of purely algebraic data. Among the many examples arising this way is the presheaf of elliptic homology theories on the moduli stack of elliptic curves. In this paper we incite the use of such techniques in the algebro-geometric setting of motivic homotopy theory.

In what follows we shall state some of the main results in the paper, comment on the proofs and discuss some of the background and relation to previous works. Throughout we employ a stacky viewpoint of the subject which originates with formulations in stable homotopy theory pioneered by Morava and Hopkins. Let S be a regular noetherian base scheme of finite Krull dimension and $\mathbf{SH}(S)$ the corresponding motivic stable homotopy category. A complex point $\mathsf{Spec}(\mathbf{C}) \to S$ induces a functor $\mathbf{SH}(S) \to \mathbf{SH}$ to the classical stable homotopy category. Much of the work in this paper is guidelined by the popular quest of hoisting results in $\mathbf{SH}$ to the more complicated motivic category.

To set the stage, denote by MGL the algebraic cobordism spectrum introduced by Voevodsky [39]. By computation we show $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$ is a flat Hopf algebroid in Adams graded abelian groups. (Our standard conventions concerning graded objects are detailed in Section 3. Recall that $\mathsf{MGL}_* \equiv \mathsf{MGL}_{2*,*}$.) The useful fact that MGL gives rise to an algebraic stack $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ comes to bear. (This apparatus is reviewed in Section 2.) By comparing with the complex cobordism spectrum MU we deduce a 2-categorical commutative diagram:

$$\begin{tikzcd} \mathsf{Spec}(\mathsf{MGL}_*) \arrow[r] \arrow[d] & \mathsf{Spec}(\mathsf{MU}_*) \arrow[d] \\ {[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]} \arrow[r] & {[\mathsf{MU}_*/\mathsf{MU}_*\mathsf{MU}]} \end{tikzcd} \tag{1}$$

The right hand part of the diagram is well-known: Milnor's computation of MU_* and Quillen's identification of the canonical formal group law over MU_* with the universal formal group law are early success stories in modern algebraic topology. As a $\mathbf{G}_{\mathfrak{m}}$-stack the lower right hand corner identifies with the moduli stack of strict graded formal groups. Our plan from the get-go was to prove (1) is cartesian and use that description of the algebraic cobordism part of the diagram to deduce motivic analogs of theorems in stable homotopy theory. It turns out this strategy works for general base schemes.

Recall that an MU_*-module M_* is Landweber exact if $v_0^{(p)}, v_1^{(p)}, \ldots$ forms a regular sequence in M_* for every prime p. Here $v_0^{(p)} = p$ and the $v_i^{(p)}$ for $i > 0$ are indecomposable elements of degree $2p^i - 2$ in MU_* with Chern numbers divisible by p. Using the cartesian diagram (1) we show the following result for Landweber exact motivic homology theories, see Theorem 7.3 for a more precise statement.

Theorem: *Suppose A_* is a Landweber exact graded MU_*-algebra. Then*

$$\mathsf{MGL}_{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{A}_*$$

is a bigraded ring homology theory on $\mathbf{SH}(S)$.

Using the theorem we deduce that

$$\mathsf{MGL}^{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{A}_*$$

is a ring cohomology theory on the subcategory of strongly dualizable objects of $\mathbf{SH}(S)$. In the case of the Laurent polynomial ring $\mathbf{Z}[\beta, \beta^{-1}]$ on the Bott element, this observation forms part of the proof in [35] of the motivic Conner-Floyd isomorphism

$$\mathsf{MGL}^{**}(-) \otimes_{\mathsf{MU}_*} \mathbf{Z}[\beta, \beta^{-1}] \xrightarrow{\cong} \mathsf{KGL}^{**}(-)$$

for the motivic spectrum KGL representing homotopy algebraic K-theory.

Define the category of Tate objects $\mathbf{SH}(S)_{\mathcal{T}}$ as the smallest localizing triangulated subcategory of the motivic stable homotopy category containing the set $\mathcal{T}$ of all mixed motivic spheres

$$\mathsf{S}^{p,q} \equiv S_s^{p-q} \wedge \mathbf{G}_{\mathfrak{m}}^q$$

of smash products of the simplicial circle S^1_s and the multiplicative group scheme $\mathbf{G}_{\mathfrak{m}}$. The Tate objects are precisely the cellular spectra in the terminology of [7]. Our choice of wording is deeply rooted in the theory of motives. Since the inclusion $\mathbf{SH}(S)_{\mathcal{T}} \subseteq \mathbf{SH}(S)$ preserves sums and $\mathbf{SH}(S)$ is compactly generated, a general result for triangulated categories shows that it acquires a right adjoint functor $\mathsf{p}_{\mathbf{SH}(S),\mathcal{T}} \colon \mathbf{SH}(S) \to \mathbf{SH}(S)_{\mathcal{T}}$, which we call the Tate projection. When E is a Tate object and F a motivic spectrum there is thus an isomorphism

$$\mathsf{E}_{**}(\mathsf{F}) \cong \mathsf{E}_{**}(\mathsf{p}_{\mathbf{SH}(S),\mathcal{T}}\mathsf{F}).$$

As in topology, it follows that the E_{**}-homology of F is determined by the E_{**}-homology of mixed motivic spheres. This observation is a key input in showing $(\mathsf{E}_*, \mathsf{E}_*\mathsf{E})$ is a flat Hopf algebroid in Adams graded abelian groups provided one - and hence both - of the canonical maps $\mathsf{E}_{**} \to \mathsf{E}_{**}\mathsf{E}$ is flat and the canonical map $\mathsf{E}_*\mathsf{E} \otimes_{\mathsf{E}_*} \mathsf{E}_{**} \to \mathsf{E}_{**}\mathsf{E}$ is an isomorphism. Specializing to the example of algebraic cobordism allows us to form the algebraic stack $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ and (1).

Our motivic analog of Landweber's exact functor theorem takes the following form, see Theorem 8.7.

Theorem: *Suppose M_* is an Adams graded Landweber exact MU_*-module. Then there exists a motivic spectrum E in $\mathbf{SH}(S)_{\mathcal{T}}$ and a natural isomorphism*

$$\mathsf{E}_{**}(-) \cong \mathsf{MGL}_{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{M}_*$$

of homology theories on $\mathbf{SH}(S)$.

In addition, if M_* is a graded MU_*-algebra then E acquires a quasi-multiplication which represents the ring structure on the corresponding Landweber exact theory.

When the base scheme is the integers $\mathbf{Z}$ we use motivic Landweber exactness and Voevodsky's result that $\mathbf{SH}(\mathbf{Z})_{\mathcal{T}}$ is a Brown category [39], so that all homology theories are representable, to conclude the proof of the motivic exact functor theorem. For more details and a proof of the fact that $\mathbf{SH}(\mathbf{Z})_{\mathcal{T}}$ is a Brown category we refer to [26]. For a general base scheme we provide base change results which allow us to reduce to the case of the integers. The subcategory of Tate objects of the derived category of modules over MGL - relative to $\mathbf{Z}$ - turns also out to be a Brown category. This suffices to show the above remains valid when translated verbatim to the setting of highly structured MGL-modules. Recall MGL is a motivic symmetric spectrum and the monoid axiom introduced in [33] holds for the motivic stable structure [17, Proposition 4.19]. Hence the modules over MGL acquire a closed symmetric monoidal model structure. Moreover, for every cofibrant replacement of MGL

in commutative motivic symmetric ring spectra there is a Quillen equivalence between the corresponding module categories.
We wish to emphasize the close connection between our results and the classical Landweber exact functor theorem. In particular, if M_* is concentrated in even degrees there exists a commutative ring spectrum $\mathsf{E}^{\mathsf{Top}}$ in $\mathbf{SH}$ which represents the corresponding topological Landweber exact theory. Although E and $\mathsf{E}^{\mathsf{Top}}$ are objects in widely different categories of spectra, it turns out there is an isomorphism

$$\mathsf{E}_{**}\mathsf{E} \cong \mathsf{E}_{**} \otimes_{\mathsf{E}_*^{\mathsf{Top}}} \mathsf{E}_*^{\mathsf{Top}}\mathsf{E}^{\mathsf{Top}}.$$

In the last part of the paper we describe (co)operations and phantom maps between Landweber exact motivic spectra. Using a universal coefficient spectral sequence we show that every MGL-module E gives rise to a surjection

$$\mathsf{E}^{p,q}(\mathsf{M}) \longrightarrow \mathsf{Hom}^{p,q}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{M}, \mathsf{E}_{**}), \tag{2}$$

and the kernel of (2) identifies with the Ext-term

$$\mathsf{Ext}^{1,(p-1,q)}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{M}, \mathsf{E}_{**}). \tag{3}$$

Imposing the assumption that $\mathsf{E}_*^{\mathsf{Top}}\mathsf{E}^{\mathsf{Top}}$ be a projective $\mathsf{E}_*^{\mathsf{Top}}$-module implies the given Ext-term in (3) vanishes, and hence (2) is an isomorphism. The assumption on $\mathsf{E}^{\mathsf{Top}}$ holds for unitary topological K-theory KU and localizations of Johnson-Wilson theories. By way of example we compute the KGL-cohomology of KGL. That is, using the completed tensor product we show there is an isomorphism of KGL^{**}-algebras

$$\mathsf{KGL}^{**}\mathsf{KGL} \xrightarrow{\cong} \mathsf{KGL}^{**}\widehat{\otimes}_{\mathsf{KU}^*}\mathsf{KU}^*\mathsf{KU}.$$

By [2] the group $\mathsf{KU}^1\mathsf{KU}$ is trivial and $\mathsf{KU}^0\mathsf{KU}$ is uncountable. We also show that KGL does not support any nontrivial phantom map. Adopting the proof to $\mathbf{SH}$ reproves the analogous result for KU. The techniques we use can further be utilized to construct a Chern character in $\mathbf{SH}(S)$ from KGL to the periodized rational motivic Eilenberg-MacLane spectrum representing rational motivic cohomology. For smooth schemes over fields we prove there is an isomorphism between rational motivic cohomology $\mathsf{M}\mathbf{Q}$ and the Landweber spectrum representing the additive formal group law over $\mathbf{Q}$. This leads to explicit computations of rational algebraic cobordism groups, cf. Corollary 10.6.

Theorem: *If X is a smooth scheme over a field and L^* denotes the (graded) Lazard ring, then there is an isomorphism*

$$\mathsf{MGL}^{**}(X) \otimes_{\mathbf{Z}} \mathbf{Q} \cong \mathsf{M}\mathbf{Q}^{**}(X) \otimes_{\mathbf{Z}} \mathsf{L}^*.$$

For finite fields it follows that $\mathsf{MGL}^{2**} \otimes_{\mathbf{Z}} \mathbf{Q} \cong \mathbf{Q} \otimes_{\mathbf{Z}} \mathsf{L}^*$ and $\mathsf{MGL}^{**} \otimes_{\mathbf{Z}} \mathbf{Q}$ is the trivial group if $(*, *) \notin \mathbf{Z}(2,1)$. Number fields provide other examples for which $\mathsf{MGL}^{**} \otimes_{\mathbf{Z}} \mathbf{Q}$ can now be computed explicitly (in terms of the number of real and complex embeddings). The theorem suggests the spectral sequence associated to the slice tower of the algebraic cobordism spectrum takes the expected form, and that it degenerates rationally, cf. the works of Hopkins-Morel reviewed in [19] and Voevodsky [41].

Inspired by the results herein we make some rather speculative remarks concerning future works. The all-important chromatic approach to stable homotopy theory acquires deep interplays with the algebraic geometry of formal groups. Landweber exact algebras over Hopf algebroids represent a central theme in this endeavor, leading for example to the bicomplete closed symmetric monoidal abelian category of $\mathsf{BP}_*\mathsf{BP}$-comodules. The techniques in this paper furnish a corresponding Landweber exact motivic Brown-Peterson spectrum MBP equivalent to the constructions in [16] and [38]. The object $\mathsf{MBP}_*\mathsf{MBP}$ and questions in motivic chromatic theory at large can be investigated along the lines of this paper. An exact analog of Bousfield's localization machinery in motivic stable homotopy theory was worked out in [32, Appendix A], cf. also [13] for a discussion of the chromatic viewpoint. In a separate paper [27] we dispense with the regularity assumption on S. The results in this paper remain valid for noetherian base schemes of finite Krull dimension. Since this generalization uses arguments which are independent of the present work, we deferred it to *loc. cit.* The slices of motivic Landweber spectra are studied in [34] by the third author.

Acknowledgments. We wish to thank J. Hornbostel, O. Röndigs and the referee for helpful comments on this paper.

2 Preliminaries on algebraic stacks

By a stack we shall mean a category fibered in groupoids over the site comprised by the category of commutative rings endowed with the fpqc-topology. A stack $\mathfrak{X}$ is algebraic if its diagonal is representable and affine, and there exists an affine scheme U together with a faithfully flat map $U \to \mathfrak{X}$, called a presentation of $\mathfrak{X}$. We refer to [12], [25] and [11] for motivation and basic properties of these notions.

Lemma 2.1: *Suppose there are 2-commutative diagrams of algebraic stacks*

$$\begin{array}{ccc} \mathfrak{Z} & \longrightarrow & \mathfrak{Z}' \\ \downarrow & & \downarrow \\ \mathfrak{X} & \longrightarrow & \mathfrak{X}' \end{array} \qquad \begin{array}{ccc} \mathfrak{Y} & \longrightarrow & \mathfrak{Y}' \\ {\scriptstyle\pi}\downarrow & & \downarrow \\ \mathfrak{X} & \longrightarrow & \mathfrak{X}' \end{array} \tag{4}$$

where π is faithfully flat. Then the left hand diagram in (4) is cartesian if and only if the naturally induced commutative diagram

$$\begin{tikzcd} \mathfrak{Z} \times_{\mathfrak{X}} \mathfrak{Y} \arrow[r] \arrow[d] & \mathfrak{Z}' \times_{\mathfrak{X}'} \mathfrak{Y}' \arrow[d] \\ \mathfrak{Y} \arrow[r] & \mathfrak{Y}' \end{tikzcd} \tag{5}$$

is cartesian.

Proof. The base change of the canonical 1-morphism $\mathfrak{c} : \mathfrak{Z} \to \mathfrak{Z}' \times_{\mathfrak{X}'} \mathfrak{X}$ over $\mathfrak{X}$ along π identifies with the canonically induced 1-morphism

$$\mathfrak{Z} \times_{\mathfrak{X}} \mathfrak{Y} \xrightarrow{\mathfrak{c}\times 1} (\mathfrak{Z}' \times_{\mathfrak{X}'} \mathfrak{X}) \times_{\mathfrak{X}} \mathfrak{Y} \cong \mathfrak{Z}' \times_{\mathfrak{X}'} \mathfrak{Y} \cong (\mathfrak{Z}' \times_{\mathfrak{X}'} \mathfrak{Y}') \times_{\mathfrak{Y}'} \mathfrak{Y}.$$

This is an isomorphism provided (5) is cartesian; hence so is $\mathfrak{c} \times 1$. By faithful flatness of π it follows that $\mathfrak{c}$ is an isomorphism. The reverse implication holds trivially. $\square$

Corollary 2.2: *Suppose $\mathfrak{X}$ and $\mathfrak{Y}$ are algebraic stacks, $U \to \mathfrak{X}$ and $V \to \mathfrak{Y}$ are presentations and there is a 2-commutative diagram:*

$$\begin{tikzcd} U \arrow[r] \arrow[d] & V \arrow[d] \\ \mathfrak{X} \arrow[r] & \mathfrak{Y} \end{tikzcd} \tag{6}$$

Then (6) is cartesian if and only if one - and hence both - of the commutative diagrams ($i = 1, 2$)

$$\begin{tikzcd} U \times_{\mathfrak{X}} U \arrow[r] \arrow[d, "pr_i"'] & V \times_{\mathfrak{Y}} V \arrow[d, "pr_i"] \\ U \arrow[r] & V \end{tikzcd} \tag{7}$$

is cartesian.

Proof. Follows from Lemma 2.1 since presentations are faithfully flat. $\square$

A presentation $U \to \mathfrak{X}$ yields a Hopf algebroid or cogroupoid object in commutative rings $(\Gamma(\mathcal{O}_U), \Gamma(\mathcal{O}_{U\times_{\mathfrak{X}} U}))$. Conversely, if (A, B) is a flat Hopf algebroid, denote by $[\mathsf{Spec}(A)/\mathsf{Spec}(B)]$ the associated algebraic stack. We note that by [25, Theorem 8] there is an equivalence of 2-categories between flat Hopf algebroids and presentations of algebraic stacks.

Let $\mathsf{Qc}_{\mathfrak{X}}$ denote the category of quasi-coherent $\mathcal{O}_{\mathfrak{X}}$-modules and $\mathcal{A} \in \mathsf{Qc}_{\mathfrak{X}}$ a monoid, or quasi-coherent sheaf of $\mathcal{O}_{\mathfrak{X}}$-algebras. If X_0 is a scheme and

$\pi : X_0 \to \mathfrak{X}$ faithfully flat, then $\mathcal{A}$ is equivalent to the datum of the $\mathcal{O}_{X_0}$-algebra $\mathcal{A}(X_0) \equiv \pi^*\mathcal{A}$ combined with a descent datum with respect to $X_1 \equiv X_0 \times_{\mathfrak{X}} X_0 \rightrightarrows X_0$. When $X_0 = \mathsf{Spec}(A)$ is affine, $X_1 = \mathsf{Spec}(\Gamma)$ is affine, (A, Γ) a flat Hopf algebroid and $\mathcal{A}(X_0)$ a Γ-comodule algebra.
Denote the adjunction between left $\mathcal{A}$-modules in $\mathsf{Qc}_{\mathfrak{X}}$ and left $\mathcal{A}(X_0)$-modules in Qc_{X_0} by:

$$\pi^* : \mathcal{A} - \mathsf{mod} \rightleftarrows \mathcal{A}(X_0) - \mathsf{mod} : \pi_*$$

Since π_* has an exact left adjoint π^* it preserves injectives and there are isomorphisms

$$\mathsf{Ext}^n_{\mathcal{A}}(\mathcal{M}, \pi_*\mathcal{N}) \cong \mathsf{Ext}^n_{\mathcal{A}(X_0)}(\pi^*\mathcal{M}, \mathcal{N}) \tag{8}$$

between Ext-groups in the categories of quasi-coherent left $\mathcal{A}$- and $\mathcal{A}(X_0)$-modules.
Now assume that $i : \mathcal{U} \hookrightarrow \mathfrak{X}$ is the inclusion of an open algebraic substack. Then [25, Propositions 20, 22] imply $i_* : \mathsf{Qc}_{\mathcal{U}} \hookrightarrow \mathsf{Qc}_{\mathfrak{X}}$ is an embedding of a thick subcategory; see also [25, section 3.4] for a discussion of the functoriality of $\mathsf{Qc}_{\mathfrak{X}}$ with respect to $\mathfrak{X}$. For $\mathcal{F}, \mathcal{G} \in \mathsf{Qc}_{\mathcal{U}}$ the Yoneda description of Ext-groups gives isomorphisms

$$\mathsf{Ext}^n_{\mathcal{A}}(\mathcal{A} \otimes_{\mathcal{O}_{\mathfrak{X}}} i_*\mathcal{F}, \mathcal{A} \otimes_{\mathcal{O}_{\mathfrak{X}}} i_*\mathcal{G}) \cong \mathsf{Ext}^n_{i_*\mathcal{A}}(i^*\mathcal{A} \otimes_{\mathcal{O}_{\mathcal{U}}} \mathcal{F}, i^*\mathcal{A} \otimes_{\mathcal{O}_{\mathcal{U}}} \mathcal{G}). \tag{9}$$

We shall make use of the following general result in the context in motivic homotopy theory, cf. the proof of Theorem 9.7.

Proposition 2.3: *Suppose there is a 2-commutative diagram of algebraic stacks*

$$\begin{tikzcd}
 & X_0 \arrow[dd, "\pi"] & \\
X \arrow[ur, "\alpha"] \arrow[rr, "f"] \arrow[dr, "f_X"'] \arrow[dd, "\pi_X"'] & & Y \arrow[dl, "f_Y"] \arrow[dd, "\pi_Y"] \\
 & \mathfrak{X} & \\
\mathcal{U} \arrow[ur, hook, "i_X"] \arrow[rr, hook, "i"] & & \mathcal{U}' \arrow[ul, hook', "i_Y"']
\end{tikzcd}$$

where X, Y, X_0 are schemes, π, π_X, π_Y faithfully flat, and i_X, i_Y (hence also i) open inclusions of algebraic substacks. If $\pi_Y^*\pi_{Y,*}\mathcal{O}_Y \in \mathsf{Qc}_Y$ is projective then

$$\mathsf{Ext}^n_{\mathcal{A}(X_0)}(\mathcal{A}(X_0) \otimes_{\mathcal{O}_{X_0}} \pi^* f_{Y,*}\mathcal{O}_Y, \mathcal{A}(X_0) \otimes_{\mathcal{O}_{X_0}} \alpha_*\mathcal{O}_X)$$

$$\cong \begin{cases} 0 & n \geq 1, \\ \mathsf{Hom}_{\mathcal{O}_Y}(\pi_Y^*\pi_{Y,*}\mathcal{O}_Y, \mathcal{A}(Y) \otimes_{\mathcal{O}_Y} f_*\mathcal{O}_X) & n = 0. \end{cases}$$

Proof. By (8) the group $\mathsf{Ext}^n_{\mathcal{A}(X_0)}(\pi^*(\mathcal{A}\otimes_{\mathcal{O}_\mathfrak{X}} f_{Y,*}\mathcal{O}_Y),\mathcal{A}(X_0)\otimes_{\mathcal{O}_{X_0}}\alpha_*\mathcal{O}_X)$ is isomorphic to $\mathsf{Ext}^n_{\mathcal{A}}(\mathcal{A}\otimes_{\mathcal{O}_\mathfrak{X}} f_{Y,*}\mathcal{O}_Y,\pi_*(\pi^*\mathcal{A}\otimes_{\mathcal{O}_{X_0}}\alpha_*\mathcal{O}_X))$, which the projection formula identifies with $\mathsf{Ext}^n_{\mathcal{A}}(\mathcal{A}\otimes_{\mathcal{O}_\mathfrak{X}} i_{Y,*}\pi_{Y,*}\mathcal{O}_Y,\mathcal{A}\otimes_{\mathcal{O}_\mathfrak{X}} i_{Y,*}i_*\pi_{X,*}\mathcal{O}_X)$. By (9) the latter Ext-group is isomorphic to $\mathsf{Ext}^n_{i_Y^*\mathcal{A}}(i_Y^*\mathcal{A}\otimes_{\mathcal{O}_{\mathcal{U}'}}\pi_{Y,*}\mathcal{O}_Y, i_Y^*\mathcal{A}\otimes_{\mathcal{O}_{\mathcal{U}'}} i_*\pi_{X,*}\mathcal{O}_X)$. Replacing $i_*\pi_{X,*}\mathcal{O}_X$ by $\pi_{Y,*}f_*\mathcal{O}_X$ and applying (8) gives an isomorphism to $\mathsf{Ext}^n_{\mathcal{A}(Y)}(\pi_Y^*(i_Y^*\mathcal{A}\otimes_{\mathcal{O}_{\mathcal{U}'}}\pi_{Y,*}\mathcal{O}_Y),\mathcal{A}(Y)\otimes_{\mathcal{O}_Y} f_*\mathcal{O}_X) = \mathsf{Ext}^n_{\mathcal{A}(Y)}(\mathcal{A}(Y)\otimes_{\mathcal{O}_Y}\pi_Y^*\pi_{Y,*}\mathcal{O}_Y,\mathcal{A}(Y)\otimes_{\mathcal{O}_Y} f_*\mathcal{O}_X)$. Now $\mathcal{A}(Y)\otimes_{\mathcal{O}_Y}\pi_Y^*\pi_{Y,*}\mathcal{O}_Y$ is a projective left $\mathcal{A}(Y)$-module by the assumption on $\pi_Y^*\pi_{Y,*}\mathcal{O}_Y$. Hence the Ext-term vanishes in every positive degree, while for $n=0$, we get

$$\begin{aligned}\mathsf{Hom}_{\mathcal{A}(Y)}(\mathcal{A}(Y)\otimes_{\mathcal{O}_Y}\pi_Y^*\pi_{Y,*}\mathcal{O}_Y,\mathcal{A}(Y)\otimes_{\mathcal{O}_Y} f_*\mathcal{O}_X)&\cong\\ \cong\mathsf{Hom}_{\mathcal{O}_Y}(\pi_Y^*\pi_{Y,*}\mathcal{O}_Y,\mathcal{A}(Y)\otimes_{\mathcal{O}_Y} f_*\mathcal{O}_X).&\end{aligned}$$

□

3 Conventions

The category of graded objects in an additive tensor category $\mathcal{A}$ refers to integer-graded objects subject to the Koszul sign rule $x\otimes y=(-1)^{|x||y|}y\otimes x$. However, in the motivic setting, $\mathcal{A}$ will often have a supplementary graded structure. The category of Adams graded objects in $\mathcal{A}$ refers to integer-graded objects in $\mathcal{A}$, but no sign rule for the tensor product is introduced as a consequence of the Adams grading. It is helpful to think of the Adams grading as being even. We will deal with graded abelian groups, Adams graded graded abelian groups, or $\mathbf{Z}^2$-graded abelian groups with a sign rule in the first but not in the second variable, and Adams graded abelian groups. For an Adams graded graded abelian group A_{**} we define $A_i\equiv A_{2i,i}$ and let A_* denote the corresponding Adams graded abelian group. It will be convenient to view evenly graded MU_*-modules as being Adams graded, and implicitly divide the grading by a factor of 2.

The smash product induces a closed symmetric monoidal structure on $\mathbf{SH}(S)$. We denote the internal function spectrum from E to F by $\underline{\mathsf{Hom}}(\mathsf{E},\mathsf{F})$ and the tensor unit or sphere spectrum by $\mathbf{1}$. The Spanier-Whitehead dual of E is by definition $\mathsf{E}^\vee\equiv\underline{\mathsf{Hom}}(\mathsf{E},\mathbf{1})$. Note that E_{**} with the usual indexing is an Adams graded graded abelian group. Let E_i be short for $\mathsf{E}_{2i,i}$. When E is a ring spectrum, i.e. a commutative monoid in $\mathbf{SH}(S)$, we implicitly assume E_{**} is a commutative monoid in Adams graded graded abelian groups. This latter holds true for orientable ring spectra [16, Proposition 2.16] in view of [24, Theorem 3.2.23].

4 Homology and cohomology theories

An object F of $\mathbf{SH}(S)$ is called finite (another term is compact) if $\mathsf{Hom}_{\mathbf{SH}(S)}(\mathsf{F}, -)$ respects sums. Using the 5-lemma one shows the subcategory of finite objects $\mathbf{SH}(S)_{\mathsf{f}}$ of $\mathbf{SH}(S)$ is thick [14, Definition 1.4.3(a)]. For a set $\mathcal{R}$ of objects in $\mathbf{SH}(S)_{\mathsf{f}}$ let $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$ denote the smallest thick triangulated subcategory of $\mathbf{SH}(S)_{\mathsf{f}}$ containing $\mathcal{R}$ and $\mathbf{SH}(S)_{\mathcal{R}}$ the smallest localizing subcategory of $\mathbf{SH}(S)$ containing $\mathcal{R}$ [14, Definition 1.4.3(b)]. The examples we will deal with are the sets of mixed motivic spheres $\mathcal{T}$, the set of (isomorphism classes of) strongly dualizable objects $\mathcal{D}$ and the set $\mathbf{SH}(S)_{\mathsf{f}}$.

Remark 4.1: *According to [7, Remark 7.4]* $\mathbf{SH}(S)_{\mathcal{T}} \subseteq \mathbf{SH}(S)$ *is the full subcategory of cellular motivic spectra introduced in loc. cit.*

Recall $\mathsf{F} \in \mathbf{SH}(S)$ is strongly dualizable if for every $\mathsf{G} \in \mathbf{SH}(S)$ the canonical map

$$\mathsf{F}^{\vee} \wedge \mathsf{G} \longrightarrow \underline{\mathsf{Hom}}(\mathsf{F}, \mathsf{G})$$

is an isomorphism. A strongly dualizable object is finite since $\mathbf{1}$ is finite.

Lemma 4.2: $\mathbf{SH}(S)_{\mathcal{D},\mathsf{f}}$ *is the full subcategory of* $\mathbf{SH}(S)_{\mathsf{f}}$ *of strongly dualizable objects of* $\mathbf{SH}(S)$.

Proof. Since $\mathcal{D}$ is stable under cofiber sequences and retracts, every object of $\mathbf{SH}(S)_{\mathcal{D},\mathsf{f}}$ is strongly dualizable. □

Lemma 4.3: $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$ *is the full subcategory of compact objects of* $\mathbf{SH}(S)_{\mathcal{R}}$ *and the latter is compactly generated.*

Proof. Note $\mathbf{SH}(S)_{\mathcal{R}}$ is compactly generated since $\mathbf{SH}(S)$ is so [28, Theorem 2.1, 2.1.1]. If $(-)^c$ indicates a full subcategory of compact objects [28, Theorem 2.1, 2.1.3] implies

$$\mathbf{SH}(S)_{\mathcal{R}}^c = \mathbf{SH}(S)_{\mathcal{R}} \cap \mathbf{SH}(S)^c = \mathbf{SH}(S)_{\mathcal{R}} \cap \mathbf{SH}(S)_{\mathsf{f}}.$$

Hence it suffices to show $\mathbf{SH}(S)_{\mathcal{R}} \cap \mathbf{SH}(S)_{\mathsf{f}} = \mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$. The inclusion "$\supseteq$" is obvious and to prove "$\subseteq$" let $\mathcal{R}'$ be the smallest set of objects closed under suspension, retract and cofiber sequences containing $\mathcal{R}$. Then $\mathcal{R}' \subseteq \mathbf{SH}(S)_{\mathsf{f}}$ and

$$\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}} = \mathbf{SH}(S)_{\mathcal{R}',\mathsf{f}} \subseteq \mathbf{SH}(S)_{\mathsf{f}}, \mathbf{SH}(S)_{\mathcal{R}} = \mathbf{SH}(S)_{\mathcal{R}'}.$$

By applying [28, Theorem 2.1, 2.1.3] to $\mathcal{R}'$ it follows that

$$\mathbf{SH}(S)_{\mathcal{R}} \cap \mathbf{SH}(S)_{\mathsf{f}} = \mathbf{SH}(S)_{\mathcal{R}'} \cap \mathbf{SH}(S)_{\mathsf{f}} = \mathcal{R}' \subseteq \mathbf{SH}(S)_{\mathcal{R}',\mathsf{f}} = \mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}.$$

□

COROLLARY 4.4: *If $\mathcal{R} \subseteq \mathcal{R}'$ are as above, the inclusion $\mathbf{SH}(S)_{\mathcal{R}} \subseteq \mathbf{SH}(S)_{\mathcal{R}'}$ has a right adjoint $\mathsf{p}_{\mathcal{R},\mathcal{R}'}$.*

Proof. Since $\mathbf{SH}(S)_{\mathcal{R}}$ is compactly generated and the inclusion preserves sums the claim follows from [28, Theorem 4.1]. □

DEFINITION 4.5: *The Tate projection is the functor*

$$\mathsf{p}_{\mathbf{SH}(S)_f,\mathcal{T}} : \mathbf{SH}(S) \longrightarrow \mathbf{SH}(S)_{\mathcal{T}}.$$

LEMMA 4.6: *In the situation of Corollary 4.4, the right adjoint $\mathsf{p}_{\mathcal{R}',\mathcal{R}}$ preserves sums.*

Proof. Using [28, Theorem 5.1] it suffices to show that $\mathbf{SH}(S)_{\mathcal{R}} \subseteq \mathbf{SH}(S)_{\mathcal{R}'}$ preserves compact objects. Hence by Lemma 4.3 we are done provided $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}} \subseteq \mathbf{SH}(S)_{\mathcal{R}',\mathsf{f}}$. Clearly this holds since $\mathcal{R} \subseteq \mathcal{R}'$. □

LEMMA 4.7: *Suppose $\mathcal{R}$ as above contains $\mathcal{T}$. Then*

$$\mathsf{p}_{\mathcal{R},\mathcal{T}} : \mathbf{SH}(S)_{\mathcal{R}} \longrightarrow \mathbf{SH}(S)_{\mathcal{T}}$$

is an $\mathbf{SH}(S)_{\mathcal{T}}$-module functor.

Proof. Let $\iota : \mathbf{SH}(S)_{\mathcal{T}} \to \mathbf{SH}(S)_{\mathcal{R}}$ be the inclusion and $\mathsf{F} \in \mathbf{SH}(S)_{\mathcal{T}}$, $\mathsf{G} \in \mathbf{SH}(S)_{\mathcal{R}}$. Then the counit of the adjunction between ι and $\mathsf{p}_{\mathcal{R},\mathcal{T}}$ yields the canonical map

$$\iota(\mathsf{F} \wedge \mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G})) \cong \iota(\mathsf{F}) \wedge \iota(\mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G})) \longrightarrow \iota(\mathsf{F}) \wedge \mathsf{G},$$

adjoint to

$$\mathsf{F} \wedge \mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G}) \longrightarrow \mathsf{p}_{\mathcal{R},\mathcal{T}}(\iota(\mathsf{F}) \wedge \mathsf{G}). \tag{10}$$

We claim (10) is an isomorphism for all F, G. In effect, the full subcategory of $\mathbf{SH}(S)_{\mathcal{T}}$ generated by the objects F for which (10) is an isomorphism for all $\mathsf{G} \in \mathbf{SH}(S)_{\mathcal{R}}$ is easily seen to be localizing, and hence we may assume $\mathsf{F} = \mathsf{S}^{p,q}$ for $p, q \in \mathbf{Z}$. The sphere $\mathsf{S}^{p,q}$ is invertible, so $\mathbf{SH}(S)_{\mathcal{T}}(-, \mathsf{p}_{\mathcal{R},\mathcal{T}}(\iota(\mathsf{S}^{p,q}) \wedge \mathsf{G})) \cong \mathbf{SH}(S)_{\mathcal{R}}(\iota(-), \mathsf{S}^{p,q} \wedge \mathsf{G})$ is isomorphic to $\mathbf{SH}(S)_{\mathcal{R}}(\iota(-) \wedge \mathsf{S}^{-p,-q}, \mathsf{G}) \cong \mathbf{SH}(S)_{\mathcal{T}}(- \wedge \mathsf{S}^{-p-q}, \mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G})) \cong \mathbf{SH}(S)_{\mathcal{T}}(-, \mathsf{S}^{p,q} \wedge \mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G}))$. This shows $\mathsf{p}_{\mathcal{R},\mathcal{T}}(\iota(\mathsf{S}^{p,q}) \wedge \mathsf{G})$ and $\mathsf{S}^{p,q} \wedge \mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G})$ are isomorphic, as desired. □

REMARK 4.8: *(i) For every $\mathsf{G} \in \mathbf{SH}(S)$ the counit $\mathsf{p}_{\mathcal{R},\mathcal{T}}(\mathsf{G}) \to \mathsf{G}$, where ι is omitted from the notation, is an π_{**}-isomorphism. Using $\mathsf{p}_{\mathbf{SH}(S),\mathcal{T}}$ rather than the cellular functor introduced in [7] refines Proposition 7.3 of loc. cit.*

(ii) *If* $\mathsf{E} \in \mathbf{SH}(S)_{\mathcal{T}}$ *and* $\mathsf{F} \in \mathbf{SH}(S)$ *then* $\mathsf{E}_{p,q}(\mathsf{F}) \cong \mathsf{E}_{p,q}(\mathsf{p}_{\mathbf{SH}(S),\mathcal{T}}(\mathsf{F}))$ *on account of the isomorphisms between* $\mathbf{SH}(S)(\mathsf{S}^{p,q}, \mathsf{E} \wedge \mathsf{F})$ *and*

$$\mathbf{SH}(S)_{\mathcal{T}}(\mathsf{S}^{p,q}, \mathsf{p}_{\mathbf{SH}(S),\mathcal{T}}(\mathsf{E} \wedge \mathsf{F})) \cong \mathbf{SH}(S)_{\mathcal{T}}(\mathsf{S}^{p,q}, \mathsf{E} \wedge \mathsf{p}_{\mathbf{SH}(S),\mathcal{T}}(\mathsf{F})).$$

In [7] it is argued that most spectra should be non-cellular. On the other hand, the E*-homology of* F *agrees with the* E*-homology of some cellular spectrum. We note that many conspicuous motivic (co)homology theories are representable by cellular spectra: Landweber exact theories, including algebraic cobordism and homotopy algebraic* K*-theory, and also motivic (co)homology over fields of characteristic zero according to work of Hopkins and Morel.*

Definition 4.9: *A homology theory on a triangulated subcategory* T *of* $\mathbf{SH}(S)$ *is a homological functor* $\mathsf{T} \to \mathsf{Ab}$ *which preserves sums. Dually, a cohomology theory on* T *is a homological functor* $\mathsf{T}^{\mathrm{op}} \to \mathsf{Ab}$ *which takes sums to products.*

Lemma 4.10: *Suppose* $\mathcal{R} \subseteq \mathcal{D}$ *is closed under duals. Then every homology theory on* $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$ *extends uniquely to a homology theory on* $\mathbf{SH}(S)_{\mathcal{R}}$.

Proof. In view of Lemma 4.3 we can apply [14, Corollary 2.3.11] which we refer to for a more detailed discussion. □

Homology and cohomology theories on $\mathbf{SH}(S)_{\mathcal{D},\mathsf{f}}$ are interchangeable according to the categorical duality equivalence $\mathbf{SH}(S)_{\mathcal{D},\mathsf{f}}^{\mathrm{op}} \cong \mathbf{SH}(S)_{\mathcal{D},\mathsf{f}}$. The same holds for every $\mathcal{R}$ for which $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$ is contained in $\mathbf{SH}(S)_{\mathcal{D},\mathsf{f}}$ and closed under duality, e.g. $\mathbf{SH}(S)_{\mathcal{T},\mathsf{f}}$. We shall address the problem of representing homology theories on $\mathbf{SH}(S)$ in Section 8. Cohomology theories are always defined on $\mathbf{SH}(S)_{\mathsf{f}}$ unless specified to the contrary.

Definition 4.11: *Let* $\mathsf{T} \subset \mathbf{SH}(S)$ *be a triangulated subcategory closed under the smash product. A multiplicative or ring (co)homology theory on* T*, always understood to be commutative, is a (co)homology theory* E *on* T *together with maps* $\mathbf{Z} \to E(\mathsf{S}^{0,0})$ *and* $E(\mathsf{F}) \otimes E(\mathsf{G}) \to E(\mathsf{F} \wedge \mathsf{G})$ *which are natural in* $\mathsf{F}, \mathsf{G} \in \mathsf{T}$*. These maps are subject to the usual unitality, associativity and commutativity constraints [36, pg. 269].*

Ring spectra in $\mathbf{SH}(S)$ give rise to ring homology and cohomology theories. We shall use the following bigraded version of (co)homology theories.

Definition 4.12: *Let* $\mathsf{T} \subset \mathbf{SH}(S)$ *be a triangulated subcategory closed under shifts by all mixed motivic spheres* $\mathsf{S}^{p,q}$*. A bigraded homology theory on* T *is a homological functor* Φ *from* T *to Adams graded graded abelian groups which* *preserves sums together with natural isomorphisms*

$$\Phi(X)_{p,q} \cong \Phi(\Sigma^{1,0}X)_{p+1,q}$$

and

$$\Phi(X)_{p,q} \cong \Phi(\Sigma^{0,1}X)_{p,q+1}$$

such that the diagram

$$\begin{array}{ccc} \Phi(X)_{p,q} & \longrightarrow & \Phi(\Sigma^{1,0}X)_{p+1,q} \\ \downarrow & & \downarrow \\ \Phi(\Sigma^{0,1}X)_{p,q+1} & \longrightarrow & \Phi(\Sigma^{1,1}X)_{p+1,q+1} \end{array}$$

commutes for all p and q.
Bigraded cohomology theories are defined likewise.

We note there is an equivalence of categories between (co)homology theories on T and bigraded (co)homology theories on T.

5 Tate objects and flat Hopf algebroids

Guided by stable homotopy theory, we wish to associate flat Hopf algebroids to suitable motivic ring spectra. By a Hopf algebroid we shall mean a cogroupoid object in the category of commutative rings over either abelian groups, Adams graded abelian groups or Adams graded graded abelian groups. Throughout this section E is a ring spectrum in $\mathbf{SH}(S)_{\mathcal{T}}$. We call E_{**} flat provided one - and hence both - of the canonical maps $\mathsf{E}_{**} \to \mathsf{E}_{**}\mathsf{E}$ is flat, and similarly for E_* and $\mathsf{E}_* \to \mathsf{E}_*\mathsf{E}$.

Lemma 5.1: *(i) If E_{**} is flat then for every motivic spectrum F the canonical map*

$$\mathsf{E}_{**}\mathsf{E} \otimes_{\mathsf{E}_{**}} \mathsf{E}_{**}\mathsf{F} \longrightarrow (\mathsf{E} \wedge \mathsf{E} \wedge \mathsf{F})_{**}$$

is an isomorphism.

(ii) If E_* is flat and the canonical map $\mathsf{E}_*\mathsf{E} \otimes_{\mathsf{E}_*} \mathsf{E}_{**} \to \mathsf{E}_{**}\mathsf{E}$ is an isomorphism, then for every motivic spectrum F the canonical map

$$\mathsf{E}_*\mathsf{E} \otimes_{\mathsf{E}_*} \mathsf{E}_*\mathsf{F} \longrightarrow (\mathsf{E} \wedge \mathsf{E} \wedge \mathsf{F})_*$$

is an isomorphism.

Proof. (i): Using Lemma 4.7 we may assume that F is a Tate object. The proof follows now along the same lines as in topology by first noting that the statement clearly holds when F is a mixed motivic sphere, and secondly that we are comparing homology theories on $\mathbf{SH}(S)_{\mathcal{T}}$ which respect sums. (ii): The two assumptions imply we may refer to (i). Hence there is an isomorphism

$$\mathsf{E}_{**}\mathsf{E} \otimes_{\mathsf{E}_{**}} \mathsf{E}_{**}\mathsf{F} \longrightarrow (\mathsf{E} \wedge \mathsf{E} \wedge \mathsf{F})_{**}.$$

By the second assumption the left hand side identifies with

$$(\mathsf{E}_{*}\mathsf{E} \otimes_{\mathsf{E}_{*}} \mathsf{E}_{**}) \otimes_{\mathsf{E}_{**}} \mathsf{E}_{**}\mathsf{F} \cong \mathsf{E}_{*}\mathsf{E} \otimes_{\mathsf{E}_{*}} \mathsf{E}_{**}\mathsf{F}.$$

Restricting to bidegrees which are multiples of $(2,1)$ yields the claimed isomorphism. $\square$

Corollary 5.2: *(i) If E_{**} is flat then $(\mathsf{E}_{**}, \mathsf{E}_{**}\mathsf{E})$ is canonically a flat Hopf algebroid in Adams graded graded abelian groups and for every $\mathsf{F} \in \mathbf{SH}(S)$ the module $\mathsf{E}_{**}\mathsf{F}$ is an $(\mathsf{E}_{**}, \mathsf{E}_{**}\mathsf{E})$-comodule.*

(ii) If E_{*} is flat and the canonical map $\mathsf{E}_{*}\mathsf{E} \otimes_{\mathsf{E}_{*}} \mathsf{E}_{**} \rightarrow \mathsf{E}_{**}\mathsf{E}$ is an isomorphism, then $(\mathsf{E}_{*}, \mathsf{E}_{*}\mathsf{E})$ is canonically a flat Hopf algebroid in Adams graded abelian groups and for every $\mathsf{F} \in \mathbf{SH}(S)$ the modules $\mathsf{E}_{**}\mathsf{F}$ and $\mathsf{E}_{*}\mathsf{F}$ are $(\mathsf{E}_{*}, \mathsf{E}_{*}\mathsf{E})$-comodules.

The second part of Corollary 5.2 is really a statement about Hopf algebroids:

Lemma 5.3: *Suppose (A_{**}, Γ_{**}) is a flat Hopf algebroid in Adams graded graded abelian groups and the natural map $\Gamma_{*} \otimes_{A_{*}} A_{**} \rightarrow \Gamma_{**}$ is an isomorphism. Then (A_{*}, Γ_{*}) has the natural structure of a flat Hopf algebroid in Adams graded abelian groups, and for every comodule M_{**} over (A_{**}, Γ_{**}) the modules M_{**} and M_{*} are (A_{*}, Γ_{*})-comodules.*

6 The stacks of topological and algebraic cobordism

6.1 The algebraic stack of MU

Denote by $\underline{\mathsf{FG}}$ the moduli stack of one-dimensional commutative formal groups [25]. It is algebraic and a presentation is given by the canonical map $\mathsf{FGL} \rightarrow \underline{\mathsf{FG}}$, where FGL is the moduli scheme of formal group laws. The stack $\underline{\mathsf{FG}}$ has a canonical line bundle ω, and $[\mathsf{MU}_{*}/\mathsf{MU}_{*}\mathsf{MU}]$ is equivalent to the corresponding $\mathbf{G}_{\mathfrak{m}}$-torsor $\underline{\mathsf{FG}}^{s}$ over $\underline{\mathsf{FG}}$.

6.2 The algebraic stack of MGL

In this section we first study the (co)homology of finite Grassmannians over regular noetherian base schemes of finite Krull dimension. Using this computational input we relate the algebraic stacks of MU and MGL. A key result is the isomorphism

$$\mathsf{MGL}_{**}\mathsf{MGL} \cong \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{MU}_*\mathsf{MU}.$$

When S is a field this can easily be extracted from [6, Theorem 5]. Since it is crucial for the following, we will give a rather detailed argument for the generalization.

We recall the notion of oriented motivic ring spectra formulated by Morel [23], cf. [16], [29] and [38]: If E is a motivic ring spectrum, the unit map $\mathbf{1} \to \mathsf{E}$ yields a class $1 \in \mathsf{E}^{0,0}(\mathbf{1})$ and hence by smashing with the projective line a class $\mathsf{c}_1 \in \mathsf{E}^{2,1}(\mathbf{P}^1)$. An orientation on E is a class $\mathsf{c}_\infty \in \mathsf{E}^{2,1}(\mathbf{P}^\infty)$ that restricts to c_1. Note that KGL and MGL are canonically oriented.

For $0 \leq d \leq n$ define the ring

$$\mathsf{R}_{n,d} \equiv \mathbf{Z}[\mathsf{x}_1, \ldots, \mathsf{x}_{n-d}]/(\mathsf{s}_{d+1}, \ldots, \mathsf{s}_n), \tag{11}$$

where s_i is given by

$$1 + \sum_{n=1}^{\infty} \mathsf{s}_n t^n \equiv (1 + \mathsf{x}_1 t + \mathsf{x}_2 t^2 + \ldots + \mathsf{x}_{n-d} t^{n-d})^{-1} \text{ in } \mathbf{Z}[\mathsf{x}_1, \ldots, \mathsf{x}_{n-d}][[t]]^\times.$$

By assigning weight i to x_i every $\mathsf{s}_k \in \mathbf{Z}[\mathsf{x}_1, \ldots, \mathsf{x}_k]$ is homogeneous of degree k. In (11), $\mathsf{s}_j = \mathsf{s}_j(\mathsf{x}_1, \ldots, \mathsf{x}_{n-d}, 0, \ldots)$ by convention when $d+1 \leq i \leq n$.

We note that $\mathsf{R}_{n,d}$ is a free $\mathbf{Z}$-module of rank $\binom{n}{d}$. For every sequence $\underline{a} = (a_1, \ldots, a_d)$ subject to the inequalities $n - d \geq a_1 \geq a_2 \geq \ldots \geq a_d \geq 0$, set:

$$\Delta_{\underline{a}} \equiv \det \begin{pmatrix} \mathsf{x}_{a_1} & \mathsf{x}_{a_1+1} & \cdots & \mathsf{x}_{a_1+d-1} \\ \mathsf{x}_{a_2-1} & \mathsf{x}_{a_2} & \cdots & \mathsf{x}_{a_2+d-2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathsf{x}_{a_d-d+1} & \cdots & \cdots & \mathsf{x}_{a_d} \end{pmatrix}$$

Here $\mathsf{x}_0 \equiv 1$ and $\mathsf{x}_i \equiv 0$ for $i < 0$ or $i > n-d$. The Schur polynomials $\{\Delta_{\underline{a}}\}$ form a basis for $\mathsf{R}_{n,d}$ as a $\mathbf{Z}$-module. Let $\pi : \mathsf{R}_{n+1,d+1} \to \mathsf{R}_{n,d+1}$ be the unique surjective ring homomorphism where $\pi(\mathsf{x}_i) = \mathsf{x}_i$ for $1 \leq i \leq n-d-1$ and $\pi(\mathsf{x}_{n-d}) = 0$. It is easy to see that $\pi(\Delta_{\underline{a}}) = \Delta_{\underline{a}}$ if $a_1 \leq n-d-1$ and $\pi(\Delta_{\underline{a}}) = 0$ for $a_1 = n-d$. Hence the kernel of π is the principal ideal generated by x_{n-d}. That is,

$$\ker(\pi) = \mathsf{x}_{n-d} \cdot \mathsf{R}_{n+1,d+1}.$$

Moreover, let $\iota : \mathsf{R}_{n,d} \to \mathsf{R}_{n+1,d+1}$ be the unique monomorphism of abelian groups such that for every $\underline{a}$, $\iota(\Delta_{\underline{a}}) = \Delta_{\underline{a}'}$ where $\underline{a}' = (n-d, \underline{a}) \equiv (n-d, a_1, \ldots, a_d)$. Clearly we get

$$\operatorname{im}(\iota) = \ker(\pi). \tag{12}$$

Note that ι is a map of degree $n-d$. We will also need the unique ring homomorphism $f : \mathsf{R}_{n+1,d+1} \to \mathsf{R}_{n,d} = \mathsf{R}_{n+1,d+1}/(\mathsf{s}_{d+1})$ where $f(\mathsf{x}_i) = \mathsf{x}_i$ for all $1 \leq i \leq n-d$. Elementary matrix manipulations establish the equalities

$$f(\Delta_{(a_1,\ldots,a_d,0)}) = \Delta_{(a_1,\ldots,a_d)} \tag{13}$$

and

$$\iota(\Delta_{(a_1,\ldots,a_d)}) = \mathsf{x}_{n-d} \cdot \Delta_{(a_1,\ldots,a_d,0)}. \tag{14}$$

Next we discuss some geometric constructions involving Grassmannians.
For $0 \leq d \leq n$, denote by $\mathbf{Gr}_{n-d}(\mathbb{A}^n)$ the scheme parametrizing subvector bundles of rank $n-d$ of the trivial rank n bundle such that the inclusion of the subbundle is locally split. Similarly, $\mathsf{G}(n,d)$ denotes the scheme parametrizing locally free quotients of rank d of the trivial bundle of rank n; $\mathsf{G}(n,d) \cong \mathbf{Gr}_{n-d}(\mathbb{A}^n)$ is smooth of relative dimension $d(n-d)$. If

$$0 \longrightarrow \mathcal{K}_{n,d} \longrightarrow \mathcal{O}^n_{\mathsf{G}(n,d)} \longrightarrow \mathcal{Q}_{n,d} \longrightarrow 0 \tag{15}$$

is the universal short exact sequence of vector bundles on $\mathsf{G}(n,d)$ and $\mathcal{K}'_{n,d}$ denotes the dual of $\mathcal{K}_{n,d}$, then the tangent bundle

$$\mathcal{T}_{\mathsf{G}(n,d)} \cong \mathcal{Q}_{n,d} \otimes \mathcal{K}'_{n,d}. \tag{16}$$

The map

$$i : \mathsf{G}(n,d) \cong \mathbf{Gr}_{n-d}(\mathbb{A}^n) \hookrightarrow \mathbf{Gr}_{n-d}(\mathbb{A}^{n+1}) \cong \mathsf{G}(n+1,d+1)$$

classifying $\mathcal{K}_{n,d} \subseteq \mathcal{O}^n_{\mathsf{G}(n,d)} \hookrightarrow \mathcal{O}^{n+1}_{\mathsf{G}(n,d)}$ is a closed immersion. From (16) it follows that the normal bundle $\mathcal{N}(i)$ of i identifies with $\mathcal{K}_{n,d}$. Next consider the composition on $\mathsf{G}(n+1,d+1)$

$$\alpha : \mathcal{O}^n_{\mathsf{G}(n+1,d+1)} \hookrightarrow \mathcal{O}^{n+1}_{\mathsf{G}(n+1,d+1)} \longrightarrow \mathcal{Q}_{n+1,d+1}$$

for the inclusion into the first n factors. The complement of the support of $\operatorname{coker}(\alpha)$ is an open subscheme $U \subseteq \mathsf{G}(n+1,d+1)$ and there is a map $\pi : U \to \mathsf{G}(n,d+1)$ classifying $\alpha|_U$. It is easy to see that π is an affine bundle of dimension d, and hence

$$\pi \text{ is a motivic weak equivalence.} \tag{17}$$

An argument with geometric points reveals that $U = \mathsf{G}(n+1,d+1) \smallsetminus i(\mathsf{G}(n,d))$. We summarize the above with the diagram

$$\mathsf{G}(n,d) \xhookrightarrow{i} \mathsf{G}(n+1,d+1) \hookleftarrow U \xrightarrow{\pi} \mathsf{G}(n,d+1). \tag{18}$$

With these precursors out of the way we are ready to compute the (co)homology of finite Grassmannians with respect to any oriented motivic ring spectrum. For every $0 \leq d \leq n$ there is a unique morphism of E^{**}-algebras $\varphi_{n,d} : \mathsf{E}^{**} \otimes_{\mathbf{Z}} \mathsf{R}_{n,d} \to \mathsf{E}^{**}(\mathsf{G}(n,d))$ such that $\varphi_{n,d}(\mathsf{x}_i) = \mathsf{c}_i(\mathcal{K}_{n,d})$ for $1 \leq i \leq n-d$. This follows from (15) and the standard calculus of Chern classes in E-cohomology. Note that $\varphi_{n,d}$ is bigraded if we assign degree $(2i,i)$ to $\mathsf{x}_i \in \mathsf{R}_{n,d}$.

Proposition 6.1: *For $0 \leq d \leq n$ the map of E^{**}-algebras*

$$\varphi_{n,d} : \mathsf{E}^{**} \otimes_{\mathbf{Z}} \mathsf{R}_{n,d} \longrightarrow \mathsf{E}^{**}(\mathsf{G}(n,d))$$

is an isomorphism.

Proof. We observe that the result holds when $d = 0$ and $d = n$, since then $\mathsf{G}(n,d) = S$. By induction it suffices to show that if $\varphi_{n,d}$ and $\varphi_{n,d+1}$ are isomorphisms, then so is $\varphi_{n+1,d+1}$. To that end we contemplate the diagram:

$$\begin{tikzcd}
\mathsf{E}^{*-2r,*-r}(\mathsf{G}(n,d)) \arrow[r, "\alpha"] & \mathsf{E}^{**}(\mathsf{G}(n+1,d+1)) \arrow[r, "\beta"] & \mathsf{E}^{**}(\mathsf{G}(n,d+1)) \\
(\mathsf{E}^{**} \otimes_{\mathbf{Z}} \mathsf{R}_{n,d})(-2r,-r) \arrow[u, "\varphi_{n,d}(-2r,-r)", "\cong"'] \arrow[r, "1\otimes\iota"] & \mathsf{E}^{**} \otimes_{\mathbf{Z}} \mathsf{R}_{n+1,d+1} \arrow[u, "\varphi_{n+1,d+1}"] \arrow[r, "1\otimes\pi"] & \mathsf{E}^{**} \otimes_{\mathbf{Z}} \mathsf{R}_{n,d+1} \arrow[u, "\varphi_{n,d+1}", "\cong"']
\end{tikzcd} \tag{19}$$

Here $r \equiv \operatorname{codim}(i) = n-d$ and $(-2r,-r)$ indicates a shift. The top row is part of the long exact sequence in E-cohomology associated with (18) using the Thom isomorphism $\mathsf{E}^{*+2r,*+r}(\mathsf{Th}(\mathcal{N}(i))) \cong \mathsf{E}^{**}(\mathsf{G}(n,d))$ and the fact that $\mathsf{E}^{**}(U) \cong \mathsf{E}^{**}(\mathsf{G}(n,d+1))$ by (17). The lower sequence is short exact by (12). Since $\mathcal{K}_{n+1,d+1}|_U \cong \pi^*(\mathcal{K}_{n,d+1}) \oplus \mathcal{O}_U$ we get $\beta(\varphi_{n+1,d+1}(\mathsf{x}_i)) = \beta(\mathsf{c}_i(\mathcal{K}_{n+1,d+1})) = \mathsf{c}_i(\mathcal{K}_{n+1,d+1}|_U) = \pi^*(\mathsf{c}_i(\mathcal{K}_{n,d+1})) = \varphi_{n,d+1}(1 \otimes \pi(\mathsf{x}_i))$. Therefore, the right hand square in (19) commutes, β is surjective and the top row in (19) is short exact. Next we study the Gysin map α.
Since $i^*(\mathcal{K}_{n+1,d+1}) = \mathcal{K}_{n,d}$ there is a cartesian square of projective bundles:

$$\begin{tikzcd}
\mathbf{P}(\mathcal{K}_{n,d} \oplus \mathcal{O}) \arrow[r, "i'"] \arrow[d, "p"] & \mathbf{P}(\mathcal{K}_{n+1,d+1} \oplus \mathcal{O}) \arrow[d] \\
\mathsf{G}(n,d) \arrow[r, "i"] & \mathsf{G}(n+1,d+1)
\end{tikzcd}$$

By the induction hypothesis $\varphi_{n,d}$ is an isomorphism. Thus the projective bundle theorem gives

$$\mathsf{E}^{**}(\mathbf{P}(\mathcal{K}_{n,d} \oplus \mathcal{O})) \cong (\mathsf{E}^{**} \otimes_{\mathbf{Z}} \mathsf{R}_{n,d})[\mathsf{x}]/(\mathsf{x}^{r+1} + \sum_{i=1}^{r} (-1)^i \varphi_{n,d}(\mathsf{x}_i)\mathsf{x}^{r+1-i}),$$

where $\mathsf{x} \equiv \mathsf{c}_1(\mathcal{O}_{\mathbf{P}(\mathcal{K}_{n,d}\oplus\mathcal{O})}(1)) \in \mathsf{E}^{2,1}(\mathbf{P}(\mathcal{K}_{n,d}\oplus\mathcal{O}))$. Similarly,

$$\mathsf{E}^{**}(\mathbf{P}(\mathcal{K}_{n+1,d+1}\oplus\mathcal{O})) \cong$$
$$\cong \mathsf{E}^{**}(\mathsf{G}(n+1,d+1))[\mathsf{x}']/(\mathsf{x}'^{r+1} + \sum_{i=1}^{r}(-1)^i \varphi_{n+1,d+1}(\mathsf{x}'_i)\mathsf{x}'^{r+1-i}),$$

where $\mathsf{x}' \equiv \mathsf{c}_1(\mathcal{O}_{\mathbf{P}(\mathcal{K}_{n+1,d+1}\oplus\mathcal{O})}(1))$ and $\mathsf{x}'_i = \mathsf{c}_i(\mathcal{K}_{n+1,d+1}) \in \mathsf{R}_{n+1,d+1}$. (We denote the canonical generators of $\mathsf{R}_{n+1,d+1}$ by x'_i in order to distinguish them from $\mathsf{x}_i \in \mathsf{R}_{n,d}$.) Recall the Thom class of $\mathcal{K}_{n,d} \cong \mathcal{N}(i)$ is constructed from

$$\mathsf{th} \equiv \mathsf{c}_r(p^*(\mathcal{K}_{n,d}) \otimes \mathcal{O}_{\mathbf{P}(\mathcal{K}_{n,d}\oplus\mathcal{O})}(1)) =$$
$$= \mathsf{x}^r + \sum_{i=1}^{r}(-1)^i \varphi_{n,d}(\mathsf{x}_i)\mathsf{x}^{r-i} \in \mathsf{E}^{2r,r}(\mathbf{P}(\mathcal{K}_{n,d}\oplus\mathcal{O})).$$

Using $i'^*(\mathsf{x}') = \mathsf{x}$ and $i^*(\varphi_{n+1,d+1}(\mathsf{x}'_i)) = \varphi_{n,d}(\mathsf{x}_i)$ for $1 \le i \le r$, we get that

$$\tilde{\mathsf{th}} \equiv \mathsf{x}'^r + \sum_{i=1}^{r}(-1)^i \varphi_{n+1,d+1}(\mathsf{x}'_i)\mathsf{x}'^{r-i} \in \mathsf{E}^{2r,r}(\mathbf{P}(\mathcal{K}_{n+1,d+1}\oplus\mathcal{O}))$$

satisfies $i^{'*}(\tilde{\mathsf{th}}) = \mathsf{th}$, and if $z : \mathsf{G}(n+1,d+1) \to \mathbf{P}(\mathcal{K}_{n+1,d+1}\oplus\mathcal{O})$ denotes the zero-section, then

$$z^*(\tilde{\mathsf{th}}) = (-1)^{n-d}\varphi_{n+1,d+1}(\mathsf{x}'_{n-d}) \in \mathsf{E}^{2(n-d),n-d}(\mathsf{G}(n+1,d+1)). \tag{20}$$

Moreover, since $i^*(\mathcal{K}_{n+1,d+1}) = \mathcal{K}_{n,d}$ we conclude

$$\mathsf{E}^{**}(i) \circ \varphi_{n+1,d+1} = \varphi_{n,d} \circ (1 \otimes f). \tag{21}$$

By inspection of the construction of the Thom isomorphism, it follows that

$$\alpha \circ \mathsf{E}^{**}(i) \text{ equals multiplication by } z^*(\tilde{\mathsf{th}}). \tag{22}$$

And for every partition $\underline{a}$ as above,

$$\alpha \circ \varphi_{n,d}(\Delta_{\underline{a}}) \overset{(13)}{=} \alpha \circ \varphi_{n,d} \circ (1 \otimes f)(\Delta_{(\underline{a},0)}) \overset{(21)}{=} \alpha \circ \mathsf{E}^{**}(i) \circ \varphi_{n+1,d+1}(\Delta_{(\underline{a},0)})$$
$$\overset{(22)}{=} z^*(\tilde{\mathsf{th}}) \cdot \varphi_{n+1,d+1}(\Delta_{(\underline{a},0)}) \overset{(20)}{=} \varphi_{n+1,d+1}((-1)^{n-d}\mathsf{x}'_{n-d} \cdot \Delta_{(\underline{a},0)})$$
$$\overset{(14)}{=} (-1)^{n-d} \cdot \varphi_{n+1,d+1}((1 \otimes \iota)(\Delta_{\underline{a}})).$$

This verifies that the left hand square in (19) commutes up to a sign. Hence, by the 5-lemma, $\varphi_{n+1,d+1}$ is an isomorphism. $\square$

Since $\Sigma^\infty_+ \mathsf{G}(n,d) \in \mathbf{SH}(S)$ is dualizable and E is oriented we see that for all $0 \leq d \leq n$ the Kronecker product

$$\mathsf{E}^{**}(\mathsf{G}(n,d)) \otimes_{\mathsf{E}_{**}} \mathsf{E}_{**}(\mathsf{G}(n,d)) \longrightarrow \mathsf{E}_{**} \tag{23}$$

is a perfect pairing of finite free E_{**}-modules.

Proposition 6.2: *(i)* $\mathsf{E}^{**}(\mathsf{B}\mathbf{GL}_d) = \mathsf{E}^{**}[[\mathsf{c}_1, \ldots, \mathsf{c}_d]]$ *where* $\mathsf{c}_i \in \mathsf{E}^{2i,i}(\mathsf{B}\mathbf{GL}_d)$ *is the* i*th Chern class of the tautological rank* d *vector bundle.*

(ii) *a)* $\mathsf{E}^{**}(\mathsf{B}\mathbf{GL}) = \mathsf{E}^{**}[[\mathsf{c}_1, \mathsf{c}_2, \ldots]]$ *where* c_i *is the* i*th Chern class of the universal bundle.*

b) $\mathsf{E}_{**}(\mathsf{B}\mathbf{GL}) = \mathsf{E}_{**}[\beta_0, \beta_1, \ldots]/(\beta_0 = 1)$ *as* E_{**}*-algebras where* $\beta_i \in \mathsf{E}_{2i,i}(\mathsf{B}\mathbf{GL})$ *is the image of the dual of* $\mathsf{c}_1^i \in \mathsf{E}^{2i,i}(\mathsf{B}\mathbf{GL}_1)$*.*

(iii) *There are Thom isomorphisms of* E^{**}*-modules*

$$\mathsf{E}^{**}(\mathsf{B}\mathbf{GL}) \xrightarrow{\cong} \mathsf{E}^{**}(\mathsf{MGL})$$

and E_{**}*-algebras*

$$\mathsf{E}_{**}(\mathsf{MGL}) \xrightarrow{\cong} \mathsf{E}_{**}(\mathsf{B}\mathbf{GL}).$$

Proof. Parts (i) and $(ii)a)$ are clear from the above. From (23) we conclude there are canonical isomorphisms

$$\mathsf{E}^{**}(\mathsf{B}\mathbf{GL}_d) \xrightarrow{\cong} \mathsf{Hom}_{\mathsf{E}_{**}}(\mathsf{E}_{**}(\mathsf{B}\mathbf{GL}_d), \mathsf{E}_{**}),$$

$$\mathsf{E}_{**}(\mathsf{B}\mathbf{GL}_d) \xrightarrow{\cong} \mathsf{Hom}_{\mathsf{E}_{**},c}(\mathsf{E}^{**}(\mathsf{B}\mathbf{GL}_d), \mathsf{E}_{**}).$$

The notation $\mathsf{Hom}_{\mathsf{E}_{**},c}$ refers to continuous E_{**}-linear maps with respect to the inverse limit topology on $\mathsf{E}^{**}(\mathsf{B}\mathbf{GL}_d)$ and the discrete topology on E_{**}. Using this, the proofs of parts $(ii)b)$ and (iii) carry over verbatim from topology. ☐

Corollary 6.3: *(i) The tuple* $(\mathsf{MGL}_{**}, \mathsf{MGL}_{**}\mathsf{MGL})$ *is a flat Hopf algebroid in Adams graded graded abelian groups. For every motivic spectrum* F *the module* $\mathsf{MGL}_{**}\mathsf{F}$ *is an* $(\mathsf{MGL}_{**}, \mathsf{MGL}_{**}\mathsf{MGL})$*-comodule.*

(ii) By restriction of structure the tuple $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$ *is a flat Hopf algebroid in Adams graded abelian groups. For every motivic spectrum* F *the modules* $\mathsf{MGL}_{**}\mathsf{F}$ *and* $\mathsf{MGL}_*\mathsf{F}$ *are* $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$*-comodules.*

Proof. (i): We note MGL is a Tate object by [7, Theorem 6.4], Remark 4.1 and MGL_{**} is flat by Proposition 6.2(iii) with $\mathsf{E} = \mathsf{MGL}$. Hence the statement follows from Corollary 5.2(i). (ii): The bidegrees of the generators β_i in Proposition 6.2 are multiples of $(2,1)$. This implies the assumptions in Corollary 5.1(ii) hold, and the statement follows. □

The flat Hopf algebroid $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$ gives rise to the algebraic stack

$$[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}].$$

Although the grading is not required for the definition, it defines a $\mathbf{G}_{\mathfrak{m}}$-action on the stack and we may therefore form the quotient stack $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}}$. For $\mathsf{F} \in \mathbf{SH}(S)$, let $\mathfrak{F}(\mathsf{F})$ be the $\mathbf{G}_{\mathfrak{m}}$-equivariant quasi-coherent sheaf on $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ associated with the comodule structure on $\mathsf{MGL}_*\mathsf{F}$ furnished by Corollary 6.3(ii). Denote by $\mathfrak{F}/\mathbf{G}_{\mathfrak{m}}(\mathsf{F})$ the descended quasi-coherent sheaf on $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}}$.

Lemma 6.4: *(i)*

$$\mathsf{MGL}_{**}\mathsf{MGL} \cong \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{MU}_*\mathsf{MU} \cong \mathsf{MGL}_{**}[b_0, b_1, \ldots]/(b_0 = 1).$$

(ii) Let x, x' *be the images of the orientation on* MGL *with respect to the two natural maps* $\mathsf{MGL}_* \to \mathsf{MGL}_*\mathsf{MGL}$. *Then* $\mathsf{x}' = \sum_{i\geq 0} b_i \mathsf{x}^{i+1}$ *(where* $b_0 = 1$*).*

Proof. Here b_i is the image under the Thom isomorphism of β_i in Proposition 6.2. Part (i) follows by comparing the familiar computation of $\mathsf{MU}_*\mathsf{MU}$ with our computation of $\mathsf{MGL}_{**}\mathsf{MGL}$. For part (ii), the computations leading up to [1, Corollary 6.8] carry over to the algebraic cobordism spectrum. □

6.3 Formal groups and stacks

A graded formal group over an evenly graded ring A_* or more generally over an algebraic $\mathbf{G}_{\mathfrak{m}}$-stack is a group object in formal schemes over the base with a compatible $\mathbf{G}_{\mathfrak{m}}$-action such that locally in the Zariski topology it looks like $\mathrm{Spf}(R_*[\![\mathsf{x}]\!])$, as a formal scheme with $\mathbf{G}_{\mathfrak{m}}$-action, where x has weight -1. (Note that every algebraic $\mathbf{G}_{\mathfrak{m}}$-stack can be covered by affine $\mathbf{G}_{\mathfrak{m}}$-stacks.) This is equivalent to demanding that x has weight 0 (or any other fixed weight) by looking at the base change $R \to R[\mathsf{y}, \mathsf{y}^{-1}]$, y of weight 1. A strict graded formal group is a graded formal group together with a trivialization of the line bundle of invariant vector fields with the trivial line bundle of weight 1. The strict graded formal group associated with the formal group law over MU_* inherits a coaction of $\mathsf{MU}_*\mathsf{MU}$ compatible with the grading and the trivialization; thus,

it descends to a strict graded formal group over $\underline{\mathsf{FG}}^s$. As a stack, $\underline{\mathsf{FG}}^s$ is the moduli stack of formal groups with a trivialization of the line bundle of invariant vector fields, while as a $\mathbf{G}_{\mathfrak{m}}$-stack it is the moduli stack of strict graded formal groups. It follows that $\underline{\mathsf{FG}}$ (with trivial $\mathbf{G}_{\mathfrak{m}}$-action) is the moduli stack of graded formal groups. For a $\mathbf{G}_{\mathfrak{m}}$-stack $\mathfrak{X}$ the space of $\mathbf{G}_{\mathfrak{m}}$-maps to $\underline{\mathsf{FG}}$ is the space of maps from the stack quotient $\mathfrak{X}/\mathbf{G}_{\mathfrak{m}}$ to $\underline{\mathsf{FG}}$. Hence a graded formal group is tantamount to a formal group over $\mathfrak{X}/\mathbf{G}_{\mathfrak{m}}$.
An orientable theory gives rise to a strict graded formal group over the coefficients:

Lemma 6.5: *If $\mathsf{E} \in \mathbf{SH}(S)$ is an oriented ring spectrum satisfying the assumptions in Corollary 5.2(ii) then the corresponding strict graded formal group over E_* inherits a compatible $\mathsf{E}_*\mathsf{E}$-coaction and there is a descended strict graded formal group over the stack $[\mathsf{E}_*/\mathsf{E}_*\mathsf{E}]$. In particular, the flat Hopf algebroid $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$ acquires a well defined strict graded formal group, $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ a strict graded formal group and the quotient stack $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}}$ a formal group.*

Proof. Functoriality of $\mathsf{E}^*(\mathsf{F})$ in E and F ensures the formal group over E_* inherits an $\mathsf{E}_*\mathsf{E}$-coaction. For example, compatibility with the comultiplication of the formal group amounts to commutativity of the diagram:

$$\begin{tikzcd}
(\mathsf{E} \wedge \mathsf{E})^*(\mathbb{P}^\infty) \arrow[r] \arrow[d] & (\mathsf{E} \wedge \mathsf{E} \wedge \mathsf{E})^*(\mathbb{P}^\infty) \arrow[d] \\
(\mathsf{E} \wedge \mathsf{E})^*(\mathbb{P}^\infty \times \mathbb{P}^\infty) \arrow[r] & (\mathsf{E} \wedge \mathsf{E} \wedge \mathsf{E})^*(\mathbb{P}^\infty \times \mathbb{P}^\infty)
\end{tikzcd}$$

All maps respect gradings, so there is a graded formal group over the Hopf algebroid. Different orientations yield formal group laws which differ by a strict isomorphism, so there is an enhanced strict graded formal group over the Hopf algebroid. It induces a strict graded formal group over the $\mathbf{G}_{\mathfrak{m}}$-stack $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ and quotienting out by the $\mathbf{G}_{\mathfrak{m}}$-action yields a formal group over the quotient stack. $\square$

For oriented motivic ring spectra E and F, denote by $\varphi(\mathsf{E},\mathsf{F})$ the strict isomorphism of formal group laws over $(\mathsf{E} \wedge \mathsf{F})_*$ from the pushforward of the formal group law over E_* to the one of the formal group law over F_* given by the orientations on $\mathsf{E} \wedge \mathsf{F}$ induced by E and F.

Lemma 6.6: *Suppose $\mathsf{E}, \mathsf{F}, \mathsf{G}$ are oriented spectra and let $p\colon (\mathsf{E} \wedge \mathsf{F})_* \to (\mathsf{E} \wedge \mathsf{F} \wedge \mathsf{G})_*$, $q\colon (\mathsf{F} \wedge \mathsf{G})_* \to (\mathsf{E} \wedge \mathsf{F} \wedge \mathsf{G})_*$ and $r\colon (\mathsf{E} \wedge \mathsf{G})_* \to (\mathsf{E} \wedge \mathsf{F} \wedge \mathsf{G})_*$ denote the natural maps. Then $r_*\varphi(\mathsf{E},\mathsf{G}) = \mathsf{p}_*\varphi(\mathsf{E},\mathsf{F}) \circ q_*\varphi(\mathsf{F},\mathsf{G})$.*

Corollary 6.7: *If* $\mathsf{E} \in \mathbf{SH}(S)$ *is an oriented ring spectrum and satisfies the assumptions in Corollary 5.2(i), there is a map of Hopf algebroids* $(\mathsf{MU}_*, \mathsf{MU}_*\mathsf{MU}) \to (\mathsf{E}_{**}, \mathsf{E}_{**}\mathsf{E})$ *such that* $\mathsf{MU}_* \to \mathsf{E}_{**}$ *classifies the formal group law on* E_{**} *and* $\mathsf{MU}_*\mathsf{MU} \to \mathsf{E}_{**}\mathsf{E}$ *the strict isomorphism* $\varphi(\mathsf{E}, \mathsf{E})$. *If* E *satisfies the assumptions in Corollary 5.2(ii) then this map factors through a map of Hopf algebroids* $(\mathsf{MU}_*, \mathsf{MU}_*\mathsf{MU}) \to (\mathsf{E}_*, \mathsf{E}_*\mathsf{E})$. *The induced map of stacks classifies the strict graded formal group on* $[\mathsf{E}_*/\mathsf{E}_*\mathsf{E}]$.

6.4 A map of stacks

Corollary 6.7 and the orientation of MGL furnish a map of flat Hopf algebroids

$$(\mathsf{MU}_*, \mathsf{MU}_*\mathsf{MU}) \longrightarrow (\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$$

such that the induced map of $\mathbf{G}_{\mathfrak{m}}$-stacks $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}] \to \underline{\mathsf{FG}}^s$ classifies the strict graded formal group on $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$. Thus there is a 2-commutative diagram:

$$\begin{tikzcd} \mathsf{Spec}(\mathsf{MGL}_*) \arrow[r] \arrow[d] & \mathsf{Spec}(\mathsf{MU}_*) \arrow[d] \\ {[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]} \arrow[r] & \underline{\mathsf{FG}}^s \end{tikzcd} \tag{24}$$

Quotienting out by the $\mathbf{G}_{\mathfrak{m}}$-action yields a map of stacks $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}} \to \underline{\mathsf{FG}}$ which classifies the formal group on $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}}$.

Proposition 6.8: *The diagram (24) is cartesian.*

Proof. Combine Corollary 2.2 and Lemma 6.4. Part (*ii*) of the lemma is needed to ensure that the left and right units of $(\mathsf{MU}_*, \mathsf{MU}_*\mathsf{MU})$ and $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$ are suitably compatible. □

Corollary 6.9: *The diagram*

$$\begin{tikzcd} \mathsf{Spec}(\mathsf{MGL}_*) \arrow[r] \arrow[d] & \mathsf{Spec}(\mathsf{MU}_*) \arrow[d] \\ {[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}}} \arrow[r] & \underline{\mathsf{FG}} \end{tikzcd}$$

is cartesian.

7 Landweber exact theories

Recall the Lazard ring L is isomorphic to MU_*. For a prime p we fix a regular sequence

$$v_0^{(p)} = p, v_1^{(p)}, \ldots \in \mathsf{MU}_*$$

where $v_n^{(p)}$ has degree $2(p^n-1)$ as explained in the introduction. An (ungraded) L-module M is Landweber exact if $(v_0^{(p)}, v_1^{(p)}, \ldots)$ is a regular sequence on M for every p. An Adams graded MU_*-module M_* is Landweber exact if the underlying ungraded module is Landweber exact as an L-module [15, Definition 2.6]. In stacks this translates as follows: An L-module M gives rise to a quasi-coherent sheaf $\mathsf{M}^\sim$ on $\mathsf{Spec}(\mathsf{L})$ and M is Landweber exact if and only if $\mathsf{M}^\sim$ is flat over $\underline{\mathsf{FG}}$ with respect to $\mathsf{Spec}(\mathsf{L}) \to \underline{\mathsf{FG}}$, see [25, Proposition 7].

Lemma 7.1: *Let M_* be an Adams graded MU_*-module and $\mathsf{M}_*^\sim$ the associated quasi-coherent sheaf on $\mathsf{Spec}(\mathsf{MU}_*)$. Then M_* is Landweber exact if and only if $\mathsf{M}_*^\sim$ is flat over $\underline{\mathsf{FG}}^s$ with respect to $\mathsf{Spec}(\mathsf{MU}_*) \to \underline{\mathsf{FG}}^s$.*

Proof. We need to prove the "only if" implication. Assume M_* is Landweber exact so that $\mathsf{M}^\sim$ has a compatible $\mathbf{G}_{\mathfrak{m}}$-action. Let $\mathsf{q}\colon \mathsf{Spec}(\mathsf{MU}_*) \to [\mathsf{Spec}(\mathsf{MU}_*)]/\mathbf{G}_{\mathfrak{m}}$ denote the quotient map and $\mathsf{N}_*^\sim$ the descended quasi-coherent sheaf of $\mathsf{M}_*^\sim$ on $[\mathsf{Spec}(\mathsf{MU}_*)/\mathbf{G}_{\mathfrak{m}}]$. There is a canonical map $\mathsf{N}_*^\sim \to \mathsf{q}_*\mathsf{M}_*^\sim$, which is the inclusion of the weight zero part of the $\mathbf{G}_{\mathfrak{m}}$-action. By assumption, $\mathsf{M}_*^\sim$ is flat over $\underline{\mathsf{FG}}$, i.e. $\mathsf{q}_*\mathsf{M}_*^\sim$ is flat over $\underline{\mathsf{FG}}$. Since $\mathsf{N}_*^\sim$ is a direct summand of $\mathsf{q}_*\mathsf{M}_*^\sim$ it is flat over $\underline{\mathsf{FG}}$. Hence $\mathsf{M}_*^\sim$ is flat over $\underline{\mathsf{FG}}^s$ since there is a cartesian diagram:

$$\begin{tikzcd}
\mathsf{Spec}(\mathsf{MU}_*) \arrow[r] \arrow[d] & \underline{\mathsf{FG}}^s \arrow[d] \\
{[\mathsf{Spec}(\mathsf{MU}_*)]/\mathbf{G}_{\mathfrak{m}}} \arrow[r] & \underline{\mathsf{FG}}
\end{tikzcd}$$

□

Remark 7.2: *Lemma 7.1 does not hold for (ungraded) L-modules: The map $\mathsf{Spec}(\mathbf{Z}) \to \underline{\mathsf{FG}}^s$ classifying the strict formal multiplicative group over the integers is not flat, whereas the corresponding L-module $\mathbf{Z}$ is Landweber exact.*

In the following statements we view Adams graded abelian groups as Adams graded graded abelian groups via the line $\mathbf{Z}(2,1)$. For example an MU_*-module structure on an Adams graded graded abelian group M_{**} is an MU_*-module in this way. In particular, $\mathsf{MGL}_{**}\mathsf{F}$ is an MU_*-module for every motivic spectrum F.

Theorem 7.3: *Suppose* A_* *is a Landweber exact* MU_**-algebra, i.e. there is a map of commutative algebras* $\mathsf{MU}_* \to \mathsf{A}_*$ *in Adams graded abelian groups such that* A_* *viewed as an* MU_**-module is Landweber exact. Then the functor* $\mathsf{MGL}_{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{A}_*$ *is a bigraded ring homology theory on* $\mathbf{SH}(S)$.

Proof. By Corollary 6.8 there is a projection p from

$$\mathsf{Spec}(\mathsf{A}_*) \times_{\underline{\mathsf{FG}}^s} [\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}] \cong \mathsf{Spec}(\mathsf{A}_*) \times_{\mathsf{Spec}(\mathsf{MU}_*)} \mathsf{Spec}(\mathsf{MGL}_*)$$

to $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ such that

$$\mathsf{MGL}_*\mathsf{F} \otimes_{\mathsf{MU}_*} \mathsf{A}_* \cong \Gamma(\mathsf{Spec}(\mathsf{A}_*) \times_{\underline{\mathsf{FG}}^s} [\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}], \mathsf{p}^*\mathfrak{F}(\mathsf{F})). \tag{25}$$

(This is an isomorphism of Adams graded abelian groups, but we won't use that fact.) The assignment $\mathsf{F} \mapsto \mathfrak{F}(\mathsf{F})$ is a homological functor since $\mathsf{F} \mapsto \mathsf{MGL}_*\mathsf{F}$ is a homological functor, and p is flat since it is the pullback of $\mathsf{Spec}(\mathsf{A}_*) \to \underline{\mathsf{FG}}^s$ which is flat by Lemma 7.1. Thus p^* is exact. Taking global sections over an affine scheme is an exact functor [37, Corollary 4.23]. Therefore, $\mathsf{F} \mapsto \Gamma(\mathsf{Spec}(\mathsf{A}_*) \times_{\underline{\mathsf{FG}}^s} [\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}], \mathsf{p}^*\mathfrak{F}(\mathsf{F}))$ is a homological functor on $\mathbf{SH}(S)$, so that by (25) $\mathsf{F} \mapsto \mathsf{MGL}_*\mathsf{F} \otimes_{\mathsf{MU}_*} \mathsf{A}_*$ is a homological functor with values in Adams graded abelian groups. It follows that $\mathsf{F} \mapsto (\mathsf{MGL}_*\mathsf{F} \otimes_{\mathsf{MU}_*} \mathsf{A}_*)_0$, the degree zero part in the Adams graded abelian group, is a homological functor, and it preserves sums. Hence it is a homology theory on $\mathbf{SH}(S)$. The associated bigraded homology theory is clearly the one formulated in the theorem. Finally, the ring structure is induced by the ring structures on the homology theory represented by MGL and on A_*. ☐

We note the proof works using $\mathfrak{F}/\mathbf{G}_{\mathfrak{m}}(\mathsf{F})$ instead of $\mathfrak{F}(\mathsf{F})$; this makes the reference to Lemma 7.1 superfluous since neglecting the grading does not affect the proof.

Corollary 7.4: *The functor* $\mathsf{MGL}^{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{A}_*$ *is a ring cohomology theory on strongly dualizable motivic spectra.*

Proof. Applying the functor in Theorem 7.3 to the Spanier-Whitehead duals of strongly dualizable motivic spectra yields the cohomology theory on display. Its ring structure is induced by the ring structure on A_*. ☐

Proposition 7.5: *The maps* $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}] \to \underline{\mathsf{FG}}^s$ *and* $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]/\mathbf{G}_{\mathfrak{m}} \to \underline{\mathsf{FG}}$ *are affine.*

Proof. Use Proposition 6.8, Corollary 6.9 and the fact that being an affine morphism can be tested after faithfully flat base change. ☐

Remark 7.6: *We may formulate the above reasoning in more sheaf theoretic terms: Namely, denoting by* $i\colon [\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}] \to \underline{\mathsf{FG}}^s$ *the canonical map, the Landweber exact theory is given by taking sections of* $i_*\mathfrak{F}(\mathsf{F})$ *over* $\mathsf{Spec}(\mathsf{A}_*) \to \underline{\mathsf{FG}}^s$. *It is a homology theory by Proposition 7.5 since* $\mathsf{Spec}(\mathsf{A}_*) \to \underline{\mathsf{FG}}^s$ *is flat.*

Next we give the versions of the above theorems for MU_*-modules.

Proposition 7.7: *Suppose* M_* *is an Adams graded Landweber exact* MU_*-*module. Then* $\mathsf{MGL}_{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{M}_*$ *is a homology theory on* $\mathbf{SH}(S)$ *and* $\mathsf{MGL}^{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{M}_*$ *a cohomology theory on strongly dualizable spectra.*

Proof. The map $i\colon [\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}] \to \underline{\mathsf{FG}}^s$ is affine according to Proposition 7.5. With $p\colon \mathsf{Spec}(\mathsf{MU}_*) \to \underline{\mathsf{FG}}^s$ the canonical map, the first functor in the proposition is given by

$$\mathsf{F} \longmapsto \Gamma(\mathsf{Spec}(\mathsf{MU}_*), \mathsf{M}_* \otimes_{\mathsf{MU}_*} p^* i_* \mathfrak{F}(\mathsf{F})),$$

which is exact by assumption.
The second statement is proven by taking Spanier-Whitehead duals. □

A Landweber exact theory refers to a homology or cohomology theory constructed as in Proposition 7.7. There are periodic versions of the previous results:

Proposition 7.8: *Suppose* M *is a Landweber exact* L-*module. Then* $\mathsf{MGL}_*(-) \otimes_{\mathsf{L}} \mathsf{M}$ *is a* $(2,1)$-*periodic homology theory on* $\mathbf{SH}(S)$ *with values in ungraded abelian groups. The same statement holds for cohomology of strongly dualizable objects. These are ring theories if* M *is a commutative* L-*algebra.*

Next we formulate the corresponding results for (highly structured) MGL-modules. In stable homotopy theory this viewpoint is emphasized in [20] and it plays an important role in this paper, cf. Section 9.

Proposition 7.9: *Suppose* M_* *is a Landweber exact Adams graded* MU_*-*module. Then* $\mathsf{F} \mapsto \mathsf{F}_{**} \otimes_{\mathsf{MU}_*} \mathsf{M}_*$ *is a bigraded homology theory on the derived category* $\mathcal{D}_{\mathsf{MGL}}$ *of* MGL-*modules.*

Proof. The proof proceeds along a now familiar route. What follows reviews the main steps. We wish to construct a homological functor from $\mathcal{D}_{\mathsf{MGL}}$ to quasi-coherent sheaves on $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$. Our first claim is that for every $\mathsf{F} \in \mathcal{D}_{\mathsf{MGL}}$ the Adams graded MGL_*-module F_* is an $(\mathsf{MGL}_*, \mathsf{MGL}_*\mathsf{MGL})$-comodule. As in Lemma 5.1,

$$\mathsf{MGL}_{**}\mathsf{MGL} \otimes_{\mathsf{MGL}_{**}} \mathsf{F}_{**} \longrightarrow (\mathsf{MGL} \wedge \mathsf{F})_{**}$$

is an isomorphism restricting to an isomorphism

$$\mathsf{MGL}_*\mathsf{MGL} \otimes_{\mathsf{MGL}_*} \mathsf{F}_* \longrightarrow (\mathsf{MGL} \wedge \mathsf{F})_*.$$

This is proven by first observing that it holds for "spheres" $\Sigma^{p,q}\mathsf{MGL}$, and secondly that both sides are homological functors which commute with sums. This establishes the required comodule structure. Next, the proof of Proposition 7.7 using flatness of M_* viewed as a quasi-coherent sheaf on $[\mathsf{MGL}_*/\mathsf{MGL}_*\mathsf{MGL}]$ shows the functor in question is a homology theory. The remaining parts are clear. □

Remark 7.10: *We leave the straightforward formulations of the cohomology, algebra and periodic versions of Proposition 7.9 to the reader.*

8 Representability and base change

Here we deal with the question when a motivic (co)homology theory is representable. Let $\mathcal{R}$ be a subset of $\mathbf{SH}(S)_{\mathsf{f}}$ such that $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$ consists of strongly dualizable objects, is closed under smash products and duals and contains the unit.

First, recall the notions of unital algebraic stable homotopy categories and Brown categories from [14, Definition 1.1.4 and next paragraph]: A stable homotopy category is a triangulated category equipped with sums, a compatible closed tensor product, a set $\mathcal{G}$ of strongly dualizable objects generating the triangulated category as a localizing subcategory, and such that every cohomological functor is representable. It is unital algebraic if the tensor unit is finite (thus the objects of $\mathcal{G}$ are finite) and a Brown category if homology functors and natural transformations between them are representable.

A map between objects in a stable homotopy category is phantom if the induced map between the corresponding cohomology functors on the full subcategory of finite objects is the zero map. In case the category is unital algebraic this holds if and only if the map between the induced homology theories is the zero map.

Lemma 8.1: *The category $\mathbf{SH}(S)_{\mathcal{R}}$ is a unital algebraic stable homotopy category. The set $\mathcal{G}$ can be chosen to be (representatives of) the objects of $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$.*

Proof. This is an immediate application of [14, Theorem 9.1.1]. □

Remark 8.2: *If $S = \mathsf{Spec}(k)$ for a field k admitting resolutions of singularities, then $\mathbf{SH}(S)$ itself is unital algebraic, essentially because every smooth k-scheme is strongly dualizable in $\mathbf{SH}(S)$, cf. [31, Theorem 52]. For S the spectrum of* *a discrete valuation ring R with quotient field K, $U_+ := \mathsf{Spec}(K)_+ \in \mathbf{SH}(S)$ is compact but not strongly dualizable, hence by [14, Theorem 2.1.3,d)] $\mathbf{SH}(S)$ is not unital algebraic. We sketch a proof of the fact that U_+ is not dualizable which arose in discussion with J. Riou: Assume U_+ was dualizable, and consider the trace of its identity, an element of $\pi_{0,0}(\mathbf{1}_R)$ which restricts to $1 \in \pi_{0,0}(\mathbf{1}_K)$ and to $0 \in \pi_{0,0}(\mathbf{1}_\kappa)$ (κ the residue field of R). To obtain a contradiction, it would thus suffice to know that $\pi_{0,0}(\mathbf{1}_R)$ is simple, which seems plausible but is open to the authors' knowledge. However, it suffices to construct a tensor-functor $\mathbf{SH}(S) \to D$ (a "realization") such that the corresponding statements hold in D. Taking for D the category of $\mathsf{L}\mathbf{Q}$-modules (cf. Section 10) is easily seen to work.*

Lemma 8.3: *Suppose S is covered by Zariski spectra of countable rings. Then $\mathbf{SH}(S)_\mathcal{R}$ is a Brown category and the category of homology functors on $\mathbf{SH}(S)_\mathcal{R}$ is naturally equivalent to $\mathbf{SH}(S)_\mathcal{R}$ modulo phantom maps.*

Proof. The first part follows by combining [14, Theorem 4.1.5] and [39, Proposition 5.5], [26, Theorem 1] and the second part by the definition of a Brown category. ☐

Suppose $\mathcal{R}, \mathcal{R}'$ are as above and $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}} \subset \mathbf{SH}(S)_{\mathcal{R}',\mathsf{f}}$. Then a cohomology theory on $\mathbf{SH}(S)_{\mathcal{R}',\mathsf{f}}$ represented by F restricts to a cohomology theory on $\mathbf{SH}(S)_{\mathcal{R},\mathsf{f}}$ represented by $\mathsf{p}_{\mathcal{R}',\mathcal{R}}(\mathsf{F})$. For Landweber exact theories the following holds:

Proposition 8.4: *Suppose a Landweber exact homology theory restricted to $\mathbf{SH}(S)_{\mathcal{T},\mathsf{f}}$ is represented by a Tate spectrum E. Then E represents the theory on $\mathbf{SH}(S)$.*

Proof. Let M_* be a Landweber exact Adams graded MU_*-module affording the homology theory under consideration. By assumption there is an isomorphism on $\mathbf{SH}(S)_{\mathcal{T},\mathsf{f}}$

$$\mathsf{E}_{**}(-) \cong \mathsf{MGL}_{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{M}_*.$$

By Lemma 4.10 the isomorphism extends to $\mathbf{SH}(S)_\mathcal{T}$. Since MGL is cellular, an argument as in Remark 4.8 shows that both sides of the isomorphism remain unchanged when replacing a motivic spectrum by its Tate projection. ☐

Next we consider a map $f\colon S' \to S$ of base schemes. The derived functor $\mathbf{L}f^*$, see [30, Proposition A.7.4], sends the class of compact generators $\Sigma^{p,q}\Sigma^\infty X_+$ of $\mathbf{SH}(S)$ - X a smooth S-scheme - to compact objects of $\mathbf{SH}(S')$. Hence [28, Theorem 5.1] implies $\mathbf{R}f_*$ preserves sums, and the same result shows $\mathbf{L}f^*$

preserves compact objects in general. A modification of the proof of Lemma 4.7 shows $\mathbf{R}f_*$ is an $\mathbf{SH}(S)_{\mathcal{T}}$-module functor, i.e. there is an isomorphism

$$\mathbf{R}f_*(\mathsf{F}' \wedge \mathbf{L}f^*\mathsf{G}) \cong \mathbf{R}f_*(\mathsf{F}') \wedge \mathsf{G} \tag{26}$$

in $\mathbf{SH}(S)$, which is natural in $\mathsf{F}' \in \mathbf{SH}(S')$, $\mathsf{G} \in \mathbf{SH}(S)_{\mathcal{T}}$.

Proposition 8.5: *Suppose a Landweber exact homology theory over S determined by the Adams graded MU_*-module M_* is representable by $\mathsf{E} \in \mathbf{SH}(S)_{\mathcal{T}}$. Then $\mathbf{L}f^*\mathsf{E} \in \mathbf{SH}(S')_{\mathcal{T}}$ represents the Landweber exact homology theory over S' determined by M_*.*

Proof. For an object F' of $\mathbf{SH}(S')$, adjointness, the assumption on E and (26) imply $(\mathbf{L}f^*\mathsf{E})_{**}(\mathsf{F}') = \pi_{**}(\mathsf{F}' \wedge \mathbf{L}f^*\mathsf{E})$ is isomorphic to

$$\pi_{**}(\mathbf{R}f_*(\mathsf{F}' \wedge \mathbf{L}f^*\mathsf{E})) \cong \pi_{**}(\mathbf{R}f_*\mathsf{F}' \wedge \mathsf{E}) \cong \pi_{**}(\mathsf{MGL} \wedge \mathbf{R}f_*\mathsf{F}') \otimes_{\mathsf{MU}_*} \mathsf{M}_*.$$

Again by adjointness and (26) there is an isomorphism with

$$\pi_{**}(\mathsf{MGL}_{S'} \wedge \mathsf{F}') \otimes_{\mathsf{MU}_*} \mathsf{M}_* = \mathsf{MGL}_{S',**}\mathsf{F}' \otimes_{\mathsf{MU}_*} \mathsf{M}_*.$$

□

In the next lemma we show the pullback from Proposition 8.5 respects multiplicative structures. In general one cannot expect that ring structures on the homology theory lift to commutative monoid structures on representing spectra. Instead we will consider quasi-multiplications on spectra, by which we mean maps $\mathsf{E} \wedge \mathsf{E} \to \mathsf{E}$ rendering the relevant diagrams commutative up to phantom maps.

Lemma 8.6: *Suppose a Landweber exact homology theory afforded by the Adams graded MU_*-algebra A_* is represented by a Tate object $\mathsf{E} \in \mathbf{SH}(S)_{\mathcal{T}}$ with quasi-multiplication $m\colon \mathsf{E} \wedge \mathsf{E} \to \mathsf{E}$. Then $\mathbf{L}f^*m\colon \mathbf{L}f^*\mathsf{E} \wedge \mathbf{L}f^*\mathsf{E} \to \mathbf{L}f^*\mathsf{E}$ is a quasi-multiplication and represents the ring structure on the Landweber exact homology theory determined by A_* over S'.*

Proof. Let $\phi\colon \mathsf{F}_1 \wedge \mathsf{F}_2 \to \mathsf{F}_3$ be a map in $\mathbf{SH}(S)_{\mathcal{T}}$. Let F'_i be the base change of F_i to S'. If $\mathsf{F}', \mathsf{G}' \in \mathbf{SH}(S')$ there are isomorphisms $\mathsf{F}'_{i,**}\mathsf{F}' \cong \mathsf{F}_{i,**}\mathbf{R}f_*\mathsf{F}'$ employed in the proof of Proposition 8.5, and likewise for G'. These isomorphisms are compatible with ϕ in the sense provided by the commutative diagram:

$$\begin{tikzcd}
\mathsf{F}'_{1,**}\mathsf{F}' \otimes \mathsf{F}'_{2,**}\mathsf{G}' \arrow[rr] & & \mathsf{F}'_{3,**}(\mathsf{F}' \wedge \mathsf{G}') \\
 & & \mathsf{F}_{3,**}(\mathbf{R}f_*(\mathsf{F}' \wedge \mathsf{G}')) \arrow[u, "\cong"'] \\
\mathsf{F}_{1,**}\mathbf{R}f_*\mathsf{F}' \otimes \mathsf{F}_{2,**}\mathbf{R}f_*\mathsf{G}' \arrow[rr] \arrow[uu, "\cong"'] & & \mathsf{F}_{3,**}(\mathbf{R}f_*\mathsf{F}' \wedge \mathbf{R}f_*\mathsf{G}') \arrow[u]
\end{tikzcd}$$

Applying the above to the quasi-multiplication m implies $\mathbf{L}f^*m$ represents the ring structure on the Landweber theory over S'. Hence $\mathbf{L}f^*m$ is a quasi-multiplication since the commutative diagrams exist for the homology theories, i.e. up to phantom maps. □

We are ready to prove the motivic analog of Landweber's exact functor theorem.

Theorem 8.7: *Suppose M_* is an Adams graded Landweber exact MU_*-module. Then there exists a Tate object $\mathsf{E} \in \mathbf{SH}(S)_{\mathcal{T}}$ and an isomorphism of homology theories on $\mathbf{SH}(S)$*

$$\mathsf{E}_{**}(-) \cong \mathsf{MGL}_{**}(-) \otimes_{\mathsf{MU}_*} \mathsf{M}_*.$$

In addition, if M_* is a graded MU_*-algebra, then E acquires a quasi-multiplication which represents the ring structure on the Landweber exact theory.

Proof. First, let $S = \mathsf{Spec}(\mathbf{Z})$. By Landweber exactness, see Proposition 7.7, the right hand side of the claimed isomorphism is a homology theory on $\mathbf{SH}(\mathbf{Z})$. Its restriction to $\mathbf{SH}(\mathbf{Z})_{\mathcal{T},\mathsf{f}}$ is represented by some $\mathsf{E} \in \mathbf{SH}(\mathbf{Z})_{\mathcal{T}}$ since $\mathbf{SH}(\mathbf{Z})_{\mathcal{T}}$ is a Brown category by Lemma 8.3. We may conclude in this case using Proposition 8.4. The general case follows from Proposition 8.5 since $\mathbf{L}f^*(\mathbf{SH}(\mathbf{Z})_{\mathcal{T}}) \subseteq \mathbf{SH}(S)_{\mathcal{T}}$ for $f : S \to \mathsf{Spec}(\mathbf{Z})$.
Now assume M_* is a graded MU_*-algebra. We claim that the representing spectrum $\mathsf{E} \in \mathbf{SH}(\mathbf{Z})_{\mathcal{T}}$ has a quasi-multiplication representing the ring structure on the Landweber theory: The corresponding ring cohomology theory on $\mathbf{SH}(\mathbf{Z})_{\mathcal{T},\mathsf{f}}$ can be extended to ind-representable presheaves on $\mathbf{SH}(\mathbf{Z})_{\mathcal{T},\mathsf{f}}$. Evaluating $\mathsf{E}(\mathsf{F}) \otimes \mathsf{E}(\mathsf{G}) \to \mathsf{E}(\mathsf{F} \wedge \mathsf{G})$ with $\mathsf{F} = \mathsf{G}$ the ind-representable presheaf given by E on $\mathsf{id}_\mathsf{E} \otimes \mathsf{id}_\mathsf{E}$ gives a map $(\mathsf{E} \wedge \mathsf{E})_0(-) \to \mathsf{E}_0(-)$ of homology theories. Since $\mathbf{SH}(\mathbf{Z})_{\mathcal{T}}$ is a Brown category this map lifts to a map $\mathsf{E} \wedge \mathsf{E} \to \mathsf{E}$ of spectra which is a quasi-multiplication since it represents the multiplication of the underlying homology theory. The general case follows from Lemma 8.6. □

Remark 8.8: *A complex point* $\mathsf{Spec}(\mathbf{C}) \to S$ *induces a sum preserving* $\mathbf{SH}(S)_{\mathcal{T}}$*-module realization functor* $r\colon \mathbf{SH}(S) \to \mathbf{SH}$ *to the stable homotopy category. By the proof of Proposition 8.5 it follows that the topological realization of a Landweber exact theory is the corresponding topological Landweber exact theory, as one would expect.*

Proposition 8.9: *Suppose* M_* *is an Adams graded Landweber exact* MU_**-module. Then there exists an* MGL*-module* E *and an isomorphism of homology theories on* $\mathcal{D}_{\mathsf{MGL}}$

$$(\mathsf{E} \wedge_{\mathsf{MGL}} -)_{**} \cong (-)_{**} \otimes_{\mathsf{MU}_*} \mathsf{M}_*.$$

In addition, if M_* *is a graded* MU_**-algebra then* E *acquires a quasi-multiplication in* $\mathcal{D}_{\mathsf{MGL}}$ *which represents the ring structure on the Landweber exact theory.*

Proof. We indicate a proof. By Proposition 7.9 it suffices to show that the homology theory given by the right hand side of the isomorphism is representable. When the base scheme is $\mathsf{Spec}(\mathbf{Z})$ we claim that $\mathcal{D}_{\mathsf{MGL},\mathcal{T}}$ is a Brown category. In effect, $\mathbf{SH}(S)_{\mathsf{f}}$ is countable, cf. [39, Proposition 5.5], [26, Theorem 1], and MGL is a countable direct homotopy limit of finite spectra, so it follows that $\mathcal{D}_{\mathsf{MGL},\mathcal{T},\mathsf{f}}$ is also countable. The conclusion that $\mathcal{D}_{\mathsf{MGL},\mathcal{T}}$ be a Brown category follows now from [14, Theorem 4.1.5]. Thus there exists an object of $\mathcal{D}_{\mathsf{MGL},\mathcal{T}}$ representing the Landweber exact theory over $\mathsf{Spec}(\mathbf{Z})$. Now let $f\colon S \to \mathsf{Spec}(\mathbf{Z})$ be the unique map and $\mathbf{L}f^*_{\mathsf{MGL}}\colon \mathcal{D}_{\mathsf{MGL}_\mathbf{Z}} \to \mathcal{D}_{\mathsf{MGL}_S}$ the pullback functor between MGL-modules. It has a right adjoint $\mathbf{R}f_{\mathsf{MGL},*}$. As prior to Proposition 8.5, we conclude $\mathbf{R}f_{\mathsf{MGL},*}$ preserves sums and is a $\mathcal{D}_{\mathsf{MGL}_\mathbf{Z},\mathcal{T}}$-module functor. The proof of Proposition 8.5 shows $\mathbf{L}f^*_{\mathsf{MGL}}$ represents the Landweber theory over S.
By inferring the analog of Lemma 8.6 our claim about the quasi-multiplication is proven along the lines of the corresponding statement in Theorem 8.7. ☐

9 Operations and cooperations

Let A_* be a Landweber exact Adams graded MU_*-algebra and E a motivic spectrum with a quasi-multiplication which represents the corresponding Landweber exact theory. Denote by $\mathsf{E}^{\mathsf{Top}}$ the ring spectrum representing the corresponding topological Landweber exact theory. Then $\mathsf{E}^{\mathsf{Top}}_* \cong \mathsf{A}_*$, $\mathsf{E}^{\mathsf{Top}}$ is a commutative monoid in the stable homotopy category and there are no even degree nontrivial phantom maps between such topological spectra [15, Section 2.1].

Proposition 9.1: *In the above situation the following hold.*

(i) $\mathsf{E}_{**}\mathsf{E} \cong \mathsf{E}_{**} \otimes_{\mathsf{E}^{\mathsf{Top}}_*} \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}$.

(ii) E *satisfies the assumption of Corollary 5.2(ii).*

(iii) The flat Hopf algebroid $(\mathsf{E}_{**}, \mathsf{E}_{**}\mathsf{E})$ *is induced from* $(\mathsf{MGL}_{**}, \mathsf{MGL}_{**}\mathsf{MGL})$ *via the map* $\mathsf{MGL}_{**} \to \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{A}_* \cong \mathsf{E}_{**}$.

Proof. The isomorphism $\mathsf{E}_{**}\mathsf{F} \cong \mathsf{MGL}_{**}\mathsf{F} \otimes_{\mathsf{MU}_*} \mathsf{A}_*$ can be recast as

$$\mathsf{E}_{**}\mathsf{F} \cong \mathsf{MGL}_{**}\mathsf{F} \otimes_{\mathsf{MGL}_*} \mathsf{MGL}_* \otimes_{\mathsf{MU}_*} \mathsf{E}^{\mathsf{Top}}_* \cong \mathsf{MGL}_{**}\mathsf{F} \otimes_{\mathsf{MGL}_*} \mathsf{E}_*$$

and

$$\mathsf{E}_{**}\mathsf{F} \cong \mathsf{MGL}_{**}\mathsf{F} \otimes_{\mathsf{MGL}_{**}} \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{E}^{\mathsf{Top}}_* \cong \mathsf{MGL}_{**}\mathsf{F} \otimes_{\mathsf{MGL}_{**}} \mathsf{E}_{**}.$$

In particular, $\mathsf{E}_{**}\mathsf{E} \cong \mathsf{MGL}_{**}\mathsf{E} \otimes_{\mathsf{MGL}_{**}} \mathsf{E}_{**} \cong \mathsf{E}_{**}\mathsf{MGL} \otimes_{\mathsf{MGL}_{**}} \mathsf{E}_{**}$ is isomorphic to

$$(\mathsf{MGL}_{**}\mathsf{MGL} \otimes_{\mathsf{MGL}_{**}} \mathsf{E}_{**}) \otimes_{\mathsf{MGL}_{**}} \mathsf{E}_{**} \cong \mathsf{E}_{**} \otimes_{\mathsf{MGL}_{**}} \mathsf{MGL}_{**}\mathsf{MGL} \otimes_{\mathsf{MGL}_{**}} \mathsf{E}_{**}. \quad (27)$$

Moreover, since $\mathsf{MGL}_{**}\mathsf{MGL} \cong \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{MU}_*\mathsf{MU}$,

$$\mathsf{E}^{\mathsf{Top}}_* \otimes_{\mathsf{MU}_*} \mathsf{MGL}_{**}\mathsf{MGL} \otimes_{\mathsf{MU}_*} \mathsf{E}^{\mathsf{Top}}_* \cong \mathsf{E}^{\mathsf{Top}}_* \otimes_{\mathsf{MU}_*} \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{MU}_*\mathsf{MU} \otimes_{\mathsf{MU}_*} \mathsf{E}^{\mathsf{Top}}_*$$

is isomorphic to

$$\mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}} \cong \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{E}^{\mathsf{Top}}_* \otimes_{\mathsf{E}^{\mathsf{Top}}_*} \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}} \cong \mathsf{E}_{**} \otimes_{\mathsf{E}^{\mathsf{Top}}_*} \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}.$$

This proves the first part of the proposition. In particular,

$$\mathsf{E}_*\mathsf{E} \cong \mathsf{E}_* \otimes_{\mathsf{E}^{\mathsf{Top}}_*} \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}} \quad (28)$$

and

$$\mathsf{E}_{**}\mathsf{E} \cong \mathsf{E}_{**} \otimes_{\mathsf{E}_*} \mathsf{E}_*\mathsf{E}. \quad (29)$$

We note that $\mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}$ is flat over $\mathsf{E}^{\mathsf{Top}}_*$ by the topological analog of (27) (this equation shows $\mathsf{Spec}(\mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}) = \mathsf{Spec}(\mathsf{E}^{\mathsf{Top}}_*) \times_{\underline{\mathsf{FG}}^s} \mathsf{Spec}(\mathsf{E}^{\mathsf{Top}}_*)$). Hence by (28) $\mathsf{E}_*\mathsf{E}$ is flat over E_*. Together with (29) this is Part (*ii*) of the proposition. Part (*iii*) follows from (27). ☐

Remark 9.2: *Let* $\mathsf{E}^{\mathsf{Top}}$ *and* $\mathsf{F}^{\mathsf{Top}}$ *be evenly graded topological Landweber exact spectra,* E *and* F *the corresponding motivic spectra. Then* $\mathsf{E} \wedge \mathsf{F}$ *is Landweber exact corresponding to the* MU_**-module* $(\mathsf{E}^{\mathsf{Top}} \wedge \mathsf{F}^{\mathsf{Top}})_*$ *(with either* MU_**-module structure).*

Theorem 9.3: *(i) The map afforded by the Kronecker product*

$$\mathsf{KGL}^{**}\mathsf{KGL} \longrightarrow \mathsf{Hom}_{\mathsf{KGL}_{**}}(\mathsf{KGL}_{**}\mathsf{KGL}, \mathsf{KGL}_{**})$$

is an isomorphism of KGL^{**}*-algebras.*

(ii) With the completed tensor product there is an isomorphism of KGL^{**}*-algebras*

$$\mathsf{KGL}^{**}\mathsf{KGL} \cong \mathsf{KGL}^{**}\widehat{\otimes}_{\mathsf{KU}^*}\mathsf{KU}^*\mathsf{KU}$$

Item (i) *and the module part of* (ii) *generalize to* $\mathsf{KGL}^{**}(\mathsf{KGL}^{\wedge j})$ *for* $j > 1$.

Proof. Recall $\mathsf{KU}_*\mathsf{KU}$ is free over KU_* [2] and KGL is the Landweber theory determined by the MU_*-algebra $\mathsf{MU}_* \to \mathbf{Z}[\beta, \beta^{-1}]$ which classifies the multiplicative formal group law $\mathsf{x}+\mathsf{y}-\beta\mathsf{xy}$ over $\mathbf{Z}[\beta, \beta^{-1}]$ with $|\beta| = 2$ [35, Theorem 1.2]. The corresponding topological Landweber exact theory is KU by the Conner-Floyd theorem. Thus by Proposition 9.1 (i) $\mathsf{KGL}_{**}\mathsf{KGL}$ is free over KGL_{**}. Moreover, KGL has the structure of an E_∞-motivic ring spectrum, see [9], [35], so the Universal coefficient spectral sequence in [7, Proposition 7.7] can be applied to the KGL-modules $\mathsf{KGL} \wedge \mathsf{KGL}$ and KGL; it converges conditionally [5], [21], and with abutment $\mathsf{Hom}^{**}_{\mathsf{KGL}-\mathsf{mod}}(\mathsf{KGL} \wedge \mathsf{KGL}, \mathsf{KGL}) = \mathsf{Hom}^{**}_{\mathbf{SH}(S)}(\mathsf{KGL}, \mathsf{KGL})$. But the spectral sequence degenerates since $\mathsf{KGL}_{**}\mathsf{KGL}$ is a free KGL_{**}-module. Hence items (i) and (ii) hold for $j = 1$.
The more general statement is proved along the same lines by noting the isomorphism

$$\mathsf{E}^{\mathsf{Top}}_*((\mathsf{E}^{\mathsf{Top}})^{\wedge j}) \cong \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}} \otimes_{\mathsf{E}^{\mathsf{Top}}_*} \cdots \otimes_{\mathsf{E}^{\mathsf{Top}}_*} \mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}},$$

and similarly for the Adams graded and Adams graded graded motivic versions. □

In stable homotopy theory there is a universal coefficient spectral sequence for every Landweber exact ring theory [15, Proposition 2.21]. It appears there is no direct motivic analog: While there is a reasonable notion of evenly generated motivic spectrum as in [15, Definition 2.10] and one can show that a motivic spectrum representing a Landweber exact theory is evenly generated as in [15, Proposition 2.12], this does not have as strong consequences as in topology because the coefficient ring MGL_* is not concentrated in even degrees as MU_*, but see Theorem 9.7 below. We aim to extend the above results on homotopy algebraic K-theory to more general Landweber exact motivic spectra.

Proposition 9.4: *Suppose* M *is a Tate object and* E *an* MGL*-module. Then there is a trigraded conditionally convergent right half-plane cohomological spectral sequence*

$$\mathsf{E}_2^{a,(p,q)} = \mathsf{Ext}^{a,(p,q)}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{M}, \mathsf{E}_{**}) \Rightarrow \mathsf{E}^{a+p,q}\mathsf{M}.$$

Proof. $\mathsf{MGL} \wedge \mathsf{M}$ is a cellular MGL-module so this follows from [7, Proposition 7.10]. □

The differentials in the spectral sequence go

$$d_r\colon \mathsf{E}_r^{a,(p,q)} \longrightarrow \mathsf{E}_r^{a+r,(p-r+1,q)}.$$

Theorem 9.5: *Suppose M_* is a Landweber exact graded MU_*-module concentrated in even degrees and $\mathsf{M} \in \mathbf{SH}(S)_{\mathcal{T}}$ represents the corresponding motivic cohomology theory. Then for $p,q \in \mathbf{Z}$ and N an MGL-module spectrum there is a short exact sequence*

$$0 \to \mathsf{Ext}^{1,(p-1,q)}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{M}, \mathsf{N}_{**}) \to \mathsf{N}^{p,q}\mathsf{M} \xrightarrow{\pi} \mathsf{Hom}^{p,q}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{M}, \mathsf{N}_{**}) \to 0.$$

Proof. Let $\mathsf{M}^{\mathsf{Top}}$ be the topological spectrum associated with M_*. Then $\mathsf{MU}_*\mathsf{M}^{\mathsf{Top}}$ is a flat MU_*-module of projective dimension at most one, see [15, Propositions 2.12, 2.16]. Hence $\mathsf{MGL}_{**}\mathsf{M} = \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{MU}_*\mathsf{M}^{\mathsf{Top}}$ is a MGL_{**}-module of projective dimension at most one and consequently the spectral sequence of Proposition 9.4 degenerates at its E_2-page. This implies the derived $\lim^1$-term $\lim^1 \mathsf{E}_r^{***}$ of the spectral sequence is zero; hence it converges strongly. The assertion follows because $\mathsf{E}_\infty^{p,**} = 0$ for all $p \neq 0, 1$. □

Remark 9.6: *(i) For $p,q \in \mathbf{Z}$, the group of phantom maps $\mathsf{Ph}^{p,q}(\mathsf{M},\mathsf{N}) \subseteq \mathsf{N}^{p,q}\mathsf{M}$ is defined as $\{\mathsf{S}^{p,q} \wedge \mathsf{M} \xrightarrow{\varphi} \mathsf{N} \mid$ for all $\mathsf{E} \in \mathbf{SH}(S)_{\mathcal{T},\mathsf{f}}$ and $\mathsf{E} \xrightarrow{\nu} \mathsf{S}^{p,q} \wedge \mathsf{M} : \varphi\nu = 0\}$. It is clear that $\mathsf{Ph}^{p,q}(\mathsf{M},\mathsf{N}) \subseteq \ker(\pi)$.*

(ii) The following topological example due to Strickland shows a nontrivial Ext^1-term. The canonical map $\mathsf{KU}_{(p)} \to \mathsf{KU}_p$ from p-local to p-complete unitary topological K-theory yields a cofiber sequence

$$\mathsf{KU}_{(p)} \longrightarrow \mathsf{KU}_p \longrightarrow \mathsf{E} \xrightarrow{\delta} \Sigma\mathsf{KU}_{(p)}.$$

Here E is rational and thus Landweber exact. Thus δ is a degree 1 map between Landweber exact spectra.

However, δ is a nonzero phantom map.

Over fields embeddable into $\mathbf{C}$ the corresponding boundary map for the motivic Landweber spectra is likewise phantom and non-zero. Using the notion of heights for Landweber exact algebras from [25, Section 5], observe that E has height zero while $\Sigma\mathsf{KU}_{(p)}$ has height one, compare with the assumptions in Theorem 9.7 below.

Now fix Landweber exact MU_*-algebras E_* and F_* concentrated in even degrees and a 2-commutative diagram

$$\begin{tikzcd} \mathsf{Spec}(\mathsf{F}_*) \arrow[rr, "f"] \arrow[dr, "f_{\mathsf{F}}"'] & & \mathsf{Spec}(\mathsf{E}_*) \arrow[dl, "f_{\mathsf{E}}"] \\ & \mathfrak{X} & \end{tikzcd} \tag{30}$$

where $\mathfrak{X}$ is the stack of formal groups and f_{F} (resp. f_{E}) the map classifying the formal group G_{F} (resp. G_{E}) canonically associated with the complex orientable cohomology theory corresponding to F_* (resp. E_*). This entails an isomorphism $f^*G_{\mathsf{E}} \cong G_{\mathsf{F}}$ of formal groups over $\mathsf{Spec}(\mathsf{F}_*)$. Hence the height of F_* is less or equal to the height of E_*. Let $\mathsf{E}^{\mathsf{Top}}, \mathsf{F}^{\mathsf{Top}}$ (resp. $\mathsf{E}, \mathsf{F} \in \mathbf{SH}(S)_{\mathcal{T}}$) be the topological (resp. motivic) spectra representing the indicated Landweber exact cohomology theory.

Theorem 9.7: *With the notation above assume* $\mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}$ *is a projective* $\mathsf{E}^{\mathsf{Top}}_*$-*module.*

(i) The map from Theorem 9.5

$$\pi : \mathsf{F}^{**}\mathsf{E} \longrightarrow \mathsf{Hom}^{**}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{E}, \mathsf{F}_{**}) \cong \mathsf{Hom}_{\mathsf{E}^{\mathsf{Top}}_*}(\mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}, \mathsf{F}_{**})$$

is an isomorphism.

(ii) Under the isomorphism in (i), *the bidegree* $(0,0)$ *maps* $\mathsf{S}^{*,*} \wedge \mathsf{E} \to \mathsf{F}$ *which respect the quasi-multiplication correspond bijectively to maps of* $\mathsf{E}^{\mathsf{Top}}_*$-*algebras*

$$\mathsf{Hom}_{\mathsf{E}^{\mathsf{Top}}_*-\mathsf{alg}}(\mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}, \mathsf{F}_{**}).$$

Remark 9.8: *(i) The assumptions in Theorem 9.7 hold when* $\mathsf{E}^{\mathsf{Top}} = \mathsf{KU}$ *and for certain localizations of Johnson-Wilson theories according to [2] respectively [3]. Theorem 9.7 recovers Theorem 9.3 with no mention of an* E_∞-*structure on* KGL.

(ii) The theorem applies to the quasi-multiplication $(\mathsf{E} \wedge \mathsf{E} \to \mathsf{E}) \in \mathsf{E}^{00}(\mathsf{E} \wedge \mathsf{E})$ *and shows that this is a commutative monoid structure which lifts uniquely the multiplication on the homology theory. For example, there is a unique structure of commutative monoid on* $\mathsf{KGL}_S \in \mathbf{SH}(S)$ *representing the familiar multiplicative structure of homotopy* K-*theory, see [30] for a detailed account and an independent proof in the case* $S = \mathsf{Spec}(\mathbf{Z})$.

(iii) The composite map $\alpha : \mathsf{E}_* \xrightarrow{f} \mathsf{F}_* \to \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{F}_* = \mathsf{F}_{**}$ *yields a canonical bijection between the sets* $\mathsf{Hom}_{\mathsf{E}^{\mathsf{Top}}_*-\mathsf{alg}}(\mathsf{E}^{\mathsf{Top}}_*\mathsf{E}^{\mathsf{Top}}, \mathsf{F}_{**})$ *and* $\{(\alpha', \varphi)\}$, *where* $\alpha' \colon \mathsf{E}_* \to \mathsf{F}_{**}$ *is a ring homomorphism and* $\varphi \colon \alpha_* G_{\mathsf{E}} \to \alpha'_* G_{\mathsf{E}}$ *a strict isomorphism of strict formal groups.*

(iv) Taking $\mathsf{F} = \mathsf{E}$ *in Theorem 9.7 and using Remark 9.6*(i) *implies that* $\mathsf{Ph}^{**}(\mathsf{E}, \mathsf{E}) = 0$. *For example, there are no nontrivial phantom maps* $\mathsf{KGL} \to \mathsf{KGL}$ *of any bidegree.*

Proof. (of Theorem 9.7): We shall apply Proposition 2.3 with $X_0 \equiv \mathsf{Spec}(\mathsf{MU}_*)$, $X \equiv \mathsf{Spec}(\mathsf{F}_*)$, $Y \equiv \mathsf{Spec}(\mathsf{E}_*)$, $f_X \equiv f_\mathsf{F}$ and $f_Y \equiv f_\mathsf{E}$, $\pi : \mathsf{Spec}(\mathsf{MU}_*) \to \mathfrak{X}$ the map classifying the universal formal group, f as given by (30) and $\alpha : X = \mathsf{Spec}(\mathsf{F}_*) \to X_0 = \mathsf{Spec}(\mathsf{MU}_*)$ corresponding to the MU_*-algebra structure $\mathsf{MU}_* \to \mathsf{F}_*$. Now by [25, Theorem 26], f_X (resp. f_Y) factors as $f_X = i_X \circ \pi_X$ (resp. $f_Y = i_Y \circ \pi_Y$) with π_X and π_Y faithfully flat and i_X and i_Y inclusions of open substacks. The map i in Proposition 2.3 is induced by f. Finally, MGL_{**} is canonically an $\mathsf{MU}_*\mathsf{MU}$-comodule algebra and the $\mathcal{O}_\mathfrak{X}$-algebra $\mathcal{A}$ in Proposition 2.3 corresponds to MGL_{**}, i.e. $\mathcal{A}(X_0) = \mathsf{MGL}_{**}$ and $\pi_Y^*\pi_{Y,*}\mathcal{O}_Y \in \mathsf{Qc}_Y$ to the projective $\mathsf{E}_*^{\mathsf{Top}}$-module $\mathsf{E}_*^{\mathsf{Top}}\mathsf{E}^{\mathsf{Top}}$. Taking into account the isomorphisms

$$\mathcal{A}(X_0) \otimes_{\mathcal{O}_{X_0}} \pi^* f_{Y,*}\mathcal{O}_Y \cong \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{MU}_*^{\mathsf{Top}}\mathsf{E}^{\mathsf{Top}} \cong \mathsf{MGL}_{**}\mathsf{E}$$

$$\mathcal{A}(X_0) \otimes_{\mathcal{O}_{X_0}} \alpha_*\mathcal{O}_X \cong \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathsf{F}_*^{\mathsf{Top}} \cong \mathsf{F}_{**}$$

$$\pi_Y^*\pi_{Y,*}\mathcal{O}_Y \cong \mathsf{E}_*^{\mathsf{Top}}\mathsf{E}^{\mathsf{Top}}$$

$$\mathcal{A}(Y) \otimes_{\mathcal{O}_Y} f_*\mathcal{O}_X \cong \mathsf{F}_{**}$$

$$\mathcal{O}_Y \cong \mathsf{E}_*^{\mathsf{Top}}$$

we obtain from Proposition 2.3

$$\mathsf{Ext}^n_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{E}, \mathsf{F}_{**}) \cong \begin{cases} 0 & n \geq 1, \\ \mathsf{Hom}_{\mathsf{E}_*^{\mathsf{Top}}}(\mathsf{E}_*^{\mathsf{Top}}\mathsf{E}^{\mathsf{Top}}, \mathsf{F}_{**}) & n = 0. \end{cases}$$

Hence (i) follows from Theorem 9.5 and (ii) by unwinding the definitions. □

10 The Chern character

In what follows we define a ring map from KGL to periodized rational motivic cohomology which induces the Chern character (or regulator map) from K-theory to (higher) Chow groups when the base scheme is smooth over a field. Let $\mathsf{M}\mathbf{Z}$ denote the integral motivic Eilenberg-MacLane ring spectrum introduced by Voevodsky [39, §6.1], cf. [8, Example 3.4]. Next we give a canonical orientation on $\mathsf{M}\mathbf{Z}$, in particular the construction of a map $\mathbb{P}^\infty_+ \to K(\mathbf{Z}(1), 2) = L((\mathbb{P}^1, \infty))$.

Recall the space $L(X)$ assigns to any U the group of proper relative cycles on $U \times_S X$ over U of relative dimension 0 which have universally integral coefficients. Now the line bundle $\mathcal{O}_{\mathbb{P}^n}(1) \boxtimes \mathcal{O}_{\mathbb{P}^1}(n)$ acquires the section $l_n \equiv T_n x_0^n + T_{n-1}x_0^{n-1}x_1 + \cdots + T_0 x_1^n$, where $[T_0 : \cdots : T_n]$ denotes homogeneous coordinates on $\mathbb{P}^n$ and $[x_0 : x_1]$ on $\mathbb{P}^1$. Its zero locus is a relative divisor of degree n on $\mathbb{P}^1$ which induces a map $\mathbb{P}^n \to L(\mathbb{P}^1)$. These maps combine to give maps $\mathbb{P}^n \to L((\mathbb{P}^1, \infty))$ which are compatible with the inclusions $\mathbb{P}^n \to \mathbb{P}^{n+1}$.

Hence there is an induced map $\varphi\colon \mathbb{P}^\infty \to K(\mathbf{Z}(1),2)$. Moreover, the map $\mathbb{P}^n \to L(\mathbb{P}^1)$ is additive with respect to the maps $\mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{n+m}$ induced by multiplication by the section l_n. Hence φ is a map of commutative monoids and it restricts to the canonical map $\mathbb{P}^1 \to K(\mathbf{Z}(1),2)$. This establishes an orientation on $\mathsf{M}\mathbf{Z}$ with the additive formal group law.

Let $\mathsf{M}\mathbf{Q}$ be the rationalization of $\mathsf{M}\mathbf{Z}$. In order to apply the spectral sequence in Proposition 9.4 to $\mathsf{M}\mathbf{Q}$ we equip it with an MGL-module structure. Note that both $\mathsf{M}\mathbf{Z}$ and $\mathsf{M}\mathbf{Q}$ have canonical E_∞-structures. Thus $\mathsf{M}\mathbf{Q}\wedge\mathsf{MGL}$ is also E_∞. As an $\mathsf{M}\mathbf{Q}$-module it has the form $\mathsf{M}\mathbf{Q}[b_1,b_2,\ldots]$. For any generator b_i we let $\iota_i\colon \Sigma^{2i,i}\mathsf{M}\mathbf{Q} \to \mathsf{M}\mathbf{Q}\wedge\mathsf{MGL}$ denote the corresponding map. Taking its adjoint provides a map from the free $\mathsf{M}\mathbf{Q}$-E_∞-algebra on $\bigvee_{i>0}\mathsf{S}^{2i,i}$ to $\mathsf{M}\mathbf{Q}\wedge\mathsf{MGL}$. Since we are dealing with rational coefficients the contraction of these cells in E_∞-algebras is isomorphic to $\mathsf{M}\mathbf{Q}$. Hence there is a map $\mathsf{MGL}\to\mathsf{M}\mathbf{Q}$ in E_∞-algebras. This gives in particular an MGL-module structure on $\mathsf{M}\mathbf{Q}$.

Let $\mathsf{PM}\mathbf{Q}$ be the periodized rational Eilenberg-MacLane spectrum considered as an MGL-module, and $\mathsf{L}\mathbf{Q}$ the Landweber spectrum corresponding to the additive formal group law over $\mathbf{Q}$. By Remark 9.8 $\mathsf{L}\mathbf{Q}$ is a ring spectrum. We let $\mathsf{PL}\mathbf{Q}$ be the periodic version. Both $\mathsf{L}\mathbf{Q}$ and $\mathsf{PL}\mathbf{Q}$ have canonical structures of MGL-modules. Finally, let $\mathsf{PH}\mathbf{Q}$ be the periodized rational topological Eilenberg-MacLane spectrum.

Recall the map $\mathsf{Ch}^{\mathsf{PH}}_*\colon \mathsf{KU}_*\to\mathsf{PH}\mathbf{Q}_*$ sending the Bott element to the canonical element in degree 2. The exponential map establishes an isomorphism from the additive formal group law over $\mathsf{PH}\mathbf{Q}_*$ to the pushforward of the multiplicative formal group law over KU_* with respect to $\mathsf{Ch}^{\mathsf{PH}}_*$. By Theorem 9.7 and Remark 9.8(*iii*) there is an induced map of motivic ring spectra $\mathsf{Ch}^{\mathsf{PL}}\colon\mathsf{KGL}\to\mathsf{PL}\mathbf{Q}$.

Theorem 10.1: *The rationalization*

$$\mathsf{Ch}^{\mathsf{PL}}_{\mathbf{Q}}\colon \mathsf{KGL}_{\mathbf{Q}} \longrightarrow \mathsf{PL}\mathbf{Q}$$

of the map $\mathsf{Ch}^{\mathsf{PL}}$ *from* KGL *to* $\mathsf{PL}\mathbf{Q}$ *is an isomorphism.*

Proof. Follows directly from the fact that the rationalization of $\mathsf{Ch}^{\mathsf{PH}}_*$ is an isomorphism. □

Theorem 9.5 shows there is a short exact sequence

$$0 \longrightarrow \mathsf{Ext}^{1,(p-1,q)}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{L}\mathbf{Q},\mathsf{M}\mathbf{Q}_{**}) \longrightarrow \mathsf{M}\mathbf{Q}^{p,q}\mathsf{L}\mathbf{Q}$$

$$\xrightarrow{\pi} \mathsf{Hom}^{p,q}_{\mathsf{MGL}_{**}}(\mathsf{MGL}_{**}\mathsf{L}\mathbf{Q},\mathsf{M}\mathbf{Q}_{**}) \longrightarrow 0.$$

Now since $\mathsf{M}\mathbf{Q}$ has the additive formal group law there is a natural transformation of homology theories

$$\mathsf{L}\mathbf{Q}_{**}(-) \longrightarrow \mathsf{M}\mathbf{Q}_{**}(-). \tag{31}$$

Applying the methods of Theorem 9.7 to $\mathsf{E} = \mathsf{L}\mathbf{Q}$ and $\mathsf{F} = \mathsf{M}\mathbf{Q}$ shows that (31) lifts uniquely to a map of motivic ring spectra

$$\iota\colon \mathsf{L}\mathbf{Q} \longrightarrow \mathsf{M}\mathbf{Q}.$$

It prolongs to a map of motivic ring spectra $\mathsf{PL}\mathbf{Q} \to \mathsf{PM}\mathbf{Q}$ (denoted by the same symbol).
The composite map

$$\mathsf{Ch}^{\mathsf{PM}}\colon \mathsf{KGL} \xrightarrow{\mathsf{Ch}^{\mathsf{PL}}} \mathsf{PL}\mathbf{Q} \xrightarrow{\iota} \mathsf{PM}\mathbf{Q}$$

is called the Chern character. By construction it is functorial in the base scheme with respect to the natural map $\mathbf{L}f^*\mathsf{PM}\mathbf{Q}_S \to \mathsf{PM}\mathbf{Q}_{S'}$ for $f\colon S' \to S$.
Recall that for smooth schemes over fields motivic cohomology coincides with higher Chow groups [40].

Proposition 10.2: *Evaluated on smooth schemes over fields the map $\mathsf{Ch}^{\mathsf{PM}}$ coincides with the usual Chern character from K-theory to higher Chow groups.*

Proof. The construction of the Chern character in [4] and [18] uses the methods of [10]. We first show that the individual Chern class transformations C_i in loc. cit. from K-theory to the cohomology theory in question can be extended to a transformation between simplicial presheaves on smooth affine schemes over the given field k. Fix a cofibration

$$\mathrm{BGL}(\mathbf{Z}) \longrightarrow \mathrm{BGL}^+(\mathbf{Z}).$$

The simplicial presheaf

$$\mathsf{Spec}(A) \mapsto \Gamma(A) \equiv \mathbf{Z} \times \mathrm{BGL}(A) \cup_{\mathrm{BGL}(\mathbf{Z})} \mathrm{BGL}^+(\mathbf{Z})$$

represents K-theory, see [4]. The Chern class C_i of the universal vector bundle on the sheaf $\mathrm{BGL}(-)$ can be represented by a transformation of simplicial presheaves $\mathrm{BGL}(-) \to K(i)$, where $K(i)$ denotes an injectively fibrant presheaf of simplicial abelian groups representing motivic cohomology with coefficients in $\mathbf{Q}(i)$ with the appropriate simplicial shift. The map $\mathrm{BGL}(k) \to K(i)(k)$ extends to

$$\Gamma(k) \longrightarrow K(i)(k).$$

By definition of the presheaf Γ we get the required map. Having achieved this, the Chern class transformations C_i extend to functors on the full subcategory $\mathcal{F}$ of objects of finite type in the sense of [39] in the $\mathbf{A}^1$-local homotopy category. Denote by $j\colon \mathcal{F} \to \mathbf{SH}(k)$ the canonical functor.
With the above observations as prelude, it follows that these transformations induce a multiplicative Chern character transformation

$$\tau\colon \Gamma(-) \longrightarrow \mathsf{PM}\mathbf{Q}_{00}(-) \circ j$$

on this category. The source and target of τ are $\mathbf{P}^1$-periodic and τ is compatible with these. Hence there is an induced transformation on the Karoubian envelope of the Spanier-Whitehead stabilization with respect to the pointed $\mathbf{P}^1$, which is the full subcategory of $\mathbf{SH}(k)$ of compact objects according to [39, Propositions 5.3 and 5.5]. But as a cohomology theory on compact objects, KGL is the universal oriented theory which is multiplicative for the formal group law. To conclude the proof, it is now sufficient to note that the transformation constructed above has the same effect on the universal first Chern class as $\mathsf{Ch}^{\mathsf{PM}}$ does, which is clear. □

For smooth quasi-projective schemes over fields the Chern character is known to be an isomorphism after rationalization [4], hence our transformation $\mathsf{Ch}^{\mathsf{PM}}$ is an isomorphism after rationalization (a map $\mathsf{E} \to \mathsf{F}$ between periodic spectra is an isomorphism if it induces isomorphisms $\mathsf{E}^{-i,0}(X) \to \mathsf{F}^{-i,0}(X)$ for all smooth schemes X over S and $i \geq 0$). By Mayer-Vietoris the same holds for smooth schemes over fields.

Corollary 10.3: *For smooth schemes over fields the map*

$$\iota\colon \mathsf{L}\mathbf{Q} \longrightarrow \mathsf{M}\mathbf{Q}$$

is an isomorphism of motivic ring spectra.

Corollary 10.4: *For smooth schemes over fields*

$$\mathsf{M}\mathbf{Q}_{**}(-)$$

is the universal oriented homology theory with rational coefficients and additive formal group law.

Next we identify the rationalization $\mathsf{MGL}_{\mathbf{Q}}$ of the algebraic cobordism spectrum:

Theorem 10.5: *There are isomorphisms of motivic ring spectra*

$$\mathsf{MGL}_{\mathbf{Q}} \cong \mathsf{MGL} \wedge \mathsf{L}\mathbf{Q} \cong \mathsf{L}\mathbf{Q}[b_1, \ldots],$$

where the generator b_i has bidegree $(2i, i)$ for every $i \geq 1$.

Proof. According to Remark 9.2 $\mathsf{MGL} \wedge \mathsf{L}\mathbf{Q}$ is the motivic Landweber exact spectrum associated with $\mathsf{MU} \wedge \mathsf{H}\mathbf{Q} \cong \mathsf{MU}_{\mathbf{Q}}$; this implies the first isomorphism. In homotopy, the canonical map of ring spectra $\mathsf{L}\mathbf{Q} \to \mathsf{MGL} \wedge \mathsf{L}\mathbf{Q}$ yields

$$\pi_{**}\mathsf{L}\mathbf{Q} = \mathsf{MGL}_{**} \otimes_{\mathsf{MU}_*} \mathbf{Q} \to \pi_{**}(\mathsf{MGL} \wedge \mathsf{L}\mathbf{Q}) = \\ = \mathsf{MGL}_{**}\mathsf{MGL} \otimes_{\mathsf{MU}_*} \mathbf{Q} = \pi_{**}\mathsf{L}\mathbf{Q}[b_1, \ldots].$$

Hence there is a map of ring spectra $\mathsf{L}\mathbf{Q}[b_1, \ldots] \to \mathsf{MGL} \wedge \mathsf{L}\mathbf{Q}$ under $\mathsf{L}\mathbf{Q}$ which is an π_{**}-isomorphism. Since all spectra above are cellular the second isomorphism follows. $\square$

Corollary 10.6: *Suppose S is smooth over a field.*

(i) *There are isomorphisms of motivic ring spectra*

$$\mathsf{MGL}_{\mathbf{Q}} \cong \mathsf{MGL} \wedge \mathsf{M}\mathbf{Q} \cong \mathsf{M}\mathbf{Q}[b_1, \ldots].$$

(ii) *For X/S smooth and L^* the (graded) Lazard ring, there is an isomorphism*

$$\mathsf{MGL}^{**}(X) \otimes_{\mathbf{Z}} \mathbf{Q} \cong \mathsf{M}\mathbf{Q}^{**}(X) \otimes_{\mathbf{Z}} \mathsf{L}^*.$$

Proof. Part (i) is immediate from Theorem 10.5, specialized to smooth schemes over fields, and Corollary 10.3. Part (ii) follows from (i) using compactness of X. $\square$

As alluded to in the introduction we may now explicate the rationalized algebraic cobordism of number fields. The answer is conveniently formulated in terms of the (graded) Lazard ring $\mathsf{L}^* = \mathbf{Z}[x_1, x_2, \ldots]$ with its cohomological grading $|x_i| = (-2i, -i)$, $i \geq 1$.

Corollary 10.7: *Suppose k is a number field with r_1 real embeddings and r_2 pairs of complex embeddings. Then there are isomorphisms*

$$\mathsf{MGL}^{2i,j}(k) \otimes \mathbf{Q} \cong \begin{cases} \mathsf{L}^{2i} \otimes \mathbf{Q} & j = i \\ 0 & j \neq i \end{cases}$$

$$\mathsf{MGL}^{2i+1,j}(k) \otimes \mathbf{Q} \cong \begin{cases} \mathsf{L}^{2i} \otimes k^* \otimes \mathbf{Q} & j = i+1, i \leq 0 \\ \mathsf{L}^{2i} \otimes \mathbf{Q}^{r_2} & j - i \equiv 3\,(4), j - i > 1 \\ \mathsf{L}^{2i} \otimes \mathbf{Q}^{r_1 + r_2} & j - i \equiv 1\,(4), j - i > 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. Follows from Corollary 10.6(ii) and the well-known computation of the rational motivic cohomology of number fields. $\square$

Remark 10.8: *In Corollary 10.3 we identified the (unique) Landweber exact motivic spectrum* LQ *with rational motivic cohomology* MQ *(for base schemes smooth over some field). The topological analog of this result is a triviality because* HQ *is the Landweber exact spectrum associated with the additive formal group over* $\mathbf{Q}$*. To appreciate the content of Corollary 10.3, we offer the following remark: In stable homotopy theory it is trivial that* $S^0_{\mathbf{Q}} \cong \mathsf{HQ}$ *but the motivic analog of this result fails. Let* $\mathbf{1}_{\mathbf{Q}}$ *denote the rationalized motivic sphere spectrum. Using orthogonal idempotents, Morel [22] has constructed a splitting*

$$\mathbf{1}_{\mathbf{Q}} \cong \mathbf{1}^+_{\mathbf{Q}} \vee \mathbf{1}^-_{\mathbf{Q}}$$

and noted that $\mathbf{1}^-_{\mathbf{Q}}$ *is nontrivial for formally real fields (e.g. the rational numbers). It is easy to show that every map from the motivic sphere spectrum to an oriented motivic ring spectrum annihilates* $\mathbf{1}^-_{\mathbf{Q}}$*. In particular,* $\mathbf{1}_{\mathbf{Q}}$ *and* LQ *are not isomorphic in general.*

References

[1] J. F. Adams. *Stable homotopy and generalised homology.* University of Chicago Press, Chicago, Ill., 1974. Chicago Lectures in Mathematics.

[2] J. F. Adams and F. W. Clarke. Stable operations on complex K-theory. *Illinois J. Math.*, 21(4):826–829, 1977.

[3] A. Baker. I_n-local Johnson-Wilson spectra and their Hopf algebroids. *Doc. Math.*, 5:351–364 (electronic), 2000.

[4] S. Bloch. Algebraic cycles and higher K-theory. *Adv. in Math.*, 61(3):267–304, 1986.

[5] J. M. Boardman. Conditionally convergent spectral sequences. In *Homotopy invariant algebraic structures (Baltimore, MD, 1998)*, volume 239 of *Contemp. Math.*, pages 49–84. Amer. Math. Soc., Providence, RI, 1999.

[6] S. Borghesi. Algebraic Morava K-theories. *Invent. Math.*, 151(2):381–413, 2003.

[7] D. Dugger and D. C. Isaksen. Motivic cell structures. *Algebr. Geom. Topol.*, 5:615–652 (electronic), 2005.

[8] B. I. Dundas, O. Röndigs, and P. A. Østvær. Motivic functors. *Doc. Math.*, 8:489–525 (electronic), 2003.

[9] D. Gepner and V. Snaith. On the motivic spectra representing algebraic cobordism and algebraic K-theory. *Doc. Math.*, 14:359–396 (electronic), 2009.

[10] H. Gillet. Riemann-Roch theorems for higher algebraic K-theory. *Adv. in Math.*, 40(3):203–289, 1981.

[11] P. G. Goerss. Quasi-coherent sheaves on the moduli stack of formal groups. Preprint, arXiv 0802.0996.

[12] P. G. Goerss. (Pre-)sheaves of ring spectra over the moduli stack of formal group laws. In *Axiomatic, enriched and motivic homotopy theory*, volume 131 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 101–131. Kluwer Acad. Publ., Dordrecht, 2004.

[13] J. Hornbostel. Localizations in motivic homotopy theory. *Math. Proc. Cambridge Philos. Soc.*, 140(1):95–114, 2006.

[14] M. Hovey, J. H. Palmieri, and N. P. Strickland. Axiomatic stable homotopy theory. *Mem. Amer. Math. Soc.*, 128(610):x+114, 1997.

[15] M. Hovey and N. P. Strickland. Morava K-theories and localisation. *Mem. Amer. Math. Soc.*, 139(666):viii+100, 1999.

[16] P. Hu and I. Kriz. Some remarks on Real and algebraic cobordism. *K-Theory*, 22(4):335–366, 2001.

[17] J. F. Jardine. Motivic symmetric spectra. *Doc. Math.*, 5:445–553 (electronic), 2000.

[18] M. Levine. Bloch's higher Chow groups revisited. *Astérisque*, (226):10, 235–320, 1994. K-theory (Strasbourg, 1992).

[19] M. Levine and F. Morel. *Algebraic cobordism.* Springer Monographs in Mathematics. Springer, Berlin, 2007.

[20] J. P. May. Idempotents and Landweber exactness in brave new algebra. *Homology Homotopy Appl.*, 3(2):355–359 (electronic), 2001. Equivariant stable homotopy theory and related areas (Stanford, CA, 2000).

[21] J. McCleary. *A user's guide to spectral sequences*, volume 58 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, second edition, 2001.

[22] F. Morel. Rational stable splitting of grassmanians and the rational motivic sphere spectrum. Preprint, http://www.mathematik.uni-muenchen.de/∼morel/ Splittinggrassman.pdf.

[23] F. Morel. Some basic properties of the stable homotopy category of schemes. Preprint, http://www.mathematik.uni-muenchen.de/∼morel/Stable.pdf.

[24] F. Morel and V. Voevodsky. $\mathbf{A}^1$-homotopy theory of schemes. *Inst. Hautes Études Sci. Publ. Math.*, (90):45–143 (2001), 1999.

[25] N. Naumann. The stack of formal groups in stable homotopy theory. *Adv. Math.*, 215(2):569–600, 2007.

[26] N. Naumann and M. Spitzweck. Brown representability in $\mathbf{A}^1$-homotopy theory. K-theory Preprint Archives, 942, 2009.

[27] N. Naumann, M. Spitzweck, and P. A. Østvær. Chern classes, K-theory and landweber exactness over nonregular base schemes. In *Motives and Algebraic Cycles: A celebration in Honour of Spencer J. Bloch*, volume 56 of *Fields Institute Communications*, pages 307–317. Amer. Math. Soc., Providence, RI, 2009.

[28] A. Neeman. The Grothendieck duality theorem via Bousfield's techniques and Brown representability. *J. Amer. Math. Soc.*, 9(1):205–236, 1996.

[29] I. Panin, K. Pimenov, and O. Röndigs. A universality theorem for Voevodsky's algebraic cobordism spectrum. *Homology, Homotopy Appl.*, 10(2):211–226, 2008.

[30] I. Panin, K. Pimenov, and O. Röndigs. On Voevodsky's algebraic K-theory spectrum. In *Algebraic Topology*, The Abel Symposium 2007, pages 279–330. Springer, Berlin, 2009.

[31] O. Röndigs and P. A. Østvær. Modules over motivic cohomology. *Adv. Math.*, 219(2):689–727, 2008.

[32] O. Röndigs and P. A. Østvær. Rigidity in motivic homotopy theory. *Math. Ann.*, 341(3):651–675, 2008.

[33] S. Schwede and B. E. Shipley. Algebras and modules in monoidal model categories. *Proc. London Math. Soc. (3)*, 80(2):491–511, 2000.

[34] M. Spitzweck. Slices of motivic Landweber spectra. Preprint, arXiv 0805.3350.

[35] M. Spitzweck and P. A. Østvær. The Bott inverted infinite projective space is K-theory. *Bull. London Math. Soc.*, 41(2):281–292, 2009.

[36] R. M. Switzer. *Algebraic topology—homotopy and homology.* Classics in Mathematics. Springer-Verlag, Berlin, 2002. Reprint of the 1975 original [Springer, New York; MR0385836 (52 #6695)].

[37] K. Ueno. *Algebraic geometry. 2*, volume 197 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI, 2001. Sheaves and cohomology, Translated from the 1997 Japanese original by Goro Kato, Iwanami Series in Modern Mathematics.

[38] G. Vezzosi. Brown-Peterson spectra in stable $\mathbb{A}^1$-homotopy theory. *Rend. Sem. Mat. Univ. Padova*, 106:47–64, 2001.

[39] V. Voevodsky. $\mathbf{A}^1$-homotopy theory. In *Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998)*, number Extra Vol. I, pages 579–604 (electronic), 1998.

[40] V. Voevodsky. Motivic cohomology groups are isomorphic to higher Chow groups in any characteristic. *Int. Math. Res. Not.*, (7):351–355, 2002.

[41] V. Voevodsky. Open problems in the motivic stable homotopy theory. I. In *Motives, polylogarithms and Hodge theory, Part I (Irvine, CA, 1998)*, volume 3 of *Int. Press Lect. Ser.*, pages 3–34. Int. Press, Somerville, MA, 2002.

Niko Naumann
Fakultät für Mathematik
Universität Regensburg
Germany
niko.naumann@
mathematik.uni-regensburg.de

Markus Spitzweck
Fakultät für Mathematik
Universität Regensburg
Germany
Markus.Spitzweck@
mathematik.uni-regensburg.de

Paul Arne Østvær
Department of Mathematics
University of Oslo
Norway
paularne@math.uio.no

Hyperbolic Geometry on Noncommutative Balls

Gelu Popescu[1]

Received: March 6, 2009
Revised: October 31, 2009

Communicated by Joachim Cuntz

Abstract. In this paper, we study the noncommutative balls

$$\mathcal{C}_\rho := \{(X_1, \ldots, X_n) \in B(\mathcal{H})^n : \ \omega_\rho(X_1, \ldots, X_n) \leq 1\}, \qquad \rho \in (0, \infty],$$

where ω_ρ is the joint operator radius for n-tuples of bounded linear operators on a Hilbert space. In particular, ω_1 is the operator norm, ω_2 is the joint numerical radius, and ω_∞ is the joint spectral radius.

We introduce a Harnack type equivalence relation on $\mathcal{C}_\rho$, $\rho > 0$, and use it to define a hyperbolic distance δ_ρ on the Harnack parts (equivalence classes) of $\mathcal{C}_\rho$. We prove that the open ball

$$[\mathcal{C}_\rho]_{<1} := \{(X_1, \ldots, X_n) \in B(\mathcal{H})^n : \ \omega_\rho(X_1, \ldots, X_n) < 1\}, \qquad \rho > 0,$$

is the Harnack part containing 0 and obtain a concrete formula for the hyperbolic distance, in terms of the reconstruction operator associated with the right creation operators on the full Fock space with n generators. Moreover, we show that the δ_ρ-topology and the usual operator norm topology coincide on $[\mathcal{C}_\rho]_{<1}$. While the open ball $[\mathcal{C}_\rho]_{<1}$ is not a complete metric space with respect to the operator norm topology, we prove that it is a complete metric space with respect to the hyperbolic metric δ_ρ. In the particular case when $\rho = 1$ and $\mathcal{H} = \mathbb{C}$, the hyperbolic metric δ_ρ coincides with the Poincaré-Bergman distance on the open unit ball of $\mathbb{C}^n$.

We introduce a Carathéodory type metric on $[\mathcal{C}_\infty]_{<1}$, the set of all n-tuples of operators with joint spectral radius strictly less then 1, by setting

$$d_K(A, B) = \sup_p \|\Re p(A) - \Re p(B)\|, \qquad A, B \in [\mathcal{C}_\infty]_{<1},$$

where the supremum is taken over all noncommutative polynomials with matrix-valued coefficients $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, with

[1]Research supported in part by an NSF grant

$\Re p(0) = I$ and $\Re p(X) \geq 0$ for all $X \in \mathcal{C}_1$. We obtain a concrete formula for d_K in terms of the free pluriharmonic kernel on the noncommutative ball $[\mathcal{C}_\infty]_{<1}$. We also prove that the metric d_K is complete on $[\mathcal{C}_\infty]_{<1}$ and its topology coincides with the operator norm topology.

We provide mapping theorems, von Neumann inequalities, and Schwarz type lemmas for free holomorphic functions on noncommutative balls, with respect to the hyperbolic metric δ_ρ, the Carathéodory metric d_K, and the joint operator radius ω_ρ, $\rho \in (0, \infty]$.

2000 Mathematics Subject Classification: Primary: 46L52; 46T25; 47A20; Secondary: 32Q45; 47A56

Keywords and Phrases: Noncommutative hyperbolic geometry; Noncommutative function theory; Harnack part; Hyperbolic distance; Carathéodory distance; Free holomorphic function; Free pluriharmonic function; Fock space; Creation operator; Joint operator radius; Joint numerical radius; Joint spectral radius; von Neumann inequality; Schwarz lemma.

Contents

Introduction

In [48], we provided a generalization of the Sz.-Nagy–Foiaş theory of ρ-contractions (see [54], [55], [56]), to the multivariable setting. An n-tuple $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ of bounded linear operators acting on a Hilbert space

$\mathcal{H}$ belongs to the class $\mathcal{C}_\rho$, $\rho > 0$, if there is a Hilbert space $\mathcal{K} \supseteq \mathcal{H}$ and some isometries $V_i \in B(\mathcal{K})$, $i = 1, \dots, n$, with orthogonal ranges such that

$$T_\alpha = \rho P_\mathcal{H} V_\alpha|_\mathcal{H} \quad \text{for any } \alpha \in \mathbb{F}_n^+ \backslash \{g_0\},$$

where $P_\mathcal{H}$ is the orthogonal projection of $\mathcal{K}$ onto $\mathcal{H}$. Here, $\mathbb{F}_n^+$ stands for the unital free semigroup on n generators $g_1, \dots, g_n$, and the identity g_0, while $T_\alpha := T_{i_1} T_{i_2} \cdots T_{i_k}$ if $\alpha = g_{i_1} g_{i_2} \cdots g_{i_k} \in \mathbb{F}_n^+$ and $T_{g_0} := I_\mathcal{H}$, the identity on $\mathcal{H}$. According to the theory of row contractions (see [56] for the case $n = 1$, and [16], [7], [32], [33], [34], for $n \geq 2$) we have

$$\mathcal{C}_1 = [B(\mathcal{H})^n]_1^- := \left\{ (X_1, \dots, X_n) \in B(\mathcal{H})^n : \ \|X_1 X_1^* + \cdots + X_n X_n^*\|^{1/2} \leq 1 \right\}.$$

The results in [48] (see Section 4) can be seen as the unification of the theory of isometric dilations for row contractions [54], [56], [16], [7], [32], [33], [34] (which corresponds to the case $\rho = 1$) and Berger type dilations for n-tuples $(T_1, \dots, T_n)$ with the joint numerical radius $w(T_1, \dots, T_n) \leq 1$ (which corresponds to the case $\rho = 2$).

Following the classical case ([19], [59]), we defined the *joint operator radius* $\omega_\rho : B(\mathcal{H})^n \to [0, \infty)$, $\rho > 0$, by setting

$$\omega_\rho(T_1, \dots, T_n) := \inf \left\{ t > 0 : \ \left(\frac{1}{t} T_1, \dots, \frac{1}{t} T_n \right) \in \mathcal{C}_\rho \right\}$$

and $\omega_\infty(T_1, \dots, T_n) := \lim_{\rho \to \infty} \omega_\rho(T_1, \dots, T_n)$. In particular, $\omega_1(T_1, \dots, T_n)$ coincides with the norm of the row operator $[T_1 \cdots T_n]$, $\omega_2(T_1, \dots, T_n)$ coincides with the *joint numerical radius* $w(T_1, \dots, T_n)$, and $\omega_\infty(T_1, \dots, T_n)$ is equal to the (*algebraic*) *joint spectral radius* (see [7], [25])

$$r(T_1, \dots, T_n) := \lim_{k \to \infty} \left\| \sum_{|\alpha| = k} T_\alpha T_\alpha^* \right\|^{1/2k},$$

where the length of $\alpha \in \mathbb{F}_n^+$ is defined by $|\alpha| := 0$ if $\alpha = g_0$ and by $|\alpha| := k$ if $\alpha = g_{i_1} \cdots g_{i_k}$ and $i_1, \dots, i_k \in \{1, \dots, n\}$. In [48], we considered basic properties of the joint operator radius ω_ρ and we extended to the (noncommutative and commutative) multivariable setting several classical results obtained by Sz.-Nagy and Foiaş, Halmos, Berger and Stampfli, Holbrook, Paulsen, Badea and Cassier, and others (see [2], [3], [4], [5], [17], [18], [19], [20], [21], [29], [30], [55], and [59]).

In [49], we introduced a hyperbolic metric δ on the open noncommutative ball $[B(\mathcal{H})^n]_1$, which turned out to be a noncommutative extension of the Poincaré-Bergman ([6]) metric on the open unit ball $\mathbb{B}_n := \{z \in \mathbb{C}^n : \ \|z\|_2 < 1\}$. We proved that δ is invariant under the action of the group $Aut([B(\mathcal{H})^n]_1)$ of all free holomorphic automorphisms of $[B(\mathcal{H})^n]_1$, and showed that the δ-topology and the usual operator norm topology coincide on $[B(\mathcal{H})^n]_1$. Moreover, we proved that $[B(\mathcal{H})^n]_1$ is a complete metric space with respect to the hyperbolic metric and obtained an explicit formula for δ in terms of the reconstruction operator. A Schwarz-Pick lemma for bounded free holomorphic functions on $[B(\mathcal{H})^n]_1$, with respect to the hyperbolic metric, was also obtained. In [46], we continued to study the noncommutative hyperbolic geometry on the unit ball of $B(\mathcal{H})^n$, its connections with multivariable dilation theory, and its implications to noncommutative function theory. The results from [49] and [46] make connections between noncommutative function theory (see [41], [44], [50], [47]) and classical results in hyperbolic complex analysis (see [22], [23], [24], [52], [58]).

The present paper is an attempt to extend the results [49] concerning the noncommutative hyperbolic geometry of the unit ball $[B(\mathcal{H})^n]_1$ to the more general setting of [48]. We study the noncommutative balls

$$[\mathcal{C}_\rho]_{<1} = \{(X_1, \dots, X_n) \in B(\mathcal{H})^n : \ \omega_\rho(X_1, \dots, X_n) < 1\}, \qquad \rho \in (0, \infty],$$

and the Harnach parts of $\mathcal{C}_\rho$, $\rho > 0$, as metric spaces with respect to a hyperbolic (resp. Carathéodory) type metric that will be introduced. We provide mapping theorems for free holomorphic functions on these noncommutative balls, extending classical results from complex analysis and hyperbolic geometry.

In Section 1, we consider some preliminaries on free holomorphic (resp. pluriharmonic) functions on the open unit ball $[B(\mathcal{H})^n]_1$, and present several characterizations for the n-tuples of operators of class $\mathcal{C}_\rho$, $\rho \in (0, \infty)$. We introduce a free pluriharmonic functional calculus for the class $\mathcal{C}_\rho$ and show that a von Neumann type inequality characterizes this class. In particular, we prove that an n-tuple of operators $(T_1, \dots, T_n) \in B(\mathcal{H})^n$ is of class $\mathcal{C}_\rho$ if and only if

$$\|p(T_1, \dots, T_n)\| \leq \|\rho p(S_1, \dots, S_n) + (1 - \rho)p(0)\|$$

for any noncommutative polynomial with matrix-valued coefficients $p \in \mathbb{C}[Z_1, \dots, Z_n] \otimes M_m$, $m \in \mathbb{N}$, where $S_1, \dots, S_n$ are the left creation operators on the full Fock space with n generators.

In Section 2, we introduce a preorder relation $\overset{H}{\prec}$ on the class $\mathcal{C}_\rho$. If $A := (A_1, \dots, A_n)$ and $B := (B_1, \dots, B_n)$ are in the class $\mathcal{C}_\rho \subset B(\mathcal{H})^n$, we say that A is Harnack dominated by B (denote $A \overset{H}{\prec} B$) if there exists $c > 0$ such that

$$\Re p(A_1, \dots, A_n) + (\rho - 1)\Re p(0) \leq c^2 \left[\Re p(B_1, \dots, B_n) + (\rho - 1)\Re p(0)\right]$$

for any noncommutative polynomial with matrix-valued coefficients $p \in \mathbb{C}[X_1, \dots, X_n] \otimes M_m$, $m \in \mathbb{N}$, such that $\Re p(X) := \frac{1}{2}[p(X)^* + p(X)] \geq 0$ for any $X \in [B(\mathcal{K})^n]_1$, where $\mathcal{K}$ is an infinite dimensional Hilbert space. When we want to emphasize the constant c, we write $A \underset{c}{\overset{H}{\prec}} B$. We provide several characterizations for the Harnack domination on the noncommutative ball $\mathcal{C}_\rho$ (see Theorem 2.2), and determine the set of all elements in $\mathcal{C}_\rho$ which are Harnack dominated by 0. The results of this section will play a major role in the next sections.

The relation $\overset{H}{\prec}$ induces an equivalence relation $\overset{H}{\sim}$ on the class $\mathcal{C}_\rho$. More precisely, two n-tuples A and B are Harnack equivalent (and denote $A \overset{H}{\sim} B$) if and only if there exists $c > 1$ such that $A \underset{c}{\overset{H}{\prec}} B$ and $B \underset{c}{\overset{H}{\prec}} A$ (in this case we denote $A \underset{c}{\overset{H}{\sim}} B$). The equivalence classes with respect to $\overset{H}{\sim}$ are called Harnack parts of $\mathcal{C}_\rho$. In Section 3, we provide a Harnack type double inequality for positive free pluriharmonic functions on the noncommutative ball $\mathcal{C}_\rho$ and use it to prove that the Harnack part of $\mathcal{C}_\rho$ which contains 0 coincides with the open noncommutative ball

$$[\mathcal{C}_\rho]_{<1} := \{(X_1, \ldots, X_n) \in B(\mathcal{H})^n : \ \omega_\rho(X_1, \ldots, X_n) < 1\}.$$

We introduce a hyperbolic metric $\delta_\rho : \Delta \times \Delta \to \mathbb{R}^+$ on any Harnack part Δ of $\mathcal{C}_\rho$, by setting

$$\delta_\rho(A, B) := \ln \inf \left\{ c > 1 : \ A \underset{c}{\overset{H}{\sim}} B \right\}, \qquad A, B \in \Delta.$$

A concrete formula for the hyperbolic distance on any Harnack part of $\mathcal{C}_\rho$ is obtained. When $\Delta = [\mathcal{C}_\rho]_{<1}$, we prove that

$$\delta_\rho(A, B) = \ln \max \left\{ \left\| C_{\rho,A} C_{\rho,B}^{-1} \right\|, \left\| C_{\rho,B} C_{\rho,A}^{-1} \right\| \right\}, \qquad A, B \in [\mathcal{C}_\rho]_{<1},$$

where

$$\begin{aligned} C_{\rho,X} &:= \Delta_{\rho,X} (I - R_X)^{-1}, \\ \Delta_{\rho,X} &:= \left[\rho I + (1-\rho)(R_X^* + R_X) + (\rho - 2) R_X^* R_X\right]^{1/2}, \end{aligned}$$

and $R_X := X_1^* \otimes R_1 + \cdots + X_n^* \otimes R_n$ is the *reconstruction operator* associated with the right creation operators $R_1, \ldots, R_n$ on the full Fock space with n generators, and $X := (X_1, \ldots, X_n) \in [\mathcal{C}_\rho]_{<1}$. We recall that the reconstruction operator has played an important role in noncommutative multivariable operator theory. It appeared as a building block in the characteristic function associated to a row contraction (see [34], [45]) and also as a quantized variable (associated with the n-tuple X) in the noncommutative Cauchy, Poisson, and Berezin transform, respectively (see [41], [44], [47], [48]).

In Section 4, we study the stability of the ball $\mathcal{C}_\rho$ under contractive free holomorphic functions and provide mapping theorems, von Neumann inequalities, and Schwarz type lemmas, with respect to the hyperbolic metric δ_ρ and the operator radius ω_ρ, $\rho \in (0, \infty]$.

Let $f := (f_1, \ldots, f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$ (see [36], [40]). If an n-tuple of operators $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ is of class $\mathcal{C}_\rho$, $\rho > 0$, then we prove that, under the free pluriharmonic functional calculus, the m-tuple $f(T_1, \ldots, T_n) \in B(\mathcal{H})^m$ is of class $\mathcal{C}_{\rho_f}$, where $\rho_f > 0$ is given in terms of ρ and $f(0)$.

One of the main results of this section is the following *spectral von Neumann inequality* for n-tuples of operators. If $f := (f_1, \ldots, f_m)$ satisfies the conditions above and $(T_1,\ldots,T_n)\in B(\mathcal{H})^n$ has the joint spectral radius $r(T_1,\ldots,T_n)<1$, then $r(f(T_1,\ldots,T_n))<1$.

If, in addition, $f(0)=0$ and $\delta_\rho:\Delta\times\Delta\to[0,\infty)$ is the hyperbolic metric on a Harnack part Δ of $\mathcal{C}_\rho$, then we prove that

$$\delta_\rho(f(A),f(B))\leq\delta_\rho(A,B),\qquad A,B\in\Delta.$$

In particular, this holds when Δ is the open ball $[\mathcal{C}_\rho]_{<1}$. Moreover, in this setting, we show that

$$\omega_\rho(f(T_1,\ldots,T_n))<1,\qquad (T_1,\ldots,T_n)\in[\mathcal{C}_\rho]_{<1},$$

for any $\rho>0$. The general case when $f(0)\neq 0$ is also discussed.

In Section 5, we introduce a Carathéodory type metric on the set of all n-tuples of operators with joint spectral radius strictly less then 1, i.e.,

$$[\mathcal{C}_\infty]_{<1}:=\{(X_1,\ldots,X_n)\in B(\mathcal{H})^n:\ r(X_1,\ldots,X_n)<1\},$$

by setting

$$d_K(A,B)=\sup_p\|\Re p(A)-\Re p(B)\|,$$

where the supremum is taken over all noncommutative polynomials with matrix-valued coefficients $p\in\mathbb{C}[X_1,\ldots,X_n]\otimes M_m$, $m\in\mathbb{N}$, with $\Re p(0)=I$ and $\Re p(X)\geq 0$ for all $X\in[B(\mathcal{K})^n]_1$.

We obtain a concrete formula for d_K in terms of the free pluriharmonic kernel on the open unit ball $[\mathcal{C}_\infty]_{<1}$. More precisely, we show that

$$d_K(A,B)=\|P(A,R)-P(B,R)\|,\qquad A,B\in[\mathcal{C}_\infty]_{<1},$$

where

$$P(X,R):=\sum_{k=1}^\infty\sum_{|\alpha|=k}X_\alpha\otimes R_{\tilde\alpha}^*+\rho I\otimes I+\sum_{k=1}^\infty\sum_{|\alpha|=k}X_\alpha^*\otimes R_{\tilde\alpha},\qquad X\in[\mathcal{C}_\infty]_{<1},$$

and $\tilde\alpha$ is the reverse of $\alpha\in\mathbb{F}_n^+$. This is used to prove that the metric d_K is complete on $[\mathcal{C}_\infty]_{<1}$ and its topology coincides with the operator norm topology. We also prove that if $f:=(f_1,\ldots,f_m)$ is a contractive free holomorphic function with $\|f(0)\|<1$ such that the boundary functions $\widetilde{f}_1,\ldots,\widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$, then

$$d_K(f(A),f(B))\leq\frac{1+\|f(0)\|}{1-\|f(0)\|}d_K(A,B),\qquad A,B\in[\mathcal{C}_\infty]_{<1}.$$

As a consequence, we deduce that the map

$$[\mathcal{C}_\infty]_{<1}\ni(X_1,\ldots,X_n)\mapsto f(X_1,\ldots,X_n)\in[\mathcal{C}_\infty]_{<1}$$

is continuous in the operator norm topology.

In Section 6, we compare the hyperbolic metric δ_ρ with the Carathéodory metric d_K, and the operator metric, respectively, on Harnack parts of the unit ball $\mathcal{C}_\rho$, $\rho>0$. In particular, we prove that the hyperbolic metric δ_ρ is complete on the open unit unit ball $[\mathcal{C}_\rho]_{<1}$, while the other two metrics, mentioned above, are not complete. On the other hand, we show the δ_ρ-topology, the d_K-topology, and the operator norm topology coincide on $[\mathcal{C}_\rho]_{<1}$.

In Section 7, we consider the single variable case ($n = 1$) and show that our Harnack domination for ρ-contractions is equivalent to the one introduced and studied by G. Cassier and N. Suciu in [9] and [10]. Consequently, we recover some of their results and, moreover, we obtain some results which seem to be new even in the single variable case.

Finally, we want to acknowledge that we were influenced in writing this paper by the work of C. Foiaş ([15]), I. Suciu ([53]), and G. Cassier and N. Suciu ([9], [10]) concerning the Harnack domination and the hyperbolic distance between two ρ-contractions. It will be interesting to see to which extent the results of this paper, concerning the hyperbolic geometry on noncommutative balls, can be extended to the Hardy algebras of Muhly and Solel (see [26], [27], [28]).

1. The noncommutative ball $\mathcal{C}_\rho$ and a free pluriharmonic functional calculus

In this section, we consider some preliminaries on free holomorphic (resp. pluriharmonic) functions on the unit ball $[B(\mathcal{H})^n]_1$, and several characterizations for the n-tuples of operators of class $\mathcal{C}_\rho$. We introduce a free pluriharmonic functional calculus for the class $\mathcal{C}_\rho$ and show that a von Neumann type inequality characterizes the class $\mathcal{C}_\rho$.

Let H_n be an n-dimensional complex Hilbert space with orthonormal basis e_1, $e_2, \ldots, e_n$, where $n = 1, 2, \ldots$, or $n = \infty$. The full Fock space of H_n is defined by

$$F^2(H_n) := \mathbb{C}1 \oplus \bigoplus_{k\geq 1} H_n^{\otimes k},$$

where $H_n^{\otimes k}$ is the (Hilbert) tensor product of k copies of H_n. We define the left (resp. right) creation operators S_i (resp. R_i), $i = 1, \ldots, n$, acting on the full Fock space $F^2(H_n)$ by setting

$$S_i\varphi := e_i \otimes \varphi, \qquad \varphi \in F^2(H_n),$$

(resp. $R_i\varphi := \varphi \otimes e_i, \quad \varphi \in F^2(H_n)$). We recall that the noncommutative disc algebra $\mathcal{A}_n$ (resp. $\mathcal{R}_n$) is the norm closed algebra generated by the left (resp. right) creation operators and the identity. The noncommutative analytic Toeplitz algebra F_n^∞ (resp. $\mathcal{R}_n^\infty$) is the weakly closed version of $\mathcal{A}_n$ (resp. $\mathcal{R}_n$). These algebras were introduced in [36] in connection with a von Neumann type inequality [57], as noncommutative analogues of the disc algebra $A(\mathbb{D})$ and the Hardy space $H^\infty(\mathbb{D})$. For more information on theses noncommutative algebras we refer the reader to [35], [37], [38], [40], [12], and the references therein.

Let $\mathcal{H}$ be a Hilbert space and let $B(\mathcal{H})$ be the algebra of all bounded linear operators on $\mathcal{H}$. We identify $M_m(B(\mathcal{H}))$, the set of $m\times m$ matrices with entries from $B(\mathcal{H})$, with $B(\mathcal{H}^{(m)})$, where $\mathcal{H}^{(m)}$ is the direct sum of m copies of $\mathcal{H}$. If $\mathcal{X}$ is an operator space, i.e., a closed subspace of $B(\mathcal{H})$, we consider $M_m(\mathcal{X})$ as a subspace of $M_m(B(\mathcal{H}))$ with the induced norm. Let $\mathcal{X}, \mathcal{Y}$ be operator spaces and $u : \mathcal{X} \to \mathcal{Y}$ be a linear map. Define the map $u_m : M_m(\mathcal{X}) \to M_m(\mathcal{Y})$ by

$$u_m([x_{ij}]) := [u(x_{ij})].$$

We say that u is completely bounded if

$$\|u\|_{cb} := \sup_{m\geq 1} \|u_m\| < \infty.$$

If $\|u\|_{cb} \leq 1$ (resp. u_m is an isometry for any $m \geq 1$) then u is completely contractive (resp. isometric), and if u_m is positive for all m, then u is called completely positive. For basic results concerning completely bounded maps and operator spaces we refer to [29], [31], and [13].

A few more notations and definitions are necessary. If $\omega, \gamma \in \mathbb{F}_n^+$, we say that $\omega >_l \gamma$ if there is $\sigma \in \mathbb{F}_n^+ \backslash \{g_0\}$ such that $\omega = \gamma\sigma$ and set $\omega \backslash_l \gamma := \sigma$. We denote by $\tilde{\alpha}$ the reverse of $\alpha \in \mathbb{F}_n^+$, i.e., $\tilde{\alpha} = g_{i_k} \cdots g_{i_1}$ if $\alpha = g_{i_1} \cdots g_{i_k} \in \mathbb{F}_n^+$. An operator-valued positive semidefinite kernel on the free semigroup $\mathbb{F}_n^+$ is a map $K : \mathbb{F}_n^+ \times \mathbb{F}_n^+ \to B(\mathcal{H})$ with the property that for each $k \in \mathbb{N}$, for each choice of vectors $h_1, \ldots, h_k$ in $\mathcal{H}$, and $\sigma_1, \ldots, \sigma_k$ in $\mathbb{F}_n^+$, the inequality

$$\sum_{i,j=1}^{k} \langle K(\sigma_i, \sigma_j) h_j, h_i \rangle \geq 0$$

holds. Such a kernel is called multi-Toeplitz if it has the following properties: $K(\alpha, \alpha) = I_{\mathcal{H}}$ for any $\alpha \in \mathbb{F}_n^+$, and

$$K(\sigma, \omega) = \begin{cases} K(g_0, \omega \backslash_l \sigma) & \text{if } \omega >_l \sigma \\ K(\sigma \backslash_l \omega, g_0) & \text{if } \sigma >_l \omega \\ 0 & \text{otherwise.} \end{cases}$$

An n-tuple of operators $(T_1, \ldots, T_n)$, $T_i \in B(\mathcal{H})$, belongs to the class $\mathcal{C}_\rho$, $\rho > 0$, if there exist a Hilbert space $\mathcal{K} \supseteq \mathcal{H}$ and isometries $V_i \in B(\mathcal{K})$, $i = 1, \ldots, n$, with orthogonal ranges, such that

$$T_\alpha = \rho P_{\mathcal{H}} V_\alpha|_{\mathcal{H}}, \qquad \alpha \in \mathbb{F}_n^+ \backslash \{g_0\},$$

where $P_{\mathcal{H}}$ is the orthogonal projection of $\mathcal{K}$ onto $\mathcal{H}$. If $\mathcal{K} = \mathcal{K}_T := \bigvee_{\alpha \in \mathbb{F}_n^+} V_\alpha \mathcal{H}$, then the n-tuple $(V_1, \ldots, V_n)$ is the minimal isometric dilation of $(T_1, \ldots, T_n)$, which is unique up to an isomorphism. Note that if $(T_1, \ldots, T_n) \in \mathcal{C}_\rho$, then the joint spectral radius $r(T_1, \ldots, T_n) \leq 1$, where

$$r(T_1, \ldots, T_n) := \lim_{k\to\infty} \left\| \sum_{|\alpha|=k} T_\alpha T_\alpha^* \right\|^{1/2k}.$$

We recall (see Corollary 1.36 from [48]) that $\bigcup_{\rho>0} \mathcal{C}_\rho$ is dense (in the operator norm topology) in the set of all n-tuples of operators with joint spectral radius $r(T_1, \ldots, T_n) \leq 1$. Moreover, any n-tuple of operators with $r(T_1, \ldots, T_n) < 1$ is of class $\mathcal{C}_\rho$ for some $\rho > 0$. We should add that (see Theorem 5.9 from [43])

$(T_1,\dots,T_n)\in B(\mathcal{H})^n$ has the joint spectral radius $r(T_1,\dots,T_n)<1$ if and only if it is uniformly stable, i.e., $\|\sum_{|\alpha|=k}T_\alpha T_\alpha^*\|\to 0$, as $k\to\infty$.

Since the joint spectral radius of n-tuples of operators plays an important role in the present paper, we recall (see [7], [25]) some of its properties. The joint right spectrum $\sigma_r(T_1,\dots,T_n)$ of an n-tuple $(T_1,\dots,T_n)$ of operators in $B(\mathcal{H})$ is the set of all n-tuples $(\lambda_1,\dots,\lambda_n)$ of complex numbers such that the right ideal of $B(\mathcal{H})$ generated by the operators $\lambda_1 I-T_1,\dots,\lambda_n I-T_n$ does not contain the identity operator. We know that $\sigma_r(T_1,\dots,T_n)$ is included in the closed ball of $\mathbb{C}^n$ of radius $r(T_1,\dots,T_n)$.

If we assume that $T_1,\dots,T_n\in B(\mathcal{H})$ are mutually commuting operators and $\mathcal{B}$ is a closed subalgebra of $B(\mathcal{H})$ containing $T_1,\dots,T_n$, and the identity, then the Harte spectrum $\sigma(T_1,\dots,T_n)$ is the set of all $(\lambda_1,\dots,\lambda_n)\in\mathbb{C}^n$ such that

$$(\lambda_1 I-T_1)X_1+\cdots+(\lambda_n I-T_n)X_n\neq I$$

for all $X_1,\dots,X_n\in\mathcal{B}$. In this case, we have

$$r(T_1,\dots,T_n)=\max\{\|(\lambda_1,\dots,\lambda_n)\|_2:\ (\lambda_1,\dots,\lambda_n)\in\sigma(T_1,\dots,T_n)\}.$$

According to [25], the latter formula remains true if the Harte spectrum is replaced by the Taylor's spectrum for commuting operators.

According to Theorem 4.1 from [39] and Theorems 1.34 and 1.39 from [48], we have the following characterizations for the n-tuples of operators of class $\mathcal{C}_\rho$. We denote by $\mathbb{C}[Z_1,\dots,Z_n]$ the set of all noncommutative polynomials in n noncommuting indeterminates.

Theorem 1.1. *Let $T_1,\dots,T_n\in B(\mathcal{H})$ and let $\mathcal{S}\subset C^*(S_1,\dots,S_n)$ be the operator system defined by*

$$\mathcal{S}:=\{p(S_1,\dots,S_n)+q(S_1,\dots,S_n)^*:\ p,q\in\mathbb{C}[Z_1,\dots,Z_n]\}.$$

Then the following statements are equivalent:

(i) $(T_1,\dots,T_n)\in\mathcal{C}_\rho$.

(ii) *The map $\Psi:\mathcal{S}\to B(\mathcal{H})$ defined by*

$$\Psi\left(p(S_1,\dots,S_n)+q(S_1,\dots,S_n)^*\right):=p(T_1,\dots,T_n)+q(T_1,\dots,T_n)^* \\ +(\rho-1)(p(0)+\overline{q(0)})I$$

is completely positive.

(iii) *The joint spectral radius $r(T_1\dots,T_n)\leq 1$ and the ρ-pluriharmonic kernel defined by*

$$P_\rho(rT,R):=\sum_{k=1}^\infty\sum_{|\alpha|=k}r^{|\alpha|}T_\alpha\otimes R_{\tilde\alpha}^*+\rho I\otimes I+\sum_{k=1}^\infty\sum_{|\alpha|=k}r^{|\alpha|}T_\alpha^*\otimes R_{\tilde\alpha}$$

is positive for any $0<r<1$, where the convergence is in the operator norm topology.

(iv) *The spectral radius* $r(T_1,\dots,T_n)\le 1$ *and*

$$\rho I\otimes I+(1-\rho)r\sum_{i=1}^n (T_i\otimes R_i^*+T_i^*\otimes R_i)+(\rho-2)r^2\left(\sum_{i=1}^n T_iT_i^*\otimes I\right)\ge 0$$

for any $0<r<1$.

(v) *The multi-Toeplitz kernel* $K_{\rho,T}:\mathbb{F}_n^+\times\mathbb{F}_n^+\to B(\mathcal{H})$ *defined by*

$$K_{\rho,T}(\alpha,\beta):=\begin{cases}\frac{1}{\rho}T_{\beta\backslash_l\alpha} & \text{if } \beta>_l\alpha\\ I & \text{if } \alpha=\beta\\ \frac{1}{\rho}(T_{\alpha\backslash_l\beta})^* & \text{if } \alpha>_l\beta\\ 0 & \text{otherwise}\end{cases}$$

is positive semidefinite.

Consider $1\le m<n$ and let $(R_1',\dots,R_m')$ and $(R_1,\dots,R_n)$ be the right creation operators on $F^2(H_m)$ and $F^2(H_n)$, respectively. According to the Wold type decomposition for isometries with orthogonal ranges [33], the m-tuple $(R_1,\dots,R_m)$ is unitarily equivalent to $(R_1'\otimes I_{\mathcal{E}},\dots,R_m'\otimes I_{\mathcal{E}})$, where $\mathcal{E}$ is equal to $F^2(H_n)\ominus F^2(H_m)$. Consequently, using Theorem 1.1, one can easily deduce the following result.

Corollary 1.2. *Let* $\rho>0$, $1\le m<n$, *and consider an* m*-tuple* $(T_1,\dots,T_m)\in B(\mathcal{H})^m$ *and its extension* $(T_1,\dots,T_m,0,\dots,0)\in B(\mathcal{H})^n$. *Then the following statements hold:*

(i) $(T_1,\dots,T_m)\in\mathcal{C}_\rho$ *if and only if* $(T_1,\dots,T_m,0,\dots,0)\in\mathcal{C}_\rho$;
(ii) $\omega_\rho(T_1,\dots,T_m)=\omega_\rho(T_1,\dots,T_m,0,\dots,0))$;
(iii) $r(T_1,\dots,T_m)=r(T_1,\dots,T_m,0,\dots,0)$.

Throughout this paper, we assume that $\mathcal{E}$ is a separable Hilbert space. We recall [44] that a mapping $F:[B(\mathcal{H})^n]_1\to B(\mathcal{H})\bar{\otimes}_{min}B(\mathcal{E})$ is called *free holomorphic function* on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$ if there exist $A_{(\alpha)}\in B(\mathcal{E})$, $\alpha\in\mathbb{F}_n^+$, such that $\limsup_{k\to\infty}\left\|\sum_{|\alpha|=k}A_{(\alpha)}^*A_{(\alpha)}\right\|^{1/2k}\le 1$ and

$$F(X_1,\dots,X_n)=\sum_{k=0}^\infty\sum_{|\alpha|=k}X_\alpha\otimes A_{(\alpha)},$$

where the series converges in the operator norm topology for any $(X_1,\dots,X_n)$ in the open unit ball $[B(\mathcal{H})^n]_1:=\{(X_1,\dots,X_n):\ \|X_1X_1^*+\cdots+X_nX_n\|<1\}$. The set of all free holomorphic functions on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$ is denoted by $H_{\mathbf{ball}}(B(\mathcal{E}))$. Let $H^\infty_{\mathbf{ball}}(B(\mathcal{E}))$ denote the set of all elements F in $H_{\mathbf{ball}}(B(\mathcal{E}))$ such that

$$\|F\|_\infty:=\sup\|F(X_1,\dots,X_n)\|<\infty,$$

where the supremum is taken over all n-tuples of operators $(X_1,\dots,X_n)\in[B(\mathcal{H})^n]_1$ and any Hilbert space $\mathcal{H}$. According to [44] and [47], $H^\infty_{\mathbf{ball}}(B(\mathcal{E}))$

can be identified to the operator algebra $F_n^\infty \bar{\otimes} B(\mathcal{E})$ (the weakly closed algebra generated by the spatial tensor product), via the noncommutative Poisson transform. Due to the fact that a free holomorphic function is uniquely determined by its representation on an infinite dimensional Hilbert space, we identify, throughout this paper, a free holomorphic function with its representation on a separable infinite dimensional Hilbert space.

We say that a map $u : [B(\mathcal{H})^n]_1 \to B(\mathcal{H})\bar{\otimes}_{min} B(\mathcal{E})$ is a self-adjoint *free pluriharmonic function* on $[B(\mathcal{H})^n]_1$ if $u = \Re f := \frac{1}{2}(f^* + f)$ for some free holomorphic function f. A free pluriharmonic function on $[B(\mathcal{H})^n]_1$ has the form $H := H_1 + iH_2$, where H_1, H_2 are self-adjoint free pluriharmonic functions on $[B(\mathcal{H})^n]_1$. We recall [47] that if

$$f(Z_1, \ldots, Z_n) = \sum_{k=1}^{\infty} \sum_{|\alpha|=k} Z_\alpha^* \otimes B_{(\alpha)} + I \otimes A_{(0)} + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} Z_\alpha \otimes A_{(\alpha)}$$

is a free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$ and $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ is any n-tuple of operators with joint spectral radius $r(T_1, \ldots, T_n) < 1$, then $f(T_1, \ldots, T_n)$ is a bounded linear operator, where the corresponding series converge in norm. Moreover $\lim_{r\to 1} f(rT_1, \ldots, rT_n) = f(T_1, \ldots T_n)$ in the operator norm topology. We refer to [47] for more results on free pluriharmonic functions.

We denote by $Har^c_{\mathbf{ball}}(B(\mathcal{E}))$ the set of all free pluriharmonic functions on $[B(\mathcal{H})^n]_1$ with operator-valued coefficients in $B(\mathcal{E})$, which have continuous extensions (in the operator norm topology) to the closed ball $[B(\mathcal{H})^n]_1^-$. We assume that $\mathcal{H}$ is an infinite dimensional Hilbert space. According to Theorem 4.1 from [47], we can identify $Har^c_{\mathbf{ball}}(B(\mathcal{E}))$ with the operator space $\overline{\mathcal{A}_n(\mathcal{E})^* + \mathcal{A}_n(\mathcal{E})}^{\|\cdot\|}$, where $\mathcal{A}_n(\mathcal{E}) := \mathcal{A}_n \bar{\otimes}_{min} B(\mathcal{E})$ and $\mathcal{A}_n$ is the noncommutative disc algebra. More precisely, if $u : [B(\mathcal{H})^n]_1 \to B(\mathcal{H})\bar{\otimes}_{min} B(\mathcal{E})$, then the following statements are equivalent:

(a) u is a free pluriharmonic function on $[B(\mathcal{H})^n]_1$ which has a continuous extension (in the operator norm topology) to the closed ball $[B(\mathcal{H})^n]_1^-$;

(b) there exists $f \in \overline{\mathcal{A}_n(\mathcal{E})^* + \mathcal{A}_n(\mathcal{E})}^{\|\cdot\|}$ such that $u(X) = (P_X \otimes \mathrm{id})(f)$ for $X \in [B(\mathcal{H})^n]_1$, where P_X is the noncommutative Poisson transform at X;

(c) u is a free pluriharmonic function on $[B(\mathcal{H})^n]_1$ such that $u(rS_1, \ldots, rS_n)$ converges in the operator norm topology, as $r \to 1$.

In this case, we have $f = \lim_{r\to 1} u(rS_1, \ldots, rS_n)$, where the convergence is in the operator norm topology. Moreover, the map $\Phi : Har^c_{\mathbf{ball}}(B(\mathcal{E})) \to \overline{\mathcal{A}_n(\mathcal{E})^* + \mathcal{A}_n(\mathcal{E})}^{\|\cdot\|}$ defined by $\Phi(u) := f$ is a completely isometric isomorphism of operator spaces. We call f the *model boundary function* of u.

Now, we introduce a free pluriharmonic functional calculus for the class $\mathcal{C}_\rho$.

Theorem 1.3. *Let $T := (T_1, \ldots, T_n) \in B(\mathcal{H})^n$ be of class $\mathcal{C}_\rho$, and let $u \in Har^c_{\mathbf{ball}}(B(\mathcal{E}))$ have the standard representation*

$$u(X_1, \ldots, X_n) = \sum_{k=1}^{\infty} \sum_{|\alpha|=k} X_\alpha^* \otimes B_{(\alpha)} + I \otimes A_{(0)} + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} X_\alpha \otimes A_{(\alpha)}$$

on $[B(\mathcal{H})^n]_1$, for some $A_{(\alpha)}, B_{(\alpha)} \in B(\mathcal{E})$, where the series converge in the operator norm topology. Then

$$u(T_1, \ldots, T_n) := \lim_{r \to 1} u(rT_1, \ldots, rT_1)$$

exists in the operator norm and

$$\|u(T_1, \ldots, T_n)\| \leq \|\rho u + (1-\rho)u(0)\|_\infty.$$

Proof. Since $T := (T_1, \ldots, T_n) \in B(\mathcal{H})^n$ is an n-tuple of class $\mathcal{C}_\rho$, there is a minimal isometric dilation $V := (V_1, \ldots, V_n)$ of T on a Hilbert space $\mathcal{K}_T \supseteq \mathcal{H}$, satisfying the following properties: $V_i^* V_j = \delta_{ij} I$ for $i, j = 1, \ldots, n$, and $T_\alpha = \rho P_{\mathcal{H}} V_\alpha|_{\mathcal{H}}$ for any $\alpha \in \mathbb{F}_n^+ \backslash \{g_0\}$, and $\mathcal{K}_T = \bigvee_{\alpha \in \mathbb{F}_n^+} V_\alpha \mathcal{H}$. Taking into account that $u \in Har^c_{\mathbf{ball}}(B(\mathcal{E}))$, we have

$$u(rV_1, \ldots, rV_n) = \sum_{k=1}^{\infty} \sum_{|\alpha|=k} r^{|\alpha|} V_\alpha^* \otimes B_{(\alpha)} + I \otimes A_{(0)} + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} r^{|\alpha|} V_\alpha \otimes A_{(\alpha)},$$

where the convergence is in the operator norm. Hence, and due to the fact that

$$\sum_{|\alpha|=k} r^{|\alpha|} T_\alpha^* \otimes B_{(\alpha)} = \rho(P_{\mathcal{H}} \otimes I) \left(\sum_{|\alpha|=k} r^{|\alpha|} V_\alpha^* \otimes B_{(\alpha)} \right) |_{\mathcal{H} \otimes \mathcal{E}}, \qquad k = 1, 2, \ldots,$$

we deduce that

$$\begin{aligned} u(rT_1, \ldots, rT_n) &:= \sum_{k=1}^{\infty} \sum_{|\alpha|=k} r^{|\alpha|} T_\alpha^* \otimes B_{(\alpha)} + I \otimes A_{(0)} + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} r^{|\alpha|} T_\alpha \otimes A_{(\alpha)} \\ &= \rho(P_{\mathcal{H}} \otimes I) u(rV_1, \ldots, rV_n)|_{\mathcal{H} \otimes \mathcal{E}} - (\rho - 1)u(0). \end{aligned}$$

exists in the operator norm topology. Now, taking into account that $\lim_{r \to 1} u(rV_1, \ldots, rV_1)$ exists in the operator norm, we deduce that $\lim_{r \to 1} u(rT_1, \ldots, rT_1)$ exists in the same topology. Consequently, we can define

$$u(T_1, \ldots, T_n) := \lim_{r \to 1} u(rT_1, \ldots, rT_1).$$

Using the considerations above, and the noncommutative von Neumann inequality, we obtain

$$\|u(T_1, \ldots, T_n)\| \leq \|\rho u + (1-\rho)u(0)\|_\infty \leq (\rho + |\rho - 1|)\|u\|_\infty$$

for any $(T_1, \ldots, T_n) \in \mathcal{C}_\rho$. ∎

We will refer to the map

$$Har^c_{\mathbf{ball}}(B(\mathcal{E})) \ni u \mapsto u(T_1,\dots,T_n) \in B(\mathcal{H})\bar{\otimes}_{min}B(\mathcal{E})$$

as the free pluriharmonic functional calculus for the class $\mathcal{C}_\rho$. Since there is a completely isometric isomorphism of operator spaces $\overline{\mathcal{A}_n(\mathcal{E})^* + \mathcal{A}_n(\mathcal{E})}^{\|\cdot\|} \ni f \mapsto u \in Har^c_{\mathbf{ball}}(B(\mathcal{E}))$, given by $u = (P_X \otimes \mathrm{id})(f)$ for $X \in [B(\mathcal{H})^n]_1$, we also use the notation $f(T_1,\dots,T_n)$ for $u(T_1,\dots,T_n)$.

Now, we show that the von Neumann type inequality of Theorem 1.3 characterizes the class $\mathcal{C}_\rho$. Denote

$$\mathcal{P}(S_1,\dots,S_n) := \{p(S_1,\dots,S_n) :\ p \in \mathbb{C}[Z_1,\dots,Z_n]\},$$

where $S_1,\dots,S_n$ are the left creation operators on the full Fock space $F^2(H_n)$.

Theorem 1.4. *Let $T := (T_1,\dots,T_n) \in B(\mathcal{H})^n$ be an n-tuple of operators. Then the following statements are equivalent:*

(i) *T is of class $\mathcal{C}_\rho$;*

(ii) *the von Neumann type inequality*

$$\|p(T_1,\dots,T_n)\| \le \|\rho p(S_1,\dots,S_n) + (1-\rho)p(0)\|$$

holds for any noncommutative polynomial $p \in \mathbb{C}[Z_1,\dots,Z_n] \otimes M_m$, $m \in \mathbb{N}$;

(iii) *the map $\Psi_T : \mathcal{A}_n \to B(\mathcal{H})$ defined by*

$$\Psi_T\left(q(S_1,\dots,S_n)\right) := \frac{1}{\rho}q(T_1,\dots,T_n) + \left(1 - \frac{1}{\rho}\right)q(0)I$$

for $q(S_1,\dots,S_n) \in \mathcal{P}(S_1,\dots,S_n)$ is completely contractive.

Proof. The implication $(i) \implies (ii)$ follows, in particular, from Theorem 1.3. To prove the implication $(ii) \implies (iii)$, note that setting $p := \frac{1}{\rho}q + \left(1 - \frac{1}{\rho}\right)q(0)I$, where $q \in \mathbb{C}[Z_1,\dots,Z_n] \otimes M_m$, $m \in \mathbb{N}$, we have

$$\begin{aligned}\|\Psi_T\left(q(S_1,\dots,S_n)\right)\| &= \|p(T_1,\dots,T_n)\|\\ &\le \|\rho p(S_1,\dots,S_n) + (1-\rho)p(0)\|\\ &= \|q(S_1,\dots,S_n)\|,\end{aligned}$$

which proves that Ψ_T is completely contractive on the set of all polynomials $\mathcal{P}(S_1,\dots,S_n)$ and, consequently, extends uniquely to a completely contractive map on the noncommutative disc algebra $\mathcal{A}_n$. It remains to prove that $(iii) \implies (i)$. Due to Arveson's extension theorem, item (iii) implies the existence of a unique completely positive extension $\widetilde{\Psi}_T : \mathcal{A}_n^* + \mathcal{A}_n \to B(\mathcal{H})$ of Ψ_T. Note that

$$\begin{aligned}&\widetilde{\Psi}_T(r(S_1,\dots,S_n) + q(S_1,\dots,S_n)^*) =\\ &\qquad = \frac{1}{\rho}(r(T_1,\dots,T_n) + q(T_1,\dots,T_n)^*) + \left(1 - \frac{1}{\rho}\right)(r(0) + \overline{q(0)})I\end{aligned}$$

for any polynomials $r(S_1,\dots,S_n)$ and $q(S_1,\dots,S_n)$ in $\mathcal{P}(S_1,\dots,S_n)$. Applying Theorem 1.1 (the equivalence $(i) \leftrightarrow (ii)$), we complete the proof. □

2. Harnack domination on noncommutative balls

We introduce a preorder relation $\overset{H}{\prec}$ on the noncommutative ball $\mathcal{C}_\rho$, $\rho\in(0,\infty)$, and provide several characterizations. We determine the elements of $\mathcal{C}_\rho$ which are Harnack dominated by 0. These results will play a crucial role in the next sections.

First, we consider some preliminaries on noncommutative Poisson transforms. Let $C^*(S_1,\dots,S_n)$ be the Cuntz-Toeplitz C^*-algebra generated by the left creation operators (see [11]). The noncommutative Poisson transform at the n-tuple $T:=(T_1,\dots,T_n)\in[B(\mathcal{H})^n]_1^-$ is the unital completely contractive linear map $P_T: C^*(S_1,\dots,S_n)\to B(\mathcal{H})$ defined by

$$P_T[f]:=\lim_{r\to 1}K_{rT}^*(I_\mathcal{H}\otimes f)K_{rT},\qquad f\in C^*(S_1,\dots,S_n),$$

where the limit exists in the operator norm topology of $B(\mathcal{H})$. Here, the noncommutative Poisson kernel $K_{rT}:\mathcal{H}\to\overline{\Delta_{rT}\mathcal{H}}\otimes F^2(H_n)$, $0<r\le 1$, is defined by

$$K_{rT}h:=\sum_{k=0}^\infty\sum_{|\alpha|=k}r^{|\alpha|}\Delta_{rT}T_\alpha^*h\otimes e_\alpha,\qquad h\in\mathcal{H},$$

where $\{e_\alpha\}_{\alpha\in\mathbb{F}_n^+}$ is the orthonormal basis for the full Fock space $F^2(H_n)$, defined by $e_\alpha:=e_{i_1}\otimes\cdots\otimes e_{i_k}$ if $\alpha=g_{i_1}\cdots g_{i_k}\in\mathbb{F}_n^+$ and $e_{g_0}:=1$, and $\Delta_{rT}:=(I_\mathcal{H}-r^2T_1T_1^*-\cdots-r^2T_nT_n^*)^{1/2}$. We recall that $P_T[S_\alpha S_\beta^*]=T_\alpha T_\beta^*$, $\alpha,\beta\in\mathbb{F}_n^+$. When $T:=(T_1,\dots,T_n)$ is a pure row contraction, i.e., SOT-$\lim\limits_{k\to\infty}\sum_{|\alpha|=k}T_\alpha T_\alpha^*=0$, then we have

$$P_T[f]=K_T^*(I_{\mathcal{D}_T}\otimes f)K_T,\qquad f\in C^*(S_1,\dots,S_n)\ \text{ or }\ f\in F_n^\infty,$$

where $\mathcal{D}_T:=\overline{\Delta_T\mathcal{H}}$. We refer to [41], [42], and [48] for more on noncommutative Poisson transforms on C^*-algebras generated by isometries.

A free pluriharmonic function u on $[B(\mathcal{K})^n]_1$ with operator valued coefficients is called positive, and denote $u\ge 0$, if $u(X_1,\dots,X_n)\ge 0$ for any $(X_1,\dots,X_n)\in[B(\mathcal{K})^n]_1$, where $\mathcal{K}$ is an infinite dimensional Hilbert space. We mention that it is enough to assume that the positivity condition holds for any finite dimensional Hilbert space $\mathcal{K}$. Indeed, for each $m\in\mathbb{N}$, consider $R^{(m)}:=(R_1^{(m)},\dots,R_n^{(m)})$, where $R_i^{(m)}$ is the compression of the right creation operator R_i to the subspace $\mathcal{P}_m:=\operatorname{span}\{e_\alpha:\ \alpha\in\mathbb{F}_n^+,|\alpha|\le m\}$ of $F^2(H_n)$. We recall from [47] the following result.

Lemma 2.1. *Let u be a free pluriharmonic function on $[B(\mathcal{K})^n]_1$ with operator-valued coefficients. Then $u(X_1,\dots,X_n)\ge 0$ for any $(X_1,\dots,X_n)\in[B(\mathcal{K})^n]_1$ if and only if $u(R_1^{(m)},\dots,R_n^{(m)})\ge 0$ for any $m\in\mathbb{N}$.*

Let $A:=(A_1,\dots,A_n)$ and $B:=(B_1,\dots,B_n)$ be n-tuples of operators in $\mathcal{C}_\rho\subset B(\mathcal{H})^n$. We say that A is Harnack dominated by B, and denote $A\overset{H}{\prec}B$, if there exists $c>0$ such that

$$\Re p(A_1,\dots,A_n)+(\rho-1)\Re p(0)\le c^2\left[\Re p(B_1,\dots,B_n)+(\rho-1)\Re p(0)\right]$$

for any noncommutative polynomial with matrix-valued coefficients $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, such that $\Re p \geq 0$. When we want to emphasize the constant c, we write $A \overset{H}{\underset{c}{\prec}} B$.

According to Theorem 1.3, we can associate with each n-tuple $T := (T_1, \ldots, T_n) \in \mathcal{C}_\rho$ the completely positive map $\varphi_T : \overline{\mathcal{A}_n^* + \mathcal{A}_n}^{\|\cdot\|} \to B(\mathcal{H})$ defined by

$$\varphi_T(g) := \frac{1}{\rho} g(T_1, \ldots, T_n) + \left(1 - \frac{1}{\rho}\right) g(0), \tag{2.1}$$

where $g(T_1, \ldots, T_n)$ is defined by the free pluriharmonic functional calculus for the class $\mathcal{C}_\rho$.

Now, we present several characterizations for the Harnack domination in $\mathcal{C}_\rho$.

Theorem 2.2. *Let $A := (A_1, \ldots, A_n) \in B(\mathcal{H})^n$ and $B := (B_1, \ldots, B_n) \in B(\mathcal{H})^n$ be in the class $\mathcal{C}_\rho$ and let $c > 0$. Then the following statements are equivalent:*

(i) $A \overset{H}{\underset{c}{\prec}} B$;

(ii) $P_\rho(rA, R) \leq c^2 P_\rho(rB, R)$ *for any* $r \in [0, 1)$, *where* $P_\rho(X, R)$ *is the multi-Toeplitz kernel associated with* $X \in \mathcal{C}_\rho$;

(iii) $u(rA_1, \ldots, rA_n) + (\rho - 1)u(0) \leq c^2 [u(rB_1, \ldots, rB_n) + (\rho - 1)u(0)]$ *for any positive free pluriharmonic function* u *on* $[B(\mathcal{H})^n]_1$ *with operator-valued coefficients and any* $r \in [0, 1)$;

(iv) $K_{\rho,A} \leq c^2 K_{\rho,B}$, *where* $K_{\rho,X}$ *is the multi-Toeplitz kernel associated with* $X \in \mathcal{C}_\rho$;

(v) $c^2 \varphi_B - \varphi_A$ *is a completely positive linear map on the operator space* $\overline{\mathcal{A}_n^* + \mathcal{A}_n}^{\|\cdot\|}$, *where* φ_A, φ_B *are the c.p. maps associated with* A *and* B, *respectively.*

(vi) *there is an operator* $L_{B,A} \in B(\mathcal{K}_B, \mathcal{K}_A)$ *with* $\|L_{B,A}\| \leq c$ *such that* $L_{B,A}|_\mathcal{H} = I_\mathcal{H}$ *and*

$$L_{B,A} W_i = V_i L_{B,A}, \qquad i = 1, \ldots, n,$$

where $(V_1, \ldots, V_n)$ *on* $\mathcal{K}_A \supset \mathcal{H}$ *and* $(W_1, \ldots, W_n)$ *on* $\mathcal{K}_A \supset \mathcal{H}$ *are the minimal isometric dilations of* A *and* B, *respectively.*

Proof. First we prove that $(i) \implies (ii)$. Since $R_\alpha^{(m)} = 0$ for any $\alpha \in \mathbb{F}_n^+$ with $|\alpha| \geq m + 1$, we have

$$P_\rho(rX, R^{(m)}) = \sum_{1 \leq |\alpha| \leq m} r^{|\alpha|} X_\alpha^* \otimes R_{\tilde{\alpha}}^{(m)} + \rho I \otimes I + \sum_{1 \leq |\alpha| \leq m} r^{|\alpha|} X_\alpha \otimes R_{\tilde{\alpha}}^{(m)*}.$$

Since $X \mapsto P_1(X, R)$ is a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$, with coefficients in $B(F^2(H_n))$, so is the map

$$X \mapsto P_1(rX, R^{(m)}) = (I \otimes P_{\mathcal{P}_m}) P_1(rX, R)|_{\mathcal{H} \otimes \mathcal{P}_m}$$

for any $r \in [0,1)$. If $A \overset{H}{\underset{c}{\prec}} B$, then we have

$$P_1(rA, R^{(m)}) + (\rho - 1)P_1(0, R^{(m)}) \leq c^2 \left[P_1(rB, R^{(m)}) + (\rho - 1)P_1(0, R^{(m)})\right]$$

for any $m = 1, 2, \ldots$. Using Lemma 2.1, we deduce that

$$P_1(rA, R) + (\rho - 1)I \leq c^2 \left[P_1(rB, R) + (\rho - 1)I\right]$$

for any $r \in [0,1)$. Since $P_\rho(rY, R) = P_1(rY, R) + (\rho - 1)I$ for any n-tuple $Y \in B(\mathcal{H})^n$ with spectral radius $r(Y) \leq 1$ and $r \in [0,1)$, we deduce item (ii).
To prove the implication $(ii) \implies (iii)$, assume that condition (ii) holds and let u be a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$ of the form

$$u(Z_1, \ldots, Z_n) = \sum_{k=1}^{\infty} \sum_{|\alpha|=k} Z_\alpha^* \otimes C_{(\alpha)}^* + I \otimes C_{(0)} + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} Z_\alpha \otimes C_{(\alpha)}.$$

It is well-known (see e.g. [29]) that if $\mathcal{S} \subseteq B(F^2(H_n))$ is an operator system and $\mu : \mathcal{S} \to B(\mathcal{K})$ is a completely bounded map, then there exists a completely bounded linear map

$$\widetilde{\mu} := \mu \otimes \mathrm{id} : \mathcal{S} \bar{\otimes}_{min} B(\mathcal{H}) \to B(\mathcal{K}) \bar{\otimes}_{min} B(\mathcal{H})$$

such that $\widetilde{\mu}(f \otimes Y) := \mu(f) \otimes Y$ for $f \in \mathcal{S}$ and $Y \in B(\mathcal{H})$. Moreover, $\|\widetilde{\mu}\|_{cb} = \|\mu\|_{cb}$ and, if μ is completely positive, then so is $\widetilde{\mu}$.
Using Corollary 5.5 from [47], we find a completely positive linear map $\nu : \mathcal{R}_n^* + \mathcal{R}_n \to B(\mathcal{E})$ such that $\nu(R_{\tilde{\alpha}}) = C_{(\alpha)}^*$ if $|\alpha| \geq 1$ and $\nu(I) = C_{(0)}$. Note that

$$\begin{aligned}
&(\mathrm{id} \otimes \nu)[c^2 P_\rho(rB, R) - P_\rho(rA, R)] \\
&\quad = (\mathrm{id} \otimes \nu) \left\{ \sum_{k=1}^{\infty} \sum_{|\alpha|=k} r^{|\alpha|}(c^2 B_\alpha - A_\alpha) \otimes R_{\tilde{\alpha}}^* + \rho(c^2 - 1) I \otimes I \right. \\
&\qquad \left. + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} (c^2 B_\alpha^* - A_\alpha^*) \otimes R_{\tilde{\alpha}} \right\} \\
&\quad = \left\{ \sum_{k=1}^{\infty} \sum_{|\alpha|=k} r^{|\alpha|}(c^2 B_\alpha - A_\alpha) \otimes C_{(\alpha)} + \rho(c^2 - 1) I \otimes C_{(0)} \right. \\
&\qquad \left. + \sum_{k=1}^{\infty} \sum_{|\alpha|=k} (c^2 B_\alpha^* - A_\alpha^*) \otimes C_{(\alpha)}^* \right\} \\
&\quad = c^2 \left[u(rB_1, \ldots, rB_n) + (\rho - 1)u(0)\right] \\
&\qquad - \left[u(rA_1, \ldots, rA_n) + (\rho - 1)u(0)\right].
\end{aligned}$$

Hence, and using the fact that $c^2 P_\rho(rB, R) - P_\rho(rA, R) \geq 0$, we deduce that

$$c^2 \left[u(rB_1, \ldots, rB_n) + (\rho - 1)u(0)\right] - \left[u(rA_1, \ldots, rA_n) + (\rho - 1)u(0)\right] \geq 0,$$

which proves (iii).

Now, we prove the implication $(iii) \implies (v)$. Let $g \in \left(\overline{\mathcal{A}_n^* + \mathcal{A}_n}^{\|\cdot\|}\right) \otimes_{min} M_m$ be positive. Then, according to Theorem 4.1 from [47], the map defined by

$$g(X) := (P_X \otimes \mathrm{id})[g], \qquad X \in [B(\mathcal{H})^n]_1,$$

is a positive free pluriharmonic function. Condition (iii) implies

$$g(rA_1, \ldots, rA_n) + (\rho - 1)g(0) \le c^2 \left[g(rB_1, \ldots, rB_n) + (\rho - 1)g(0)\right]$$

for any $r \in [0, 1)$. Hence, and using relation (2.1), we get $\rho\varphi_A(g_r) \le c^2\rho\varphi_B(g_r)$. Taking $r \to 1$, we deduce item (v).

To prove the implication $(v) \implies (i)$, let $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, be a noncommutative polynomial with matrix coefficients such that $\operatorname{Re} p \ge 0$. Since

$$\rho\varphi_Y(p) = p(Y_1, \ldots, Y_n) + (\rho - 1)p(0)$$

for any $Y := (Y_1, \ldots, Y_n) \in \mathcal{C}_\rho$, it is clear that (v) implies item (i).

We prove now that $(ii) \implies (iv)$. We recall that $e_\alpha := e_{i_1} \otimes \cdots \otimes e_{i_k}$ if $\alpha = g_{i_1} \cdots g_{i_k} \in \mathbb{F}_n^+$ and $e_{g_0} := 1$, and that $\{e_\alpha\}_{\alpha \in \mathbb{F}_n^+}$ is an orthonormal basis for the full Fock space $F^2(H_n)$. First, we prove that

(2.2)
$$\left\langle P_\rho(X, rR)\left(\sum_{|\beta| \le q} h_\beta \otimes e_\beta\right), \sum_{|\gamma| \le q} h_\gamma \otimes e_\gamma \right\rangle = \rho \sum_{|\beta|, |\gamma| \le q} \left\langle K_{\rho,X,r}(\gamma, \beta)h_\beta, h_\gamma\right\rangle,$$

where the multi-Toeplitz kernel $K_{\rho,X,r} : \mathbb{F}_n^+ \times \mathbb{F}_n^+ \to B(\mathcal{H})$, $r \in (0, 1)$, is defined by

$$K_{\rho,X,r}(\alpha, \beta) := \begin{cases} \frac{1}{\rho} r^{|\beta \backslash_l \alpha|} X_{\beta \backslash_l \alpha} & \text{if } \beta >_l \alpha \\ I & \text{if } \alpha = \beta \\ \frac{1}{\rho} r^{|\alpha \backslash_l \beta|} (X_{\alpha \backslash_l \beta})^* & \text{if } \alpha >_l \beta \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $\{h_\beta\}_{|\beta| \le q} \subset \mathcal{H}$, then we have

$$\begin{aligned}
&\left\langle \left(\rho I \otimes I + \sum_{k=1}^\infty \sum_{|\alpha|=k} X_\alpha^* \otimes r^k R_{\tilde\alpha}\right)\left(\sum_{|\beta| \le q} h_\beta \otimes e_\beta\right), \sum_{|\gamma| \le q} h_\gamma \otimes e_\gamma \right\rangle \\
&= \rho \sum_{|\beta| \le q} \|h_\beta\|^2 + \sum_{k=1}^\infty \sum_{|\alpha|=k} \left\langle \sum_{|\beta| \le q} X_\alpha^* h_\beta \otimes r^k R_{\tilde\alpha} e_\beta, \sum_{|\gamma| \le q} h_\gamma \otimes e_\gamma \right\rangle \\
&= \rho \sum_{|\beta| \le q} \|h_\beta\|^2 + + \sum_{|\alpha| \ge 1} \sum_{|\beta|, |\gamma| \le q} r^{|\alpha|} \left\langle e_{\beta\alpha}, e_\gamma \right\rangle \left\langle X_\alpha^* h_\beta, h_\gamma \right\rangle \\
&= \rho \sum_{|\beta| \le q} \|h_\beta\|^2 + \sum_{\gamma > \beta;\ |\beta|, |\gamma| \le q} r^{|\gamma \backslash_l \beta|} \left\langle X_{\gamma \backslash_l \beta}^* h_\beta, h_\gamma \right\rangle \\
&= \sum_{\gamma \ge \beta;\ |\beta|, |\gamma| \le q} \left\langle \rho K_{\rho,X,r}(\gamma, \beta)h_\beta, h_\gamma \right\rangle.
\end{aligned}$$

Now, taking into account that $K_{\rho,X,r}(\gamma,\beta) = K^*_{\rho,X,r}(\beta,\gamma)$, we deduce relation (2.2). Therefore, the condition $P_\rho(rA,R) \leq c^2 P_\rho(rB,R)$, $r \in [0,1)$, implies

$$[K_{\rho,A,r}(\alpha,\beta)]_{|\alpha|,|\beta|\leq q} \leq c^2[K_{\rho,B,r}(\alpha,\beta)]_{|\alpha|,|\beta|\leq q}$$

for any $0 < r < 1$ and $q = 0, 1, \ldots$. Taking $r \to 1$ in the latter inequality, we obtain item (iv).

Assume now that (iv) holds. Since $c^2 K_{\rho,B} - K_{\rho,A}$ is a positive semidefinite multi-Toeplitz kernel, due to Theorem 3.1 from [39] (see also the proof of Theorem 5.2 from [47]), we find a completely positive linear map $\mu : C^*(S_1, \ldots, S_n) \to B(\mathcal{E})$ such that

$$\mu(S_\alpha) = c^2 K_{\rho,B}(g_0,\alpha) - K_{\rho,A}(g_0,\alpha) = \frac{1}{\rho}(c^2 B_\alpha - A_\alpha)$$

for any $\alpha \in \mathbb{F}_n^+$ with $|\alpha| \geq 1$, and $\mu(I) = (c^2-1)I$. Since

$$P(rS,R) := \sum_{k=1}^\infty \sum_{|\alpha|=k} r^k S_\alpha \otimes R^*_{\tilde\alpha} + I \otimes I + \sum_{k=1}^\infty \sum_{|\alpha|=k} r^k S^*_\alpha \otimes R_{\tilde\alpha} \geq 0$$

for $r \in [0,1)$, we deduce that

$$\begin{aligned}(\mu \otimes \mathrm{id})[P(rS,R)] &= \sum_{k=1}^\infty \sum_{|\alpha|=k} \frac{1}{\rho} r^{|\alpha|}[c^2 B^*_\alpha - A^*_\alpha] \otimes R_{\tilde\alpha} + (c^2-1) I \otimes I \\ &\quad + \sum_{k=1}^\infty \sum_{|\alpha|=k} \frac{1}{\rho} r^{|\alpha|}[c^2 B_\alpha - A_\alpha] \otimes R^*_{\tilde\alpha} \\ &= c^2 P_\rho(rB,R) - P_\rho(rA,R) \geq 0,\end{aligned}$$

which implies (ii).

Let us prove that $(iv) \implies (vi)$. Assume that (iv) holds. Then we have $K_{\rho,A} \leq c^2 K_{\rho,B}$, where $K_{\rho,X}$ is the multi-Toeplitz kernel associated with $X \in \mathcal{C}_\rho$. Let $V := (V_1, \ldots, V_n)$ be the minimal isometric dilation of $A := (A_1, \ldots, A_n)$. Then $\mathcal{K}_A = \bigvee_{\alpha \in \mathbb{F}_n^+} V_\alpha \mathcal{H}$ and $\rho P_\mathcal{H} V_\alpha|_\mathcal{H} = A_\alpha$ for any $|\alpha| \geq 1$. Similar properties hold if $W := (W_1, \ldots, W_n)$ is the minimal isometric dilation of $B := (B_1, \ldots, B_n)$. Hence, and taking into account that $V_1, \ldots, V_n$ and $W_1, \ldots, W_n$ are isometries with orthogonal ranges, respectively, we have

$$\begin{aligned}&\left\| \sum_{|\alpha|\leq m} V_\alpha h_\alpha \right\|^2 = \\ &= \sum_{\alpha >_l \beta, |\alpha|,|\beta|\leq m} \left\langle V_{\alpha \backslash_l \beta} h_\alpha, h_\beta \right\rangle + \sum_{|\alpha|\leq m} \langle h_\alpha, h_\alpha \rangle + \sum_{\beta >_l \alpha, |\alpha|,|\beta|\leq m} \left\langle V^*_{\beta \backslash_l \alpha} h_\alpha, h_\beta \right\rangle \\ &= \sum_{\alpha >_l \beta, |\alpha|,|\beta|\leq m} \left\langle \frac{1}{\rho} A_{\alpha \backslash_l \beta} h_\alpha, h_\beta \right\rangle + \sum_{|\alpha|\leq m} \langle h_\alpha, h_\alpha \rangle + \sum_{\beta >_l \alpha, |\alpha|,|\beta|\leq m} \left\langle \frac{1}{\rho} A^*_{\beta \backslash_l \alpha} h_\alpha, h_\beta \right\rangle \\ &= \sum_{|\alpha|\leq m, |\beta|\leq m} \langle K_{\rho,A}(\beta,\alpha) h_\alpha, h_\beta \rangle = \left\langle [K_{\rho,A}(\beta,\alpha)]_{|\alpha|,|\beta|\leq m} \mathbf{h}_m, \mathbf{h}_m \right\rangle\end{aligned}$$

for any $m \in \mathbb{N}$ and $\mathbf{h}_m := \oplus_{|\alpha|\leq m} h_\alpha \in \oplus_{|\alpha|\leq m} \mathcal{H}_\alpha$, where each $\mathcal{H}_\alpha$ is a copy of $\mathcal{H}$. Similarly, we obtain

$$\left\| \sum_{|\alpha|\leq m} W_\alpha h_\alpha \right\|^2 = \left\langle [K_{\rho,B}(\beta,\alpha)]_{|\alpha|,|\beta|\leq m} \mathbf{h}_m, \mathbf{h}_m \right\rangle.$$

Taking into account that $K_{\rho,A} \leq c^2 K_{\rho,B}$, we deduce that

$$\left\| \sum_{|\alpha|\leq m} V_\alpha h_\alpha \right\| \leq c \left\| \sum_{|\alpha|\leq m} W_\alpha h_\alpha \right\|.$$

Therefore, we can define an operator $L_{B,A} : \mathcal{K}_B \to \mathcal{K}_A$ by setting

$$L_{B,A}\left(\sum_{|\alpha|\leq m} W_\alpha h_\alpha \right) := \sum_{|\alpha|\leq m} V_\alpha h_\alpha \tag{2.3}$$

for any $m \in \mathbb{N}$ and $h_\alpha \in \mathcal{H}$, $\alpha \in \mathbb{F}_n^+$. Note that $L_{B,A}$ is a bounded operator with $\|L_{B,A}\| \leq c$. Since $L_{B,A}|_{\mathcal{H}} = I_{\mathcal{H}}$, we have $\|L_{B,A}\| \geq 1$. It is easy to see that $L_{B,A} W_i = V_i L_{B,A}$ for $i = 1, \ldots, n$. Therefore item (vi) holds.

Conversely, assume that there is an operator $L_{B,A} \in B(\mathcal{K}_B, \mathcal{K}_A)$ with norm $\|L_{B,A}\| \leq c$ such that $L_{B,A}|_{\mathcal{H}} = I_{\mathcal{H}}$ and $L_{B,A} W_i = V_i L_{B,A}$, $i = 1, \ldots, n$. Then, we deduce that $L_{B,A}\left(\sum_{|\alpha|\leq m} W_\alpha h_\alpha \right) = \sum_{|\alpha|\leq m} V_\alpha h_\alpha$ for any $m \in \mathbb{N}$ and $h_\alpha \in \mathcal{H}$, $\alpha \in \mathbb{F}_n^+$. The condition $\|L_{B,A}\| \leq c$ implies

$$\left\| \sum_{|\alpha|\leq m} V_\alpha h_\alpha \right\|^2 \leq c^2 \left\| \sum_{|\alpha|\leq m} W_\alpha h_\alpha \right\|^2,$$

which is equivalent to the inequality

$$\left\langle [K_{\rho,A}(\beta,\alpha)]_{|\alpha|,|\beta|\leq m} \mathbf{h}_m, \mathbf{h}_m \right\rangle \leq c^2 \left\langle [K_{\rho,B}(\beta,\alpha)]_{|\alpha|,|\beta|\leq m} \mathbf{h}_m, \mathbf{h}_m \right\rangle$$

for any $m \in \mathbb{N}$ and $\mathbf{h}_m := \oplus_{|\alpha|\leq m} h_\alpha \in \oplus_{|\alpha|\leq m} \mathcal{H}_\alpha$. Consequently, we deduce item (iv). The proof is complete. □

A closer look at the proof of Theorem 2.2 reveals that one can assume that $u(0) = I$ in part (iii), and one can also assume that $\Re p(0) = I$ in the definition of the Harnack domination $A \overset{H}{\prec} B$. We also remark that, due to Theorem 1.3, we can add an equivalence to Theorem 2.2, namely, $A \underset{c}{\overset{H}{\prec}} B$ if and only if

$$u(A_1, \ldots, A_n) + (\rho - 1)u(0) \leq c^2 \left[u(B_1, \ldots, B_n) + (\rho - 1)u(0) \right]$$

for any positive free pluriharmonic function $u \in Har^c_{\mathbf{ball}}(B(\mathcal{E}))$.

Corollary 2.3. *If $A, B \in \mathcal{C}_\rho$ and $A \overset{H}{\prec} B$, then*

$$\begin{aligned}\|L_{B,A}\| &= \inf\{c > 1 : \ A \underset{c}{\overset{H}{\prec}} B\} \\ &= \inf\{c > 1 : \ P_\rho(rA, R) \leq c^2 P_\rho(rB, R) \quad \text{for any} \quad r \in [0,1)\}.\end{aligned}$$

Moreover, $A \overset{H}{\prec} B$ if and only if $\sup_{r\in[0,1)} \|L_{rA,rB}\| < \infty$. In this case,

$$\|L_{A,B}\| = \sup_{r\in[0,1)} \|L_{rA,rB}\|$$

and the mapping $r \mapsto \|L_{rA,rB}\|$ is increasing on $[0,1)$.

Proof. Assume that $A \overset{H}{\prec} B$. Then, due to Theorem 2.2, $A \underset{c}{\overset{H}{\prec}} B$ if and only if there is an operator $L_{B,A} \in B(\mathcal{K}_B, \mathcal{K}_A)$ with $\|L_{B,A}\| \leq c$ such that $L_{B,A}|_\mathcal{H} = I_\mathcal{H}$ and $L_{B,A} W_i = V_i L_{B,A}$ for $i = 1, \ldots, n$. Consequently, taking $c = \|L_{B,A}\|$, we deduce that $A \underset{\|L_{B,A}\|}{\overset{H}{\prec}} B$, which is equivalent to

$$P_\rho(rA, R) \leq \|L_{B,A}\|^2 P_\rho(rB, R)$$

for any $r \in [0,1)$. Hence, we have $tA \underset{\|L_{B,A}\|}{\overset{H}{\prec}} tB$ for any $t \in [0,1)$. Applying again Theorem 2.2 to the operators tA and tB, we deduce that $\|L_{tA,tB}\| \leq \|L_{B,A}\|$. Conversely, suppose that $c := \sup_{r\in[0,1)} \|L_{rA,rB}\| < \infty$. Since $\|L_{rA,rB}\| \leq c$, Theorem 2.2 implies $rA \underset{c}{\overset{H}{\prec}} rB$ for any $r \in [0,1)$ and, therefore, $P_\rho(rtA, R) \leq c^2 P_\rho(rtB, R)$ for any $t, r \in [0,1)$. Hence, $A \underset{c}{\overset{H}{\prec}} B$ and, consequently, $\|L_{B,A}\| \leq c$. Therefore, $\|L_{A,B}\| = \sup_{r\in[0,1)} \|L_{rA,rB}\|$. The fact that $r \mapsto \|L_{rA,rB}\|$ is an increasing function on $[0,1)$ follows from the latter relation. This completes the proof. □

We remark that if $1 \leq m < n$ and u is a positive free pluriharmonic function on $[B(\mathcal{K})^n]_1$, then the map

$$(X_1, \ldots, X_m) \mapsto u(X_1, \ldots, X_m, 0, \ldots, 0)$$

is a positive free pluriharmonic function on $[B(\mathcal{K})^m]_1$. Moreover, if g is a positive free pluriharmonic function on $[B(\mathcal{K})^m]_1$, then the map

$$(X_1, \ldots, X_n) \mapsto g(X_1, \ldots, X_m, 0, \ldots, 0)$$

is a positive free pluriharmonic function on $[B(\mathcal{K})^n]_1$. Consequently, using Corollary 1.2, one can easily deduce the following result.

Corollary 2.4. *Let $c > 0$, $\rho > 0$, and $1 \leq m < n$. Consider two n-tuples $(A_1, \ldots, A_m) \in B(\mathcal{H})^m$ and $(B_1, \ldots, B_m) \in B(\mathcal{H})^m$ in the class $\mathcal{C}_\rho$ and let*

$(A_1,\ldots,A_m,0,\ldots,0)$ and $(B_1,\ldots,B_m,0,\ldots,0)$ be their extensions in $B(\mathcal{H})^n$, respectively. Then $(A_1,\ldots,A_m)\overset{H}{\underset{c}{\prec}}(B_1,\ldots,B_m)$ in $\mathcal{C}_\rho\subset B(\mathcal{H})^m$ if and only if

$$(A_1,\ldots,A_m,0,\ldots,0)\overset{H}{\underset{c}{\prec}}(B_1,\ldots,B_m,0,\ldots,0) \quad \text{in } \ \mathcal{C}_\rho\subset B(\mathcal{H})^n.$$

We recall (e.g. [43]) that if $(T_1,\ldots T_n)$ is an n-tuple of operators, then the joint spectral radius $r(T_1,\ldots,T_n)<1$ if and only if $\lim_{k\to\infty}\left\|\sum_{|\alpha|=k}T_\alpha T_\alpha^*\right\|=0$. In what follows, we characterize the elements of $\mathcal{C}_\rho$ which are Harnack dominated by 0.

Theorem 2.5. *Let $A:=(A_1,\ldots,A_n)$ be in $\mathcal{C}_\rho$. Then $A\overset{H}{\prec}0$ if and only if the joint spectral radius $r(A_1,\ldots,A_n)<1$.*

Proof. Note that the map $X\mapsto P_\rho(X,R)$ is a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with coefficients in $B(F^2(H_n))$ and has the factorization

(2.4)
$$\begin{aligned}
&P_\rho(X,R)=\\
&=(I-R_X)^{-1}+(\rho-2)I+(I-R_X^*)^{-1}\\
&=(I-R_X^*)^{-1}\left[I-R_X+(\rho-2)(I-R_X^*)(I-R_X)+I-R_X^*\right](I-R_X)^{-1}\\
&=(I-R_X^*)^{-1}\left[\rho I+(1-\rho)(R_X^*+R_X)+(\rho-2)R_X^*R_X\right](I-R_X)^{-1},
\end{aligned}$$

where $R_X:=X_1^*\otimes R_1+\cdots+X_n^*\otimes R_n$ is the reconstruction operator associated with the n-tuple $X:=(X_1,\ldots,X_n)\in[B(\mathcal{H})^n]_1$. We remark that, due to the fact that the spectral radius of R_X is equal to the joint spectral radius $r(X_1,\ldots,X_n)$, the factorization above holds for any $X\in\mathcal{C}_\rho$ with $r(X_1,\ldots,X_n)<1$.

Now, using Theorem 2.2 part (ii) and the above-mentioned factorization, we deduce that $A\overset{H}{\prec}0$ if and only if there exists $c>0$ such that

$$(I-R_{rA}^*)^{-1}\left[\rho I+(1-\rho)(R_{rA}^*+R_{rA})+(\rho-2)R_{rA}^*R_{rA}\right](I-R_{rA})^{-1}\le\rho c^2 I$$

for any $r\in[0,1)$. Similar inequality holds if we replace the right creation operators by the left creation operators. Then, applying the noncommutative Poisson transform $\mathrm{id}\otimes P_{e^{i\theta}R}$, where $R:=(R_1,\ldots,R_n)$, we obtain

(2.5)
$$\rho I+(1-\rho)(e^{-i\theta}R_{rA}^*+e^{i\theta}R_{rA})+(\rho-2)R_{rA}^*R_{rA}\le\rho c^2(I-re^{-i\theta}R_A^*)(I-re^{i\theta}R_A)$$

for any $r\in[0,1)$ and $\theta\in\mathbb{R}$.

On the other hand, since $A:=(A_1,\ldots,A_n)\in\mathcal{C}_\rho$, we have $r(A_1,\ldots,A_n)\le 1$. Suppose that $r(A_1,\ldots,A_n)=1$. Taking into account that $r(R_A)=r(A_1,\ldots,A_n)$, we can find $\lambda_0\in\mathbb{T}$ in the approximative spectrum of R_A. Consequently, there is a sequence $\{h_m\}$ in $\mathcal{H}\otimes F^2(H_n)$ such that $\|h_m\|=1$ and

$$\lambda_0 h_m-R_Ah_m\to 0 \quad \text{as } m\to\infty. \tag{2.6}$$

In particular, relation (2.5) implies

$$\begin{aligned}(2.7)\quad &\rho\|h_m\|^2+(1-\rho)\left[\langle\lambda_0 R_{rA}^*h_m,h_m\rangle+\left\langle\bar\lambda_0 R_{rA}h_m,h_m\right\rangle\right]+(\rho-2)\|R_{rA}h_m\|^2\\ &\qquad\le\rho c^2\|h_m-\bar\lambda_0 R_{rA}h_m\|^2\end{aligned}$$

for any $r\in(0,1)$ and $m\in\mathbb{N}$. Note that due to (2.6) and the fact that $|\lambda_0|=1$, we have

$$\left\langle\bar\lambda_0 R_A h_m,h_m\right\rangle=\bar\lambda_0\langle R_Ah_m-\lambda_0h_m,h_m\rangle+1\to1,\qquad\text{as }\ m\to\infty.$$

Since

$$\begin{aligned}\|h_m-\bar\lambda_0R_{rA}h_m\|&\le\|h_m-\bar\lambda_0R_Ah_m\|+\|\bar\lambda_0(R_Ah_m-R_{rA}h_m)\|\\ &=\|\bar\lambda_0h_m-R_Ah_m\|+(1-r)\|R_Ah_m\|\end{aligned}$$

and due to the fact that $\|R_Ah_m\|\to1$ as $m\to\infty$, we deduce that

$$\limsup_{m\to\infty}\|h_m-\bar\lambda_0R_{rA}h_m\|\le1-r$$

for any $r\in(0,1)$. Now, since $R_{rA}=rR_A$ and taking $m\to\infty$ in relation (2.7), we obtain

$$\rho+2(1-\rho)r+(\rho-2)r^2\le c^2\rho(1-r)^2$$

for any $r\in(0,1)$. Setting $r=1-\frac{1}{m}$, $m\ge2$, straightforward calculations imply $2m\le\rho c^2-\rho+2$ for any $m\in\mathbb{N}$, which is a contradiction. Therefore, we must have $r(A_1,\dots,A_n)<1$.

Conversely, assume that $A:=(A_1,\dots,A_n)\in\mathcal{C}_\rho$ has the joint spectral radius $r(A_1,\dots,A_n)<1$. Since $r(A_1,\dots,A_n)=r(R_A)$, one can see that $M:=\sup_{r\in(0,1)}\|(I-rR_A)^{-1}\|$ exists and $M\ge1$. Hence

$$M^2(I-R_{rA}^*)(I-R_{rA})\ge I\ge I-R_{rA}^*R_{rA}\tag{2.8}$$

for any $r\in(0,1)$. Now we consider the case $\rho\ge1$. Note that relation (2.8) implies

$$I-R_{rA}^*R_{rA}+(\rho-1)(I-R_{rA}^*)(I-R_{rA})\le\rho M^2(I-R_{rA}^*)(I-R_{rA}).$$

The latter inequality is equivalent to

$$\rho I+(1-\rho)(R_{rA}^*+R_{rA})+(\rho-2)R_{rA}^*R_{rA}\le\rho M^2(I-R_{rA}^*)(I-R_{rA}),$$

which, due to the factorization (2.4), is equivalent to

$$P_\rho(rA,R)\le\rho M^2=M^2P_\rho(0,R)$$

for any $r\in[0,1)$. According to Theorem 2.2, we deduce that $A\overset{H}{\prec}0$.

Now, consider the case when $\rho\in(0,1)$. Since $\|R_{rA}\|\le r\rho$ and $\delta-2<0$, we have

$$\begin{aligned}\rho I+(1-\rho)(R_{rA}^*+R_{rA})+(\rho-2)R_{rA}^*R_{rA}&\le\rho I+(1-\rho)(R_{rA}^*+R_{rA})\\ &\le\rho I+2(1-\rho)r\rho\le(3\rho-2\rho^2)I.\end{aligned}$$

Using again the factorization (2.4), we deduce that

$$P_\rho(rA,R)\le(3\rho-2\rho^2)(I-R_{rA}^*)^{-1}(I-R_{rA})^{-1}$$

for any $r \in (0,1)$. Hence and using the fact that $(I - R_{rA}^*)^{-1}(I - R_{rA})^{-1} \le M^2 I$, we obtain

$$P_\rho(rA, R) \le (3 - 2\rho) M^2 P_\rho(0, R)$$

for any $r \in (0,1)$. Using again Theorem 2.2, we get $A \overset{H}{\prec} 0$. The proof is complete. $\square$

We mention that in the particular case when $n = 1$ we can recover a result obtained by Ando, Suciu, and Timotin [1], when $\rho = 1$, and by G. Cassier and N. Suciu [9], when $\rho \neq 1$.

3. Hyperbolic metric on Harnack parts of the noncommutative ball $\mathcal{C}_\rho$

The relation $\overset{H}{\prec}$ induces an equivalence relation $\overset{H}{\sim}$ on the class $\mathcal{C}_\rho$. We provide a Harnack type double inequality for positive free pluriharmonic functions on the noncommutative ball $\mathcal{C}_\rho$ and use it to prove that the Harnack part of $\mathcal{C}_\rho$ which contains 0 coincides with the open noncommutative ball $[\mathcal{C}_\rho]_{<1}$. We introduce a hyperbolic metric on any Harnack part of $\mathcal{C}_\rho$ and obtain a concrete formula in terms of the reconstruction operator.

Since $\overset{H}{\prec}$ is a preorder relation on $\mathcal{C}_\rho$, it induces an equivalence relation $\overset{H}{\sim}$ on $\mathcal{C}_\rho$, which we call Harnack equivalence. The equivalence classes with respect to $\overset{H}{\sim}$ are called Harnack parts of $\mathcal{C}_\rho$. Let $A := (A_1, \ldots, A_n)$ and $B := (B_1, \ldots, B_n)$ be in $\mathcal{C}_\rho$. We say that A and B are Harnack equivalent (we denote $A \overset{H}{\sim} B$) if and only if there exists $c \ge 1$ such that

$$\begin{aligned}\frac{1}{c^2}\left[\Re p(B_1, \ldots, B_n) + (\rho - 1)\Re p(0)\right] &\le \Re p(A_1, \ldots, A_n) + (\rho - 1)\Re p(0)\\ &\le c^2\left[\Re p(B_1, \ldots, B_n) + (\rho - 1)\Re p(0)\right]\end{aligned}$$

for any noncommutative polynomial with matrix-valued coefficients $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, such that $\Re p(X) \ge 0$ for any $X \in [B(\mathcal{H})^n]_1$. We also use the notation $A \underset{c}{\overset{H}{\sim}} B$ when $A \underset{c}{\overset{H}{\prec}} B$ and $B \underset{c}{\overset{H}{\prec}} A$. We remark that Theorem 2.2 can be used to provide several characterizations for the Harnack parts of $\mathcal{C}_\rho$.

The first result is an extension of Harnack inequality to positive free pluriharmonic functions on the noncommutative ball $\mathcal{C}_\rho$, $\rho > 0$.

Theorem 3.1. *If u is a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with operator-valued coefficients in $B(\mathcal{E})$ and $0 \le r < 1$, then*

$$u(0)\,\frac{1 - r(2\rho - 1)}{1 + r} \le u(rX_1, \ldots, rX_n) \le u(0)\,\frac{1 + r(2\rho - 1)}{1 - r}$$

for any $(X_1, \ldots, X_n) \in \mathcal{C}_\rho$.

Proof. Let

$$u(Z_1,\ldots,Z_n)=\sum_{k=1}^{\infty}\sum_{|\alpha|=k} Z_\alpha^*\otimes A_{(\alpha)}^* + I\otimes A_{(0)} + \sum_{k=1}^{\infty}\sum_{|\alpha|=k} Z_\alpha\otimes A_{(\alpha)}$$

be a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$. According to Theorem 1.4 from [49], for any $Y\in[B(\mathcal{H})^n]_1^-$ and $r\in[0,1)$, we have

$$u(0)\,\frac{1-r}{1+r}\le u(rY_1,\ldots,rY_n)\le u(0)\,\frac{1+r}{1-r}. \tag{3.1}$$

On the other hand, let $(X_1,\ldots,X_n)\in\mathcal{C}_\rho$ and let $(V_1,\ldots,V_n)$ be the minimal isometric dilation of $(X_1,\ldots,X_n)$ on a Hilbert space $\mathcal{K}_T\supseteq\mathcal{H}$. Since $X_\alpha=\rho P_{\mathcal{H}}V_\alpha|_{\mathcal{H}}$ for any $\alpha\in\mathbb{F}_n^+\backslash\{g_0\}$, and using the free pluriharmonic functional calculus, we have

$$\begin{aligned}
&u(rX_1,\ldots,rX_n)=\\
&=\sum_{k=1}^{\infty}\sum_{|\alpha|=k} r^{|\alpha|}X_\alpha^*\otimes A_{(\alpha)}^* + I\otimes A_{(0)} + \sum_{k=1}^{\infty}\sum_{|\alpha|=k} r^{|\alpha|}X_\alpha\otimes A_{(\alpha)}\\
&=\rho(P_{\mathcal{H}}\otimes I_{\mathcal{E}})\left[\sum_{k=1}^{\infty}\sum_{|\alpha|=k} r^{|\alpha|}V_\alpha^*\otimes A_{(\alpha)}^*\right]|_{\mathcal{H}\otimes\mathcal{E}} + I_{\mathcal{H}}\otimes A_{(0)}\\
&\qquad + \rho(P_{\mathcal{H}}\otimes I_{\mathcal{E}})\left[\sum_{k=1}^{\infty}\sum_{|\alpha|=k} r^{|\alpha|}V_\alpha\otimes A_{(\alpha)}\right]|_{\mathcal{H}\otimes\mathcal{E}}\\
&=\rho(P_{\mathcal{H}}\otimes I_{\mathcal{E}})u(rV_1,\ldots,rV_n)|_{\mathcal{H}\otimes\mathcal{E}}+(1-\rho)u(0),
\end{aligned}$$

where the convergence is in the operator norm topology. Due to (3.1), we have

$$u(0)\,\frac{1-r}{1+r}\le u(rV_1,\ldots,V_n)\le u(0)\,\frac{1+r}{1-r}.$$

Consequently, we deduce that

$$\begin{aligned}
u(0)\left[\frac{\rho(1-r)}{1+r}+(1-\rho)\right]&\le \rho(P_{\mathcal{H}}\otimes I_{\mathcal{E}})u(rV_1,\ldots,rV_n)|_{\mathcal{H}\otimes\mathcal{E}}+(1-\rho)u(0)\\
&\le u(0)\left[\frac{\rho(1+r)}{1-r}+(1-\rho)\right].
\end{aligned}$$

Since

$$u(rX_1,\ldots,rX_n)=\rho(P_{\mathcal{H}}\otimes I_{\mathcal{E}})u(rV_1,\ldots,rV_n)|_{\mathcal{H}\otimes\mathcal{E}}+(1-\rho)u(0),$$

the result follows. □

Now, we can determine the Harnack part of $\mathcal{C}_\rho$ which contains 0.

THEOREM 3.2. *Let $A:=(A_1,\ldots,A_n)$ be in $\mathcal{C}_\rho$. Then the following statements are equivalent:*

(i) $\omega_\rho(A_1,\dots,A_n)<1$;

(ii) $A\overset{H}{\sim}0$;

(iii) $r(A_1,\dots,A_n)<1$ *and* $P_\rho(A,R)\geq aI$ *for some constant* $a>0$.

Proof. First, we prove that $(i)\implies(ii)$. Let $A:=(A_1,\dots,A_n)$ be in $\mathcal{C}_\rho$ and assume that $\omega_\rho(A)<1$. Then there is $r_0\in(0,1)$ such that $\omega_\rho(\frac{1}{r_0}A)=\frac{1}{r_0}\omega_\rho(A)<1$. Consequently, $\frac{1}{r_0}A\in\mathcal{C}_\rho$.
According to Theorem 3.1, we have

$$\Re p(0)\,\frac{1-r_0(2\rho-1)}{1+r_0}\leq\Re p(A_1,\dots,A_n)\leq\Re p(0)\,\frac{1+r_0(2\rho-1)}{1-r_0}$$

for any noncommutative polynomial with matrix-valued coefficients $p\in\mathbb{C}[X_1,\dots,X_n]\otimes M_m$, $m\in\mathbb{N}$, such that $\Re p\geq 0$ on $[B(\mathcal{H})^n]_1$. Hence, we deduce that $A\overset{H}{\sim}0$.

To prove that $(ii)\implies(iii)$, assume that $A\overset{H}{\sim}0$. Due to Theorem 2.5, we have $r(A)<1$. Using now Theorem 2.2, we deduce that there exists $c>0$ such that

$$P_\rho(rA,R)\geq\frac{1}{c^2}P_\rho(0,R)=\frac{\rho}{c^2}I \tag{3.2}$$

for any $r\in[0,1)$. Since $r(A)<1$, one can prove that $\lim_{r\to 1}P_\rho(rA,R)=P_\rho(A,R)$ in the operator norm topology. Consequently, taking $r\to 1$ in relation (3.2), we obtain item (iii).

It remains to show that $(iii)\implies(i)$. Assume that $r(A_1,\dots,A_n)<1$ and $P_\rho(A,R)\geq aI$ for some constant $a>0$. Note that there exists $t_0\in(0,1)$ such that the map

$$t\mapsto\left(I-\sum_{i=1}^n A_i^*\otimes tR_i\right)^{-1}+(\rho-2)I+\left(I-\sum_{i=1}^n A_i\otimes tR_i^*\right)^{-1}$$

is well-defined and continuous on $[0,1+t_0]$ in the operator norm topology. In particular, there is $\epsilon_0\in(0,t_0)$ such that

$$\|P_\rho(A,R)-P_\rho(A,tR)\|<\frac{a}{2}$$

for any $t\in(1-\epsilon_0,1+\epsilon_0)$. Consequently, if $\gamma_0\in(1,1+\epsilon_0)$, then

$$P_\rho(\gamma_0A,R)\geq P_\rho(A,R)-\|P_\rho(A,R)-P_\rho(\gamma_0A,R)\|I\geq\frac{a}{2}I>0.$$

Due to Theorem 1.1, we have $\gamma_0A\in\mathcal{C}_\rho$, which implies $\omega(\gamma_0A)\leq 1$. Therefore, $\omega(A)\leq\frac{1}{\gamma_0}<1$ and item (i) holds. The proof is complete. $\square$

We remark that, when $n=1$, we recover a result obtain by Foiaş [15] if $\rho=1$, and by Cassier and Suciu [9] if $\rho>0$.

Given $A, B \in \mathcal{C}_\rho$, $\rho > 0$, in the same Harnack part of $\mathcal{C}_\rho$, i.e., $A \overset{H}{\sim} B$, we introduce

$$\Lambda_\rho(A, B) := \inf\left\{c > 1 : \; A \underset{c}{\overset{H}{\sim}} B\right\}. \tag{3.3}$$

Note that, due to Theorem 2.2, $A \overset{H}{\sim} B$ if and only if the operator $L_{B,A}$ is invertible. In this case, $L_{B,A}^{-1} = L_{A,B}$ and

$$\Lambda_\rho(A, B) = \max\{\|L_{A,B}\|, \|L_{B,A}\|\}.$$

To prove the latter equality, assume that $A \underset{c}{\overset{H}{\sim}} B$ for some $c \geq 1$. Due to the same theorem, we have $\|L_{B,A}\| \leq c$ and $\|L_{A,B}\| \leq c$. Consequently,

$$\max\{\|L_{A,B}\|, \|L_{B,A}\|\} \leq \inf\left\{c \geq 1 : \; A \underset{c}{\overset{H}{\sim}} B\right\} = \Lambda_\rho(A, B). \tag{3.4}$$

On the other hand, setting $c_0 := \|L_{B,A}\|$ and $c_0' := \|L_{A,B}\|$, Theorem 2.2 implies $A \underset{c_0}{\overset{H}{\prec}} B$ and $B \underset{c_0'}{\overset{H}{\prec}} A$. Hence, we deduce that $A \underset{d}{\overset{H}{\sim}} B$, where $d := \max\{c_0, c_0'\}$. Consequently, $\Lambda_\rho(A, B) \leq d$, which together with relation (3.4) imply $\Lambda_\rho(A, B) = \max\{\|L_{A,B}\|, \|L_{B,A}\|\}$, which proves our assertion.

Now, we can introduce a hyperbolic (*Poincaré-Bergman* type) metric $\delta_\rho : \Delta \times \Delta \to \mathbb{R}^+$ on any Harnack part Δ of $\mathcal{C}_\rho$, by setting

$$\delta_\rho(A, B) := \ln \Lambda_\rho(A, B), \qquad A, B \in \Delta. \tag{3.5}$$

Due to our discussion above, we also have

$$\delta_\rho(A, B) = \ln \max\left\{\|L_{A,B}\|, \left\|L_{A,B}^{-1}\right\|\right\}.$$

PROPOSITION 3.3. *δ_ρ is a metric on any Harnack part of $\mathcal{C}_\rho$.*

Proof. The proof is similar to that of Proposition 2.2 from [49], but uses ρ-pluriharmonic kernels. ☐

We remark that, according to Theorem 3.2, the set

$$[\mathcal{C}_\rho]_{<1} := \{(X_1, \ldots, X_n) \in B(\mathcal{H})^n : \; \omega_\rho(X_1, \ldots, X_n) < 1\}$$

is the Harnack part of $\mathcal{C}_\rho$ containing 0.

In what follows we calculate the norm of $L_{Y,X}$ with $X, Y \in [\mathcal{C}_\rho]_{<1}$, in terms of the reconstruction operators.

THEOREM 3.4. *If $X, Y \in [\mathcal{C}_\rho]_{<1}$, then $\|L_{Y,X}\| = \|C_{\rho,X} C_{\rho,Y}^{-1}\|$, where*

$$C_{\rho,X} := \Delta_{\rho,X}(I - R_X)^{-1},$$
$$\Delta_{\rho,X} := \left[\rho I + (1-\rho)(R_X^* + R_X) + (\rho - 2)R_X^* R_X\right]^{1/2}.$$

Moreover, if $X, Y \in \mathcal{C}_\rho$ is such that $X \overset{H}{\prec} Y$, then $\|L_{Y,X}\| = \sup_{r \in [0,1)} \|C_{\rho,rX} C_{\rho,rY}^{-1}\|$.

Proof. Since $X, Y \in [\mathcal{C}_\rho]_{<1}$, Theorem 3.2 implies $X \overset{H}{\sim} Y$, $r(X) < 1$, and $r(Y) < 1$. Let $c > 1$ and assume that $P_\rho(rX, R) \leq c^2 P_\rho(rY, R)$ for any $r \in [0, 1)$. Since $r(X) < 1$ and $r(Y) < 1$, we can take the limit, as $r \to 1$, in the operator norm topology, and obtain $P_\rho(X, R) \leq c^2 P_\rho(Y, R)$. Conversely, if the latter inequality holds, then $P_\rho(X, S) \leq c^2 P_\rho(Y, S)$, where $S := (S_1, \ldots, S_n)$ is the n-tuple of left creation operators. Applying the noncommutative Poisson transform $\mathrm{id} \otimes P_{rR}$, $r \in [0, 1)$, and taking into account that it is a positive map, we deduce that $P_\rho(rX, R) \leq c^2 P_\rho(rY, R)$ for any $r \in [0, 1)$.

Therefore, due to Theorem 2.2, we have

$$P_\rho(X, R) \leq c^2 P_\rho(Y, R) \quad \text{if and only if} \quad \|L_{Y,X}\| \leq c. \tag{3.6}$$

We recall that the free pluriharmonic kernel $P_\rho(X, R)$ with $X \in [\mathcal{C}_\rho]_{<1}$, has the factorization $P(X, R) = C^*_{\rho,X} C_{\rho,X}$. Due to Theorem 3.2, $P_\rho(X, R)$ is invertible and, consequently, so is $C_{\rho,X}$. Consequently,

$$P_\rho(X, R) \leq c^2 P_\rho(Y, R) \quad \text{if and only if} \quad {C^*_{\rho,Y}}^{-1} C^*_{\rho,X} C_{\rho,X} C^{-1}_{\rho,Y} \leq c^2 I.$$

Setting $c_0 := \|C_{\rho,X} C^{-1}_{\rho,Y}\|$, we have $P_\rho(X, R) \leq c_0^2 P_\rho(Y, R)$. Now, due to relation (3.6), we obtain

$$\|L_{Y,X}\| \leq c_0 = \|C_{\rho,X} C^{-1}_{\rho,Y}\|.$$

Setting $c_0' := \|L_{Y,X}\|$ and using again (3.6), we obtain $P_\rho(X, R) \leq {c_0'}^2 P_\rho(Y, R)$. Hence, we deduce that ${C^*_{\rho,Y}}^{-1} C^*_{\rho,X} C_{\rho,X} C^{-1}_{\rho,Y} \leq {c_0'}^2 I$, which implies

$$\|C_{\rho,X} C^{-1}_{\rho,Y}\| \leq c_0' = \|L_{Y,X}\|.$$

Therefore, $\|L_{Y,X}\| = \|C_{\rho,X} C^{-1}_{\rho,Y}\|$. The last part of the theorem is now obvious. □

Combining Theorem 3.4 with the remarks preceding Proposition 3.3, we obtain a concrete formula for the hyperbolic metric δ_ρ on $[\mathcal{C}_\rho]_{<1}$ in terms of the reconstruction operator, which is the main result of this section.

Theorem 3.5. *Let $\delta_\rho : [\mathcal{C}_\rho]_{<1} \times [\mathcal{C}_\rho]_{<1} \to [0, \infty)$ be the hyperbolic metric. If $X, Y \in [\mathcal{C}_\rho]_{<1}$, then*

$$\delta_\rho(X, Y) = \ln \max\left\{ \left\| C_{\rho,X} C^{-1}_{\rho,Y} \right\|, \left\| C_{\rho,Y} C^{-1}_{\rho,X} \right\| \right\},$$

where

$$\begin{aligned} C_{\rho,X} &:= \Delta_{\rho,X}(I - R_X)^{-1}, \\ \Delta_{\rho,X} &:= \left[\rho I + (1-\rho)(R^*_X + R_X) + (\rho - 2) R^*_X R_X\right]^{1/2}, \end{aligned}$$

and $R_X := X_1^* \otimes R_1 + \cdots + X_n^* \otimes R_n$ is the reconstruction operator associated with the right creation operators $R_1, \ldots, R_n$ and $X := (X_1, \ldots, X_n) \in [\mathcal{C}_\rho]_{<1}$.

Using Theorem 2.2, one can easily obtain the following result. Since the proof is similar to that of Lemma 2.6 from [49], we shall omit it.

Lemma 3.6. *Let $X := (X_1, \dots, X_n)$ and $Y := (Y_1, \dots, Y_n)$ be in $\mathcal{C}_\rho$. Then the following properties hold.*

(i) *$X \overset{H}{\sim} Y$ if and only if $rX \overset{H}{\sim} rX$ for any $r \in [0,1)$ and $\sup_{r\in[0,1)} \Lambda_\rho(rX, rY) < \infty$. In this case,*

$$\Lambda_\rho(X,Y) = \sup_{r\in[0,1)} \Lambda_\rho(rX, rY) \quad \text{and} \quad \delta_\rho(X,Y) = \sup_{r\in[0,1)} \delta_\rho(rX, rY).$$

(ii) *If $X \overset{H}{\sim} Y$, then the functions $r \mapsto \Lambda_\rho(rX, rY)$ and $r \mapsto \delta_\rho(rX, rY)$ are increasing on $[0,1)$.*

Putting together Theorem 3.5 and Lemma 3.6, we deduce the following result.

Theorem 3.7. *Let $X := (X_1, \dots, X_n)$ and $Y := (Y_1, \dots, Y_n)$ be in $\mathcal{C}_\rho$ such that $X \overset{H}{\sim} Y$. Then the metric δ_ρ satisfies the relation*

$$\delta_\rho(X,Y) = \ln \max \left\{ \sup_{r\in[0,1)} \left\| C_{\rho,rX} C_{\rho,rY}^{-1} \right\|, \sup_{r\in[0,1)} \left\| C_{\rho,rY} C_{\rho,rX}^{-1} \right\| \right\},$$

where $C_{\rho,X} := \Delta_{\rho,X}(I - R_X)^{-1}$ and $R_X := X_1^* \otimes R_1 + \dots + X_n^* \otimes R_n$ is the reconstruction operator.

Using the Harnack type inequality of Theorem 3.1, we obtain an upper bound for the hyperbolic distance δ_ρ on $[\mathcal{C}_\rho]_{<1}$. First, we need the following result.

Proposition 3.8. *Let f be in the noncommutative disc algebra $\mathcal{A}_n$ such that $\Re f \geq 0$ and let $X := (X_1, \dots, X_n) \in \mathcal{C}_\rho$ be with $\omega_\rho(X) < 1$. Then*

$$\rho \frac{1 - \omega_\rho(X)}{1 + \omega_\rho(X)} \Re f(0) \leq \Re f(X_1, \dots, X_n) + (\rho - 1)\Re f(0) \leq \rho \frac{1 + \omega_\rho(X)}{1 - \omega_\rho(X)}.$$

Proof. Let $r := \omega_\rho(X)$ and define $Y := \frac{1}{r}X$. Since $\omega_\rho(Y) = \frac{1}{r}\omega_\rho(X) = 1$, we deduce that $Y \in \mathcal{C}_\rho$. Applying Theorem 3.1 to Y, we obtain

$$\frac{1 - \omega_\rho(X)(2\rho - 1)}{1 + \omega_\rho(X)} \Re f(0) \leq \Re f(X_1, \dots, X_n) \leq \rho \frac{1 + \omega_\rho(X)(2\rho - 1)}{1 - \omega_\rho(X)}.$$

It is easy to see that the latter inequality is equivalent to the one from the proposition. □

Now, we can deduce the following upper bound for the hyperbolic distance on $[\mathcal{C}_\rho]_{<1}$.

Corollary 3.9. *For any $X, Y \in [\mathcal{C}_\rho]_{<1}$,*

$$\delta_\rho(X,Y) \leq \frac{1}{2} \ln \frac{(1 + \omega_\rho(X))(1 + \omega_\rho(Y))}{(1 - \omega_\rho(X))(1 - \omega_\rho(Y))}.$$

Proof. Using Theorem 2.2 and the inequality of Proposition 3.8, we deduce that

$$\Lambda_\rho(X,0) \leq \left(\frac{1+\omega_\rho(X)}{1-\omega_\rho(X)}\right)^{1/2}.$$

On the other hand, since δ_ρ is a metric on $[\mathcal{C}_\rho]_{<1}$, we have $\delta(X,Y) \leq \delta(X,0) + \delta_\rho(Y,0)$. Taking into account that $\delta_\rho(X,Y) = \ln \Lambda_\rho(X,Y)$, the result follows. □

We remark that when $\rho = 1$, the inequality of Corollary 3.9 is sharper then the one obtained in Corollary 2.5 from [49].

Using Corollary 2.4, on can easily obtain the following result.

Corollary 3.10. *Let $\rho > 0$, and $1 \leq m < n$. Consider two n-tuples $A := (A_1,\ldots,A_m) \in B(\mathcal{H})^m$ and $B := (B_1,\ldots,B_m) \in B(\mathcal{H})^m$ in the class $\mathcal{C}_\rho$ and their extensions $\widetilde{A} := (A_1,\ldots,A_m,0,\ldots,0)$ and $\widetilde{B} := (B_1,\ldots,B_m,0,\ldots,0)$ in $B(\mathcal{H})^n$, respectively. Then*

$$A \overset{H}{\sim} B \quad \text{if and only if} \quad \widetilde{A} \overset{H}{\sim} \widetilde{B}.$$

Moreover, in this case,

$$\delta_\rho(A,B) = \delta_\rho(\widetilde{A},\widetilde{B}).$$

In what follows we provide a few properties for the map $\rho \mapsto \delta_\rho(A,B)$.

Lemma 3.11. *Let $A := (A_1,\ldots,A_n) \in B(\mathcal{H})^n$ and $B := (B_1,\ldots,B_n) \in B(\mathcal{H})^n$ be in the class $\mathcal{C}_\rho$ and let $c > 0$ and $0 < \rho \leq \rho'$. Then the following statements hold.*

(i) *if $A \underset{c}{\overset{H}{\prec}} B$ in $\mathcal{C}_\rho$, then $A \underset{c}{\overset{H}{\prec}} B$ in $\mathcal{C}_{\rho'}$;*

(ii) *if $A \underset{c}{\overset{H}{\sim}} B$ in $\mathcal{C}_\rho$, then if $A \underset{c}{\overset{H}{\sim}} B$ in $\mathcal{C}_\rho$ and*

$$\delta_{\rho'}(A,B) \leq \delta_\rho(A,B).$$

Proof. First recall that $\mathcal{C}_\rho \subseteq \mathcal{C}_{\rho'}$. If $A \underset{c}{\overset{H}{\prec}} B$ in $\mathcal{C}_\rho$, then

$$\Re p(A_1,\ldots,A_n) + (\rho-1)\Re p(0) \leq c^2 \left[\Re p(B_1,\ldots,B_n) + (\rho-1)\Re p(0)\right]$$

for any noncommutative polynomial with matrix-valued coefficients $p \in \mathbb{C}[X_1,\ldots,X_n] \otimes M_m$, $m \in \mathbb{N}$, such that $\Re p(X) \geq 0$ for any $X \in [B(\mathcal{H})^n]_1$. Hence, $c \geq 1$ and, consequently, the inequality above holds when we replace ρ with $\rho' \geq \rho$. This shows that $A \underset{c}{\overset{H}{\prec}} B$ in $\mathcal{C}_{\rho'}$. Part (ii) is a clear consequence of (i) and the definition of the hyperbolic metric. □

If $A := (A_1,\ldots,A_n) \in B(\mathcal{H})^n$ is a nonzero n-tuple of operators such that $A \in [C_\infty]_{<1}$, i.e., the joint spectral radius $r(A) < 1$, then

$$\rho_A := \inf\{\rho > 0 : \; A \in \mathcal{C}_\rho\} > 0.$$

Indeed, if $\rho, \rho' \in (0, \infty]$, $\rho \le \rho'$, then $\mathcal{C}_\rho \subseteq \mathcal{C}_{\rho'}$ and, moreover, we have

$$\omega_{\rho'}(A) \le \omega_\rho(A), \quad r(A) = \lim_{\rho\to\infty} \omega_\rho(A), \qquad A \in B(\mathcal{H})^n.$$

Consequently, there exists $\rho > 0$ such that $\omega_{\rho'}(A) < 1$, for any $\rho' \ge \rho$. Assume now that $\rho_A = 0$. Then $T \in \mathcal{C}_\rho$, i.e., $\omega_\rho(A) \le 1$ for any $\rho > 0$. On the other hand, we know that $\|A\| \le \rho\omega_\rho(A)$. Taking $\rho \to 0$, we deduce that $A = 0$, which is a contradiction. This proves our assertion.

Note that if $A, B \in [\mathcal{C}_\infty]_{<1}$, then

$$\rho_{A,B} := \inf\{\rho > 0 : \ A, B \in \mathcal{C}_\rho\} = \max\{\rho_A, \rho_B\}.$$

PROPOSITION 3.12. *If $A, B \in [\mathcal{C}_\infty]_{<1}$, then the map*

$$[\rho_{A,B}, \infty) \ni \rho \mapsto \delta_\rho(A, B) \in \mathbb{R}^+$$

is continuous, decreasing, and

$$\lim_{\rho\to\infty} \delta_\rho(A, B) = 0.$$

Proof. Using Theorem 3.5 and Lemma 3.11, one can easily deduce that the map $\rho \mapsto \delta_\rho(A, B)$ is continuous and decreasing. To prove the last part of the proposition, note that since $\delta_\rho(A, B) \le \delta_\rho(A, 0) + \delta_\rho(0, B)$, it is enough to show that $\lim_{\rho\to\infty} \delta_\rho(A, 0) = 0$. To this end, note that Theorem 3.5, implies

$$\delta_\rho(A, 0) = \ln \max\left\{\left\|C_{\rho,A}C_{\rho,0}^{-1}\right\|, \left\|C_{\rho,0}C_{\rho,A}^{-1}\right\|\right\}, \tag{3.7}$$

where

$$C_{\rho,A}C_{\rho,0}^{-1} = \frac{1}{\sqrt{\rho}}\left[\rho I + (1-\rho)(R_A^* + R_A) + (\rho - 2)R_A^*R_A\right]^{1/2}(I - R_A)^{-1}.$$

Hence, we deduce that

$$\begin{aligned}
\lim_{\rho\to\infty} \|C_{\rho,A}C_{\rho,0}^{-1}\| &= \left\|\left[I - (R_A^* + R_A) + R_A^*R_A\right]^{1/2}(I - R_A)^{-1}\right\| \\
&= \left\|(I - R_A^*)^{-1}\left[I - (R_A^* + R_A) + R_A^*R_A\right](I - R_A)^{-1}\right\| \\
&= \left\|(I - R_A^*)^{-1}(I - R_A^*)(I - R_A)(I - R_A)^{-1}\right\| \\
&= 1
\end{aligned}$$

Similarly, we have $\lim_{\rho\to\infty} \|C_{\rho,0}C_{\rho,A}^{-1}\| = 1$. Using now relation (3.7), we complete the proof. $\square$

4. Mapping theorems for free holomorphic functions on noncommutative balls

In this section, we provide mapping theorems, spectral von Neumann inequalities, and Schwarz type results for free holomorphic functions on noncommutative balls, with respect to the hyperbolic metric and the operator radius ω_ρ, $\rho \in (0, \infty]$.

First, we prove the following mapping theorem for the classes $\mathcal{C}_\rho$, $\rho > 0$.

Theorem 4.1. *Let $f := (f_1, \ldots, f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. If $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ is of class $\mathcal{C}_\rho$, $\rho > 0$, then $f(T_1, \ldots, T_n)$ is of class $\mathcal{C}_{\rho_f}$, where*

$$\rho_f := \begin{cases} 1 + (\rho - 1)\frac{1-\|f(0)\|}{1+\|f(0)\|} & \text{if } \rho < 1 \\ \\ 1 + (\rho - 1)\frac{1+\|f(0)\|}{1-\|f(0)\|} & \text{if } \rho \geq 1. \end{cases} \tag{4.1}$$

Proof. Let $p \in \mathbb{C}[Z_1, \ldots, Z_m] \otimes M_k$, $k \in \mathbb{N}$, be such that $\Re p \geq 0$ on the unit ball $[B(\mathcal{H})^m]_1$. This is equivalent to $\Re p(S_1', \ldots S_m') \geq 0$, where $S_1', \ldots, S_m'$ are the left creation operators on the full Fock space $F^2(H_m)$. Applying the noncommutative Poisson transform $P_{f(X_1,\ldots,X_n)} \otimes \mathrm{id}$, which is a completely positive linear map, to the inequality $\Re p(S_1', \ldots S_m') \geq 0$, we obtain

$$\Re p(f(X_1, \ldots, X_n)) \geq 0, \qquad X \in [B(\mathcal{H})^n]_1.$$

Moreover, since the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$, we deduce that the boundary function of the composition $p \circ f$ is $p(\widetilde{f}_1, \ldots, \widetilde{f}_m) \in \mathcal{A}_n \bar{\otimes}_{min} M_k$.

Assume that $(T_1, \ldots, T_n) \in \mathcal{C}_\rho$. Using the free pluriharmonic functional calculus of Theorem 1.3 and Theorem 1.1, we deduce that

$$\Re(p \circ f)(T_1, \ldots, T_n) + (\rho - 1)\Re(p \circ f)(0) \geq 0. \tag{4.2}$$

On the other hand, according to the Harnack type inequality of Theorem 1.4 from [49] applied to the positive free pluriharmonic function $\Re p$ at the point $f(0) = (f_1(0), \ldots, f_m(0))$, we have

$$\Re p(0)\frac{1 - \|f(0)\|}{1 + \|f(0)\|} \leq \Re p(f(0)) \leq \Re p(0)\frac{1 + \|f(0)\|}{1 - \|f(0)\|}. \tag{4.3}$$

Let $\gamma > 0$ and note that

$$\Re p(f(T_1, \ldots, T_n)) + (\gamma - 1)\Re p(0) = A + B, \tag{4.4}$$

where

$$\begin{aligned} A &:= \Re p(f(T_1, \ldots, T_n)) + (\rho - 1)p(f(0)) \\ B &:= (\gamma - 1)\Re p(0) - (\rho - 1)p(f(0)). \end{aligned} \tag{4.5}$$

Assume now that $\rho \geq 1$. Using the second inequality in (4.3), we obtain

$$\begin{aligned} B &\geq (\gamma-1)\Re p(0) - (\rho-1)\Re p(0)\frac{1+\|f(0)\|}{1-\|f(0)\|} \\ &= \left[(\gamma-1) - (\rho-1)\frac{1+\|f(0)\|}{1-\|f(0)\|}\right]\Re p(0), \end{aligned}$$

which is positive if $\gamma \geq 1+(\rho-1)\frac{1+\|f(0)\|}{1-\|f(0)\|}$. In this case, using relation (4.4) and (4.2), we obtain

$$\Re p(f(T_1,\dots,T_n)) + (\gamma-1)\Re p(0) \geq 0$$

for any $p \in \mathbb{C}[Z_1,\dots,Z_n]\otimes M_k$, $k\in\mathbb{N}$, be such that $\Re p \geq 0$ on the unit ball $[B(\mathcal{H})^m]_1$. Applying Theorem 1.1, we deduce that $f(T_1,\dots,T_n)\in\mathcal{C}_\gamma$. In particular, we have $f(T_1,\dots,T_n)\in\mathcal{C}_{\delta_f}$ where

$$\delta_f := 1+(\rho-1)\frac{1+\|f(0)\|}{1-\|f(0)\|}.$$

Now, we consider the case $\rho\in(0,1)$. Using the first inequality in (4.3), we obtain

$$B \geq \left[(\gamma-1) - (\rho-1)\frac{1-\|f(0)\|}{1+\|f(0)\|}\right]\Re p(0),$$

which is positive if $\gamma \geq 1+(\rho-1)\frac{1-\|f(0)\|}{1+\|f(0)\|}$. As above, using relations (4.4) and (4.2), we obtain

$$\Re p(f(T_1,\dots,T_n)) + (\gamma-1)\Re p(0) \geq 0$$

for any $p \in \mathbb{C}[Z_1,\dots,Z_n]\otimes M_k$, $k\in\mathbb{N}$, be such that $\Re p \geq 0$ on the unit ball $[B(\mathcal{H})^m]_1$. Theorem 1.1 implies $f(T_1,\dots,T_n)\in\mathcal{C}_\gamma$. In particular, we have $f(T_1,\dots,T_n)\in\mathcal{C}_{\delta_f}$ where

$$\delta_f := 1+(\rho-1)\frac{1-\|f(0)\|}{1+\|f(0)\|}.$$

The proof is complete. □

Note that under the conditions of Theorem 4.1, $\rho\leq\rho_f$ and $\rho=1 \implies \rho_f=1$. Moreover, if $\rho\neq 1$, then $\rho_f=\rho$ if and only if $f(0)=0$. On can also show that $\rho_f\leq 1$ if $\rho\leq 1$.

We remark that, under the conditions of Theorem 4.1, there exists $T := (T_1,\dots,T_n)\in B(\mathcal{H})^n$ such that if $\rho>0$ is the smallest positive number such that $(T_1,\dots,T_n)\in\mathcal{C}_\rho$, then there exists a free holomorphic function f such that ρ_f is the smallest positive number with the property that $f(T_1,\dots,T_n)\in\mathcal{C}_{\rho_f}$. Indeed, if $n\leq m$, take $f(X_1,\dots,X_n)=(X_1,\dots,X_n,0,\dots,0)$ and use Corollary 2.4. When $n>m$, take $f(X_1,\dots,X_n)=(X_1,\dots,X_m)$ and $T:=(T_1,\dots,T_n,0,\dots,0)$ with $(T_1,\dots,T_n)\in\mathcal{C}_\rho$.

Corollary 4.2. *Let $f := (f_1, \dots, f_m)$ be a bounded free holomorphic function with $\|f(0)\| < \|f\|_\infty$ such that the boundary functions $\widetilde{f}_1, \dots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. If $(T_1, \dots, T_n) \in B(\mathcal{H})^n$ is of class $\mathcal{C}_\rho$, $\rho > 0$, then*

$$\omega_{\rho_f}(f(T_1, \dots, T_n)) \le \|f\|_\infty,$$

where ρ_f is given by relation (4.1). *In particular, if $f(0) = 0$ and $(T_1, \dots, T_n) \in \mathcal{C}_\rho$, then*

$$\omega_\rho(f(T_1, \dots, T_n)) \le \|f\|_\infty.$$

Proof. Applying Theorem 4.1 the function $\frac{1}{\|f\|_\infty} f$, we deduce that $\frac{1}{\|f\|_\infty} f(T_1, \dots, T_n)$ is in the class $\mathcal{C}_{\rho_f}$, which is equivalent to $\omega_{\rho_f}\left(\frac{1}{\|f\|_\infty} f(T_1, \dots, T_n)\right) \le 1$, and the first inequality of the theorem follows. Hence, and using the fact that $\rho_f = \rho$ when $f(0) = 0$, we complete the proof. □

A simple consequence of Corollary 4.2 is the following power inequality.

Corollary 4.3. *If $(T_1, \dots, T_n) \in B(\mathcal{H})^n$ is non-zero, $\rho \in (0, \infty)$, and $k \ge 1$, then*

$$\omega_\rho(T_\alpha : \ \alpha \in \mathbb{F}_n^+, |\alpha| = k) \le \omega_\rho(T_1, \dots, T_n).$$

Proof. Since $\|(T_1, \dots, T_n)\| \le \rho\omega_\rho(T_1, \dots, T_n)$, we have $\omega_\rho(T_1, \dots, T_n) \ne 0$. Applying the second part of Corollary 4.2 to the n-tuple of operators $\left(\frac{1}{\omega_\rho(T_1,\dots,T_n)} T_1, \dots, \frac{1}{\omega_\rho(T_1,\dots,T_n)} T_n\right) \in \mathcal{C}_\rho$ and to the free holomorphic function

$$f(X_1, \dots, X_n) := (X_\alpha : \ \alpha \in \mathbb{F}_n^+, |\alpha| = k), \qquad (X_1, \dots, X_n) \in [B(\mathcal{H})^n]_1,$$

we complete the proof. □

Theorem 4.4. *Let $f := (f_1, \dots, f_m)$ be a bounded free holomorphic function with $\|f(0)\| < \|f\|_\infty$ such that the boundary functions $\widetilde{f}_1, \dots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Then, for each $r \in [0, 1)$,*

$$\sup_{T \in \mathcal{C}_\rho, \ \omega_\rho(T) \le r} \omega_{\rho_f}(f(T_1, \dots, T_n)) \le \|f(rS_1, \dots, rS_n)\|,$$

where $S_1, \dots, S_n$ are the left creation operators.

Proof. Consider the free holomorphic function f_r, defined by

$$f_r(X_1, \dots, X_n) := f(rX_1, \dots, rX_n), \qquad (X_1, \dots, X_n) \in [B(\mathcal{H})^n]_1$$

and recall that $\|f_r\|_\infty = \|f(rS_1, \dots, rS_n)\|$. Applying Corollary 4.2 to f_r, we have

$$\omega_{\rho_{f_r}}(f_r(A_1, \dots, A_n)) \le \|f_r\|_\infty, \quad (A_1, \dots, A_n) \in \mathcal{C}_\rho \tag{4.6}$$

Since $f(0) = f_r(0)$, we have $\rho_f = \rho_{f_r}$. Consequently, if we assume that $\omega_\rho(T_1, \dots, T_n) \le r < 1$, then $\left(\frac{1}{r}T_1, \dots, \frac{1}{r}T_n\right) \in \mathcal{C}_\rho$ and inequality (4.6) implies

$$\omega_{\rho_f}(f(T_1, \dots, T_n)) = \omega_{\rho_f}\left(f_r\left(\frac{1}{r}T_1, \dots, \frac{1}{r}T_n\right)\right) \le \|f_r\|_\infty,$$

which completes the proof. □

Corollary 4.5. *Let $(T_1, \dots, T_n) \in B(\mathcal{H})^n$ be such that $\omega_\rho(T_1, \dots, T_n) < 1$, and let $f := (f_1, \dots, f_m)$ be a bounded free holomorphic function with the following properties:*

(i) *the boundary functions $\widetilde{f}_1, \dots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$.*
(ii) *f_j has the standard representation of the form*

$$f_j(X_1, \dots, X_n) = \sum_{|\alpha| \ge k} a_\alpha^{(j)} X_\alpha, \qquad j = 1, \dots, m.$$

Then

$$\omega_\rho(f(T_1, \dots, T_n)) \le \omega_\rho(T_1, \dots, T_n)^k \|f\|_\infty.$$

Proof. Consider the free holomorphic function $g := \frac{1}{\|f\|_\infty} f$. Note that $\|g\|_\infty = 1$ and $g(0) = 0$. According to the Schwarz lemma for free holomorphic functions (see Theorem 2.4 from [44]), we have

$$\|g(X_1, \dots, X_n)\| \le \|(X_1, \dots, X_n)\|^k, \qquad (X_1, \dots, X_n) \in [B(\mathcal{H})^n]_1.$$

Denote $r := \omega_\rho(T_1, \dots, T_n) < 1$, $\rho > 0$, and consider

$$g_r(X_1, \dots, X_n) := g(rX_1, \dots, rX_n), \qquad (X_1, \dots, X_n) \in [B(\mathcal{H})^n]_1.$$

Note that the inequality above implies $\|g_r\|_\infty \le r^k$. Applying now Theorem 4.4 to g, and using the latter inequality, we obtain

$$\omega_\rho(g(T_1, \dots, T_n)) \le \|g_r\|_\infty \le r^k = \omega_\rho(T_1, \dots, T_n)^k.$$

Hence, the result follows. □

Corollary 4.6. *Let $(T_1, \dots, T_n) \in B(\mathcal{H})^n$ be such that $\omega_\rho(T_1, \dots, T_n) < 1$, and let $f : [B(\mathcal{H})^n]_1 \to B(\mathcal{H})$ be a free holomorphic function with $\Re f \le I$ and having the standard representation*

$$f(X_1, \dots, X_n) = \sum_{|\alpha| \ge k} a_\alpha X_\alpha, \qquad \text{where} \quad k \ge 1.$$

Then

$$\omega_\rho(f(T_1, \dots, T_n)) \le \frac{2\omega_\rho(T_1, \dots, T_n)^k}{1 - \omega_\rho(T_1, \dots, T_n)^k}.$$

Proof. According to the Carathéodory type result for free holomorhic functions (see Theorem 5.1 from [51]), we have

$$\|f(X_1,\ldots,X_n)\| \leq \frac{2\|\sum_{|\beta|=k} X_\beta X_\beta^*\|^{1/2}}{1-\|\sum_{|\beta|=k} X_\beta X_\beta^*\|^{1/2}}, \qquad (X_1,\ldots,X_n)\in [B(\mathcal{H})^n]_1.$$

Hence, we deduce that $\|f_r\|_\infty \leq \frac{2r^k}{1-r^k}$ for any $r\in(0,1)$. Setting $r := \omega_\rho(T_1,\ldots,T_n) < 1$, $\rho > 0$, and applying Theorem 4.4, we obtain

$$\omega_\rho(f(T_1,\ldots,T_n)) \leq \|f_r\|_\infty \leq \frac{2\omega_\rho(T_1,\ldots,T_n)^k}{1-\omega_\rho(T_1,\ldots,T_n)^k}.$$

The proof is complete. □

Lemma 4.7. *Let $f := (f_1,\ldots,f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1,\ldots,\widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Let $A := (A_1,\ldots,A_n) \in B(\mathcal{H})^n$ and $B := (B_1,\ldots,B_n) \in B(\mathcal{H})^n$ be in the class $\mathcal{C}_\rho \subset B(\mathcal{H})^n$ and let $c \geq 1$. If $A \overset{H}{\underset{c}{\prec}} B$, then $f(A)$ and $f(B)$ are in $\mathcal{C}_{\rho_f} \subset B(\mathcal{H})^m$ and $f(A) \overset{H}{\underset{c}{\prec}} f(B)$, where ρ_f is given by relation* (4.1).

Proof. First, note that, due to Theorem 4.1, $f(A), f(B)$ are in $\mathcal{C}_{\rho_f}$, where ρ_f is given by relation (4.1). Let $p \in \mathbb{C}[Z_1,\ldots,Z_m]\otimes M_k$, $k\in\mathbb{N}$, be such that $\Re p \geq 0$ on the unit ball $[B(\mathcal{H})^m]_1$. According to the proof of Theorem 4.1, the boundary function of the composition $p\circ f$ is $p(\widetilde{f}_1,\ldots,\widetilde{f}_m) \in \mathcal{A}_n \bar{\otimes}_{min} M_k$ and $\Re(p\circ f) \geq 0$. Using the free pluriharmonic functional calculus for the class $\mathcal{C}_\rho$ and Theorem 2.2, if A, B are in $\mathcal{C}_\rho$ and $A \overset{H}{\underset{c}{\prec}} B$, $c \geq 1$, then

$$\begin{aligned}\Re(p\circ f)(A_1,\ldots,A_n) &+ (\rho-1)\Re(p\circ f)(0)\\ &\leq c^2\left[\Re(p\circ f)(B_1,\ldots,B_n) + (\rho-1)\Re(p\circ f)(0)\right].\end{aligned} \tag{4.7}$$

Assume now that $\rho \geq 1$. Due to the Harnack type inequality (4.3), the inequality (4.7) implies

$$\Re(p\circ f)(A_1,\ldots,A_n) \leq c^2\Re(p\circ f)(B_1,\ldots,B_n) + (c^2-1)(\rho-1)\Re p(0)\frac{1+\|f(0)\|}{1-\|f(0)\|},$$

which is equivalent to

$$\begin{aligned}\Re(p\circ f)(A_1,\ldots,A_n) &+ (\rho_f-1)\Re(p\circ f)(0)\\ &\leq c^2\left[\Re(p\circ f)(B_1,\ldots,B_n) + (\rho_f-1)\Re(p\circ f)(0)\right],\end{aligned}$$

where $\delta_f := 1 + (\rho-1)\frac{1+\|f(0)\|}{1-\|f(0)\|}$. Applying Theorem 2.2, we deduce that $f(A) \overset{H}{\underset{c}{\prec}} f(B)$.

Now, we consider the case $\rho \in (0,1)$. The inequality (4.7) and the Harnack type inequality (4.3) imply

$$\Re(p\circ f)(A_1,\dots,A_n) \le c^2\Re(p\circ f)(B_1,\dots,B_n)+(c^2-1)(\rho-1)\Re p(0)\frac{1-\|f(0)\|}{1+\|f(0)\|}.$$

As above, we deduce that $f(A)\overset{H}{\underset{c}{\prec}} f(B)$ in $\mathcal{C}_{\rho_f}$, where $\delta_f := 1+(\rho-1)\frac{1-\|f(0)\|}{1+\|f(0)\|}$. This completes the proof. ☐

Theorem 4.8. *Let $\delta_\rho : \Delta\times\Delta \to [0,\infty)$ be the hyperbolic metric on a Harnack part Δ of $\mathcal{C}_\rho$, and let $f := (f_1,\dots,f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1,\dots,\widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Then*

$$\delta_{\rho_f}(f(A), f(B)) \le \delta_\rho(A,B), \qquad A,B\in\Delta,$$

where ρ_f is given by relation (4.1).

Proof. Let $A,B\in\Delta\subset\mathcal{C}_\rho$, i.e., there is $c\ge 1$ such that $A\overset{H}{\underset{c}{\sim}}B$. According to Theorem 4.1 and Lemma 4.7, $f(A)$ and $f(B)$ are in $\mathcal{C}_{\rho_f}$, and $f(A)\overset{H}{\underset{c}{\sim}}f(B)$ in $\mathcal{C}_{\rho_f}$, where ρ_f is given by relation (4.1). Hence and taking into account that

$$\delta_\rho(A,B) := \ln\inf\left\{c>1:\ A\overset{H}{\underset{c}{\sim}}B\right\}, \qquad A,B\in\Delta,$$

we deduce that

$$\delta_{\rho_f}(f(A), f(B)) \le \delta_\rho(A,B), \qquad A,B\in\Delta.$$

The proof is complete. ☐

Now, we can deduce the following Schwarz type result.

Corollary 4.9. *Let $\delta_\rho : \Delta\times\Delta \to [0,\infty)$ be the hyperbolic metric on a Harnack part Δ of $\mathcal{C}_\rho$, and let $f := (f_1,\dots,f_m)$ be a contractive free holomorphic function with $f(0)=0$ such that the boundary functions $\widetilde{f}_1,\dots,\widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Then*

$$\delta_\rho(f(A), f(B)) \le \delta_\rho(A,B), \qquad A,B\in\Delta.$$

We recall that, due to Theorem 3.2, the open ball $[\mathcal{C}_\rho]_{<1}$ is the Harnack part of $\mathcal{C}_\rho$ containing 0. Consequently, Theorem 4.8 and Corollary 4.9 hold in the particular case when $\Delta := [\mathcal{C}_\rho]_{<1}$.

Ky Fan [14] showed that the von Neumann inequality [57] is equivalent to the fact that if $T\in B(\mathcal{H})$ is a strict contraction ($\|T\|<1$) and $f:\mathbb{D}\to\mathbb{D}$ is an analytic function, then $\|f(T)\|<1$. A multivariable analogue of this result was obtained in [51]. In what follows, we provide a spectral version of this result, when the norm is replaced by the joint spectral radius.

Theorem 4.10. *Let $f := (f_1, \ldots, f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. If $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ and the joint spectral radius $r(T_1, \ldots, T_n) < 1$, then*

$$r(f(T_1, \ldots, T_n)) < 1.$$

Proof. Assume that $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ has the joint spectral radius $r(T_1, \ldots, T_n) < 1$. Taking into account that $r(T_1, \ldots, T_n) = \lim_{\rho\to\infty} \omega_\rho(T_1, \ldots, T_n)$, we find $\delta > 1$ such that $\omega_\rho(T_1, \ldots, T_n) < 1$. Therefore, we have $T := (T_1, \ldots, T_n) \in \mathcal{C}_\rho$ and, due to Theorem 3.2, the n-tuple T is Harnack equivalent to 0. Consequently, $T \overset{H}{\underset{c}{\prec}} 0$ for some constant $c \geq 1$. According to Theorem 4.1, $f(T)$ and $f(0)$ are in the class $\mathcal{C}_{\rho_f}$, where ρ_f is given by relation (4.1). On the other hand, Lemma 4.7 implies $f(T) \overset{H}{\underset{c}{\prec}} f(0)$ in $\mathcal{C}_{\rho_f}$. Since $\|f(0)\| < 1$, we have the joint spectral radius $r(f(0)) < 1$. Applying Theorem 2.5, we deduce that $f(0) \overset{H}{\underset{c}{\prec}} 0$ in $\mathcal{C}_{\rho_f}$. Therefore, we have $f(T) \overset{H}{\underset{c}{\prec}} 0$ in $\mathcal{C}_{\rho_f}$. Applying again Theorem 2.5, we have $r(f(T)) < 1$. The proof is complete. $\square$

An analogue of Theorem 4.10 for n-tuples of operators with joint operator radius $\omega_\rho(T_1, \ldots, T_n) < 1$ is the following.

Theorem 4.11. *Let $f := (f_1, \ldots, f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. If $(T_1, \ldots, T_n) \in B(\mathcal{H})^n$ and $\omega_\rho(T_1, \ldots, T_n) < 1$, then*

$$\omega_{\rho_f}(f(T_1, \ldots, T_n)) < 1,$$

where ρ_f is defined by relation (4.1). In particular, if $f(0) = 0$, then $\omega_\rho(f(T_1, \ldots, T_n)) < 1$.

Proof. If $T := (T_1, \ldots, T_n) \in B(\mathcal{H})^n$ and $\omega_\rho(T_1, \ldots, T_n) < 1$, then $T \in \mathcal{C}_\rho$. According to Theorem 3.2, we have

$$r(T_1, \ldots, T_n) < 1 \quad \text{and} \quad P_\rho(T, R) \geq aI$$

for some constant $a > 0$. Applying Theorem 4.1 and Theorem 4.10, we deduce that $f(T) \in \mathcal{C}_{\rho_f}$ and $r(f(T)) < 1$. Since $\omega_\rho(T) < 1$, Theorem 3.2 implies $T \overset{H}{\sim} 0$. In particular, we have $0 \overset{H}{\underset{c}{\prec}} T$ for some constant $c \geq 1$. Applying Lemma 4.7, we deduce that $f(0) \overset{H}{\underset{c}{\prec}} f(T)$ in $\mathcal{C}_{\rho_f}$, where ρ_f is given by relation (4.1). Hence, and using Theorem 2.2 (part (ii)), we get

$$P_{\rho_f}(rf(0), R) \leq c^2 P_{\rho_f}(rf(T), R), \qquad r \in [0, 1).$$

Since $r(f(0)) < 1$ and $r(f(T)) < 1$, the latter inequality implies

$$P_{\rho_f}(f(0), R) \leq c^2 P_{\rho_f}(f(T), R), \qquad r \in [0, 1). \tag{4.8}$$

On the other hand, since the mapping $X \mapsto P_1(X,R)$ is a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$, the Harnack inequality (3.1) implies

$$P_1(f(0),R) \geq P_1(0,R)\frac{1-\|f(0)\|}{1+\|f(0)\|} = \frac{1-\|f(0)\|}{1+\|f(0)\|}I.$$

Therefore, we have

$$\begin{aligned} P_{\rho_f}(f(0),R) &= P_1(f(0),R) + (\rho_f - 1)I \\ &\geq \left(\rho_f - 1 + \frac{1-\|f(0)\|}{1+\|f(0)\|}\right) I. \end{aligned}$$

Since

$$a := \rho_f - 1 + \frac{1-\|f(0)\|}{1+\|f(0)\|} = \begin{cases} \rho\frac{1-\|f(0)\|}{1+\|f(0)\|} & \text{if } \ \rho < 1 \\ \\ (\rho-1)\frac{1+\|f(0)\|}{1-\|f(0)\|} + \frac{1-\|f(0)\|}{1+\|f(0)\|} & \text{if } \ \rho \geq 1, \end{cases}$$

we have $a > 0$. Combining the latter inequality with (4.8) we obtain

$$P_{\rho_f}(f(T),R) \geq \frac{a}{c^2}I.$$

Using again Theorem 3.2, we deduce that $\omega_{\rho_f}(f(T)) < 1$. The last part of the theorem follows from Theorem 4.1. This completes the proof. $\square$

Remark 4.12. *If $m = 1$, all the results of this section remain true when the condition $\|f(0)\| < 1$ is dropped if f is a nonconstant contractive free holomorphic function with boundary function in the noncommutative algebra $\mathcal{A}_n$.*

5. Carathéodory metric on the open noncommutative ball $[\mathcal{C}_\infty]_{<1}$ and Lipschitz mappings

In this section, we introduce a Carathéodory type metric d_K on the open ball of all n-tuples of operators $(T_1,\ldots,T_n)$ with joint spectral radius $r(T_1,\ldots,T_n) < 1$. We obtain a concrete formula for d_K in terms of the free pluriharmonic kernel on the open unit ball $[\mathcal{C}_\infty]_{<1}$. This is used to prove that the metric d_K is complete on $[\mathcal{C}_\infty]_{<1}$ and its topology coincides with the operator norm topology.

We need some notation. Consider the noncommutative balls

$$[\mathcal{C}_\rho]_{<1} := \{(X_1,\ldots,X_n) \in B(\mathcal{H})^n : \ \omega_\rho(X_1,\ldots,X_n) < 1\} \qquad \text{for } \ \rho \in (0,\infty],$$

where $\omega_\infty(X_1,\ldots,X_n) := r(X_1,\ldots,X_n)$ is the joint spectral radius of $(X_1,\ldots,X_n)$, and set

$$[\mathcal{C}_\rho]^{\prec 0} := \mathcal{C}_\rho \cap [\mathcal{C}_\infty]_{<1} \qquad \text{for } \ \rho \in (0,\infty).$$

According to Theorem 1.35 from [48], if $\rho,\rho' \in (0,\infty]$, $\rho \leq \rho'$, then $\mathcal{C}_\rho \subseteq \mathcal{C}_{\rho'}$ and, moreover, we have

$$\omega_{\rho'}(X) \leq \omega_\rho(X), \quad r(X) = \lim_{\rho\to\infty}\omega_\rho(X), \qquad X \in B(\mathcal{H})^n.$$

Consequently, we have

$$[\mathcal{C}_\rho]^{\prec 0} \subseteq [\mathcal{C}_{\rho'}]^{\prec 0}, \qquad [\mathcal{C}_\rho]_{<1} \subseteq [\mathcal{C}_{\rho'}]_{<1}.$$

Due to Theorem 2.5 and Theorem 3.2, one can easily see that

$$\{X \in \mathcal{C}_\rho : X \overset{H}{\sim} 0\} = [\mathcal{C}_\rho]_{<1} \subset [\mathcal{C}_\rho]^{\prec 0} = \left\{X \in \mathcal{C}_\rho : \ X \overset{H}{\prec} 0\right\}$$

for any $\rho \in (0, \infty)$. Note also that

$$\bigcup_{\rho>0} [\mathcal{C}_\rho]_{<1} = \bigcup_{\rho>0} [\mathcal{C}_\rho]^{\prec 0} = [\mathcal{C}_\infty]_{<1}.$$

Indeed, if $X \in [\mathcal{C}_\infty]_{<1}$, i.e., $r(X) < 1$, then taking into account that $r(X) = \lim_{\rho\to\infty} \omega_\rho(X)$, we find $\rho > 0$ such that $\omega_\rho(X) < 1$. Thus $X \in [\mathcal{C}_\rho]_{<1}$, which proves our assertion. Note also that $\bigcup_{\rho>0} [\mathcal{C}_\rho]_{<1}$ is dense (in the norm topology) in the set $\mathcal{C}_\infty$ of all n-tuples of operators $(T_1, \ldots, T_n)$ with joint spectral radius $r(T_1, \ldots, T_n) \leq 1$.

Now, we introduce the map $d_K : [\mathcal{C}_\infty]_{<1} \times [\mathcal{C}_\infty]_{<1} \to [0, \infty)$ by setting

$$d_K(A, B) = \sup_p \|\Re p(A) - \Re p(B)\|, \qquad A, B \in [\mathcal{C}_\infty]_{<1}, \tag{5.1}$$

where the supremum is taken over all polynomials $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, with $\Re p(0) = I$ and $\Re p \geq 0$ on $[B(\mathcal{H})^n]_1$. In what follows we will prove that d_K is a metric and obtain a concrete formula in terms of the free pluriharmonic kernel on the open unit ball $[\mathcal{C}_\infty]_{<1}$.

First, we need the following result.

Lemma 5.1. *Let G be a free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$, such that $G(0) = I$ and $G \geq 0$. If $A, B \in [\mathcal{C}_\infty]_{<1}$, then*

$$\|G(A) - G(B)\| \leq \|P_1(A, R) - P_1(B, R)\|,$$

where where $P_1(X, R)$ is the free pluriharmonic Poisson kernel defined by

$$P_1(X, R) := \sum_{k=1}^\infty \sum_{|\alpha|=k} X_\alpha \otimes R_{\tilde{\alpha}}^* + I \otimes I + \sum_{k=1}^\infty \sum_{|\alpha|=k} X_\alpha^* \otimes R_{\tilde{\alpha}}, \qquad X \in [\mathcal{C}_\infty]_{<1},$$

and the convergence is in the operator norm topology.

Proof. Since G is a positive free pluriharmonic function of $[B(\mathcal{H})^n]_1$ it has a unique representation of the form

$$G(X_1, \ldots, X_n) = \sum_{k=1}^\infty \sum_{|\alpha|=k} X_\alpha^* \otimes A_{(\alpha)}^* + I \otimes I + \sum_{k=1}^\infty \sum_{|\alpha|=k} X_\alpha \otimes A_{(\alpha)}, \ X \in [B(\mathcal{H})^n]_1,$$

for some $A_{(\alpha)} \in B(\mathcal{E})$, where the series converge in the operator norm topology. Applying Theorem 5.2 from [47] to G, we find a completely positive linear map $\mu : \mathcal{R}_n^* + \mathcal{R}_n \to B(\mathcal{E})$ with $\mu(I) = I$ and $\mu(R_{\tilde{\alpha}}^*) = A_{(\alpha)}$ if $|\alpha| \geq 1$.

Since $A, B \in [\mathcal{C}_\rho]_{<1}$, we have $r(A) < 1$ and $r(B) < 1$. According to the free pluriharmonic functional calculus, $P_\rho(A, R)$, $P_\rho(B, R)$, $G(A)$, and $G(B)$

are well-defined and the corresponding series converge in the operator norm topology. Consequently, we have

$$G(A) = (\mathrm{id} \otimes \mu)(P_1(A,R)) \quad \text{and} \quad G(A) = (\mathrm{id} \otimes \mu)(P_1(A,R)).$$

Taking into account that μ is completely positive linear map with $\mu(I) = I$, we have

$$\|G(A) - G(B)\| \leq \|\mu\| \|P_1(A,R) - P_1(B,R)\| = \|P_1(A,R) - P_1(B,R)\|.$$

The proof is complete. □

According to Lemma 5.1, it makes sense to define the map $d_K' : [\mathcal{C}_\infty]_{<1} \times [\mathcal{C}_\infty]_{<1} \to [0, \infty)$ by setting

$$d_K'(A,B) := \sup_u \|u(A) - u(B)\| < \infty,$$

where the supremum is taken over all free pluriharmonic functions u on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$, such that $u(0) = I$ and $u \geq 0$.

Using the the free pluriharmonic functional calculus for for n-tuples of operators $(T_1, \ldots, T_n)$ with the joint spectral radius $r(T_1, \ldots, T_n) < 1$, one can extend Proposition 3.1 from [49] and show that for any $A, B \in [\mathcal{C}_\infty]_{<1}$,

$$d_K'(A,B) = d_K(A,B),$$

where d_K is defined by relation (5.1). Since the proof is essentially the same, we shall omit it.

PROPOSITION 5.2. *d_K is a metric on $[\mathcal{C}_\infty]_{<1}$ satisfying relation*

$$d_K(A,B) = \|P_1(A,R) - P_1(B,R)\|, \qquad A, B \in [\mathcal{C}_\infty]_{<1}.$$

In addition, the map $[0,1) \ni r \mapsto d_K(rA, rB) \in \mathbb{R}^+$ *is increasing and*

$$d_K(A,B) = \sup_{r \in [0,1)} d_K(rA, rB).$$

Proof. Using Lemma 5.1 we deduce that $d_K(A,B) \leq \|P_1(A,R) - P_1(B,R)\|$. The rest of the proof is similar to that of Proposition 3.2 from [49], so we shall omit it. □

Now, we can prove the main result of this section.

THEOREM 5.3. *Let d_K be the Carathéodory metric on $[\mathcal{C}_\infty]_{<1}$. Then the following statements hold:*

(i) *the d_K-topology coincides with the norm topology on $[\mathcal{C}_\infty]_{<1}$;*
(ii) *$[\mathcal{C}_\rho]^{\prec 0}$ is a d_K-closed subset of $[\mathcal{C}_\infty]_{<1}$ for any $\rho > 0$;*
(iii) *the metric d_K is complete on $[\mathcal{C}_\infty]_{<1}$.*

Proof. We recall that the free pluriharmonic Poisson kernel is given by

$$P_1(X,R)=\sum_{k=1}^\infty\sum_{|\alpha|=k} X_\alpha\otimes R_{\tilde\alpha}^* + I\otimes I+\sum_{k=1}^\infty\sum_{|\alpha|=k} X_\alpha^*\otimes R_{\tilde\alpha}, \qquad X\in[\mathcal{C}_\infty]_{<1},$$

where the convergence is in the operator norm topology. Let $R_A:=A_1^*\otimes R_1+\cdots+A_n^*\otimes R_n$ be the reconstruction operator. Note that, due to the noncommutative von Neumann inequality, we have

$$\begin{aligned}\|A-B\| &= \|R_A-R_B\|\\ &=\left\|\frac{1}{2\pi}\int_0^{2\pi} e^{it}[P_1(A,e^{it}R)-P_1(B,e^{it}R)]dt\right\|\\ &\le \sup_{t\in[0,2\pi]}\left\|P_1(A,e^{it}R)-P_1(B,e^{it}R)\right\|\\ &\le \|P_1(A,R)-P_1(B,R)\|.\end{aligned}$$

Now, Proposition 5.2 implies

$$\|A-B\|\le d_K(A,B),\qquad A,B\in[\mathcal{C}_\infty]_{<1}, \tag{5.2}$$

which shows that the d_K-topology is stronger then the norm topology on $[\mathcal{C}_\infty]_{<1}$. Conversely, to prove that the norm topology on $[\mathcal{C}_\infty]_{<1}$ is stronger than the d_K-topology, note that since $r(R_A)=r(A)<1$ and $r(R_B)=r(B)<1$, the operators $I-R_A$ and $I-R_B$ are invertible. Thus

$$d_K(A,B)=\|P_1(A,R)-P_1(B,R)\|\le 2\|(I-R_A)^{-1}-(I-R_B)^{-1}\|$$

for any $A,B\in[\mathcal{C}_\infty]_{<1}$. Hence and due to the continuity of the maps $X\mapsto I-R_X$ on $B(\mathcal{H})^n$ and $Y\mapsto Y^{-1}$ on the group of invertible elements in $B(\mathcal{H}\otimes F^2(H_n))$, in the operator norm topology, we deduce our assertion. In conclusion, the d_K-topology coincides with the norm topology on $[\mathcal{C}_\infty]_{<1}$.

Now, to prove (ii), let $\{A^{(k)}:=(A_1^{(k)},\ldots,A_n^{(k)})\}_{k=1}^\infty$ be a d_K-Cauchy sequence in $[\mathcal{C}_\rho]^{\prec 0}\subset\mathcal{C}_\rho$. Due to inequality (5.2), we deduce that $\{A^{(k)}\}_{k=1}^\infty$ is a Cauchy sequence in the norm topology of $B(\mathcal{H})^n$. Since $\mathcal{C}_\rho$ is closed in the operator norm topology, there exists $T:=(T_1,\ldots,T_n)$ in $\mathcal{C}_\rho$ such that $\|T-A^{(k)}\|\to 0$, as $k\to\infty$.

Now let us prove that the joint spectral radius $r(T)<1$. Since $\{A^{(k)}\}_{k=1}^\infty$ is a d_K-Cauchy sequence, there exists $k_0\in\mathbb{N}$ such that $d_K(A^{(k)},A^{(k_0)})\le 1$ for any $k\ge k_0$. On the other hand, since $A^{(k_0)}\in[\mathcal{C}_\rho]^{\prec 0}$, i.e., $A^{(k_0)}\overset{H}{\prec}0$, Theorem 2.2 shows that there is $c\ge 1$ such that $P_\rho(rA^{(k_0)},R)\le c^2\delta$ for any $r\in[0,1)$. Hence, and due to the noncommutative von Neumann inequality, we deduce that

$$\begin{aligned}P_\rho(rA^{(k)},R)&\le\Big(\|P_\rho(rA^{(k)},R)-P_\rho(rA^{(k_0)},R)\|+\|P_\rho(rA^{(k_0)},R)\|\Big)I\\ &\le\Big(d_K(A^{(k)},A^{(k_0)})+\|P_\rho(rA^{(k_0)},R)\|\Big)I\le(1+c^2\delta)I\end{aligned} \tag{5.3}$$

for any $k\ge k_0$ and $r\in[0,1)$.

We show now that $\lim_{k\to\infty} P_\rho(rA^{(k)}, R) = P_\rho(rT, R)$ in the operator norm topology. First, one can easily see that, since $T, A^{(k)} \in \mathcal{C}_\rho$, we have

$$\sum_{|\alpha|=p} T_\alpha T_\alpha^* \leq \rho^2 I \quad \text{and} \quad \sum_{|\alpha|=p} A_\alpha^{(k)} A_\alpha^{(k)} \leq \rho^2 I$$

for any $p, k = 1, 2, \ldots$. Given $\epsilon > 0$ and $r \in (0,1)$, let $m \in \mathbb{N}$ be such that $\sum_{p=m}^\infty \rho r^p < \frac{\epsilon}{2}$. Note that

$$\begin{aligned}
&\|P(rA^{(k)}, R) - P(rT, R)\| \\
&\leq 2 \sum_{p=1}^{m-1} \left\| \sum_{|\alpha|=p} r^{|\alpha|} (A_\alpha^{(k)} - T_\alpha) \otimes R_{\tilde{\alpha}}^* \right\| \\
&\qquad + 2 \sum_{p=m}^{\infty} \left\| \sum_{|\alpha|=p} r^{|\alpha|} A_\alpha^{(k)} \otimes R_{\tilde{\alpha}}^* \right\| + 2 \sum_{p=m}^{\infty} \left\| \sum_{|\alpha|=p} r^{|\alpha|} T_\alpha \otimes R_{\tilde{\alpha}}^* \right\| \\
&= 2 \left\| \sum_{p=1}^{m-1} r^p \sum_{|\alpha|=p} (A_\alpha^{(k)} - T_\alpha)(A_\alpha^{(k)} - T_\alpha)^* \right\| \\
&\qquad + 2 \left\| \sum_{p=1}^{m-1} r^p \sum_{|\alpha|=p} A_\alpha^{(k)} {A_\alpha^{(k)}}^* \right\| + 2 \left\| \sum_{p=1}^{m-1} r^p \sum_{|\alpha|=p} T_\alpha {T_\alpha}^* \right\| \\
&\leq 2 \left\| \sum_{p=1}^{m-1} r^p \sum_{|\alpha|=p} (A_\alpha^{(k)} - T_\alpha)(A_\alpha^{(k)} - T_\alpha)^* \right\| + \epsilon
\end{aligned}$$

for any $k = 1, 2, \ldots$. Since $A^{(k)} \to T$ in the norm topology, as $k \to \infty$, and using the results above, one can easily deduce that $\lim_{k\to\infty} P_\rho(rA^{(k)}, R) = P_\rho(rT, R)$ for each $r \in [0,1)$. Now, taking $k \to \infty$ in inequality (5.3), we obtain $P_\rho(rT, R) \leq (1 + c^2\delta)I$ for $r \in [0,1)$. Applying Theorem 2.2, we deduce that $T \overset{H}{\prec} 0$. Now, Theorem 2.5 implies $r(T) < 1$, which shows that T is in $[\mathcal{C}_\rho]^{\prec 0}$ and, therefore, in $[\mathcal{C}_\infty]_{<1}$, which proves part (ii).

It remains to prove part (iii). To this end, let $\{A^{(k)} := (A_1^{(k)}, \ldots, A_n^{(k)})\}_{k=1}^\infty$ be a d_K-Cauchy sequence in $[\mathcal{C}_\infty]_{<1}$. Given $\epsilon > 0$, there exists $k_0 \geq 1$ such that $d_K(A^{(k)}, A^{(j)}) < \epsilon$ for any $k, j \geq k_0$. Then we have

$$d_K(A^{(k)}, 0) \leq c := d_K(A^{(k_0)}, 0) + \epsilon \quad \text{for any} \quad k \geq k_0. \tag{5.4}$$

Hence, and due to the definition of d_K, we have $\|u(A^{(k)}) - u(0)\| \leq c$ and, consequently,

$$u(A^{(k)}) \leq (\|u(A^{(k)} - u(0)\| + 1)I \leq (c+1)u(0) \quad \text{for any} \quad k \geq k_0$$

and for any positive free pluriharmonic function u on $[B(\mathcal{H})^n]_1$ with coefficients in $B(\mathcal{E})$ such that $u(0) = I$.

Now, for each $k \geq k_0$, fix $\rho_k \geq 1$ such that $A^{(k)} \in [\mathcal{C}_{\rho_k}]^{\prec 0}$. Note that the inequality above implies

$$u(A^{(k)}) + (\rho_k - 1)u(0) \leq \rho_k(c+1)u(0)$$

for all $k \geq k_0$. Applying Theorem 2.2 and using relation (5.4), we obtain

$$\|L_{0,A^{(k)}}\|^2 \leq d_K(A^{(k_0)}, 0) + \epsilon + 1, \qquad k \geq k_0.$$

Consequently, we have

$$1 \leq \epsilon_0 := \sup_{k \geq k_0} \|L_{0,A^{(k)}}\|^2 < \infty. \tag{5.5}$$

Since $\{A^{(k)}\}$ is a d_K-Cauchy sequence, there exists $m_0 \geq k_0$ such that $d_K(A^{(m')}, A^{(m)}) < \frac{1}{2\epsilon_0}$ for any $m, m' \geq m_0$. Using now relation (5.5), we obtain

$$d_K(A^{(m)}, A^{(m_0)}) < \frac{1}{2\|L_{0,A^{(m_0)}}\|^2}, \qquad k \geq m_0. \tag{5.6}$$

Since $A^{(m_0)} \in [\mathcal{C}_{\rho_{m_0}}]^{\prec 0}$, Theorem 2.5 implies $r(A^{(m_0)}) < 1$. On the other hand, since $\lim_{\rho\to\infty} \omega_\rho(A^{(m_0)}) = r(A^{(m_0)}) < 1$, there exists $\rho_{m_0} > 0$ such that $\omega_{\rho_{m_0}}(A^{(m_0)}) < 1$ for any $\rho \geq \rho_{m_0}$. We can assume that

$$\rho_{m_0} \geq \frac{\|L_{A^{(m_0)},0}\|^2}{\|L_{0,A^{(m_0)}}\|^2}. \tag{5.7}$$

Using Proposition 5.2 and relation (5.6), we deduce that

$$P_{\rho_{m_0}}(A^{(m_0)}, R) \leq P_{\rho_{m_0}}(A^{(k)}, R) + \frac{1}{2\|L_{0,A^{(k)}}\|^2} I, \qquad k \geq m_0. \tag{5.8}$$

On the other hand, since $\omega_{\rho_{m_0}}(A^{(m_0)}) < 1$, Theorem 3.2 implies $A^{(m_0)} \overset{H}{\sim} 0$ in $\mathcal{C}_{\rho_{m_0}}$. Consequently, we have $0 \overset{H}{\prec} A^{(m_0)}$, which due to Theorem 2.2, implies

$$\rho_{m_0} I = P_{\rho_{m_0}}(0, R) \leq \|L_{A^{(m_0)},0}\|^2 P_{\rho_{m_0}}(A^{(m_0)}, R).$$

Combining this with relation (5.7), we get

$$P_{\rho_{m_0}}(A^{(m_0)}, R) \geq \frac{1}{\|L_{0,A^{(m_0)}}\|^2} I.$$

Hence, and due to (5.8), we have

$$P_{\rho_{m_0}}(A^{(k)}, R) \geq \frac{1}{2\|L_{0,A^{(m_0)}}\|^2} I \geq \frac{1}{2\epsilon_0} I.$$

Applying Theorem 3.2, we deduce that $A^{(k)} \overset{H}{\sim} 0$ and $A^{(k)} \in \mathcal{C}_{\rho_{m_0}}$. Therefore, $A^{(k)} \in [\mathcal{C}_{\rho_{m_0}}]^{\prec 0}$ for all $k \geq m_0$ and the sequence $\{A^{(k)}\}_{k \geq m_0}$ is a d_K-Cauchy sequence in $[\mathcal{C}_{\rho_{m_0}}]^{\prec 0}$. Due to part (ii), there exists $A \in [\mathcal{C}_{\rho_{m_0}}]^{\prec 0} \subset [\mathcal{C}_\infty]_{<1}$ such that $d_K(A^{(k)}, A) \to 0$, as $k \to \infty$, which proves that d_K is a complete metric on $[\mathcal{C}_\infty]_{<1}$. The proof is complete. $\square$

We can provide now a class of Lipschitz functions with respect to the Carathéodory metric on $[\mathcal{C}_\infty]_{<1}$.

Theorem 5.4. *Let $f := (f_1, \dots, f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1, \dots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Then*

$$d_K(f(A), f(B)) \leq \frac{1 + \|f(0)\|}{1 - \|f(0)\|} d_K(A, B)$$

for any n-tuples $A := (A_1, \dots, A_n)$ and $B := (B_1, \dots, B_n)$ in $[\mathcal{C}_\infty]_{<1}$.

Proof. According to the maximum principle for free holomorphic functions with operator-valued coefficients (see Proposition 5.2 from [50]), the condition $\|f(0)\| < 1$ implies that $\|f(X)\| < 1$, $X \in [B(\mathcal{H})^n]_1$. If u is a free pluriharmonic function on $[B(\mathcal{H})^m]_1$, then Theorem 1.1 from [51] shows that $u \circ f$ is a free pluriharmonic function on $[B(\mathcal{H})^n]_1$. If, in addition, u is positive, then $u \circ f$ is also positive.

Assume now that A and B are in $[\mathcal{C}_\infty]_{<1}$. Due to Theorem 4.10, $f(A)$ and $f(B)$ are in $[\mathcal{C}_\infty]_{<1}$. Let $p \in \mathbb{C}[X_1, \dots, X_m] \otimes M_k$, $k \in \mathbb{N}$, be a matrix-valued noncommutative polynomial with $\Re p(0) = I$ and $\Re p \geq 0$ on $[B(\mathcal{H})^m]_1$. According to the Harnack type inequality (4.3), we have

$$\frac{1 - \|f(0)\|}{1 + \|f(0)\|} I \leq \Re p(f(0)) \leq \frac{1 + \|f(0)\|}{1 - \|f(0)\|} I.$$

Since $\|f(0)\| < 1$, we deduce that $\Re p(f(0))$ is a positive invertible operator of the form $I_\mathcal{H} \otimes A$ for some $A \in M_k$. Define the mapping $h : [B(\mathcal{H})^n]_1 \to B(\mathcal{H}) \bar{\otimes}_{min} M_k$ by setting

$$h(X) := [\Re p(f(0))]^{-1/2} \Re p(f(X)) [\Re p(f(0))]^{-1/2}, \qquad X \in [B(\mathcal{H})^n]_1.$$

Note that h is a positive free pluriharmonic function on $[B(\mathcal{H})^n]_1$ with coefficients in M_k with the property that $h(0) = I$. Now, using the above-mentioned Harnack type inequality, we have

$$\begin{aligned}
&\|\Re p(f(A)) - \Re p(f(B))\| \\
&\quad \leq \|[\Re p(f(0))]^{1/2}\| \left\| [\Re p(f(0))]^{-1/2} \left(\Re p(f(A)) - \Re p(f(B))\right) [\Re p(f(0))]^{1/2} \right\| \\
&\qquad \cdot \|[\Re p(f(0))]^{1/2}\| \\
&\quad \leq \|[\Re p(f(0))]\| \|h(A) - h(B)\| \\
&\quad \frac{1 + \|f(0)\|}{1 - \|f(0)\|} d_K(A, B).
\end{aligned}$$

Taking the supremum over all polynomials $p \in \mathbb{C}[X_1, \dots, X_m] \otimes M_k$, $k \in \mathbb{N}$, with $\Re p(0) = I$ and $\Re p \geq 0$ on $[B(\mathcal{H})^m]_1$, we obtain

$$d_K(f(A), f(B)) \leq \frac{1 + \|f(0)\|}{1 - \|f(0)\|} d_K(A, B),$$

which completes the proof. □

Corollary 5.5. *Let $f := (f_1, \ldots, f_m)$ be a contractive free holomorphic function with $f(0) = 0$ such that the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Then*

$$d_K(f(A), f(B)) \le d_K(A, B)$$

for any $A, B \in [\mathcal{C}_\infty]_{<1}$.

We remark that, using Corollary 1.2 and the remarks preceding Corollary 2.4, one can easily obtain the following result, which provides a simple example when the inequality of Theorem 5.4 is an equality.

Corollary 5.6. *If $1 \le m < n$, let $A := (A_1, \ldots, A_m) \in B(\mathcal{H})^m$ and $B := (B_1, \ldots, B_m) \in B(\mathcal{H})^m$ be in $[\mathcal{C}_\infty]_{<1}$ and let $\widetilde{A} := (A_1, \ldots, A_m, 0, \ldots, 0)$ and $\widetilde{B} := (B_1, \ldots, B_m, 0, \ldots, 0)$ be their extensions in $B(\mathcal{H})^n$, respectively. Then*

$$d_K(A, B) = d_K(\widetilde{A}, \widetilde{B}).$$

According to Theorem 5.3, the d_K-topology coincides with the norm topology on $[\mathcal{C}_\infty]_{<1}$. Due to Theorem 5.4, we deduce the following result.

Corollary 5.7. *Let $f := (f_1, \ldots, f_m)$ be a contractive free holomorphic function with $\|f(0)\| < 1$ such that the boundary functions $\widetilde{f}_1, \ldots, \widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. Then the map*

$$[\mathcal{C}_\infty]_{<1} \ni (T_1, \ldots, T_n) \mapsto f(T_1, \ldots, T_n) \in [\mathcal{C}_\infty]_{<1}$$

is continuous in the operator norm topology, where $[\mathcal{C}_\infty]_{<1}$ is the corresponding ball in $B(\mathcal{H})^n$ and $B(\mathcal{H})^m$, respectively.

6. Three metric topologies on Harnack parts of $\mathcal{C}_\rho$

In this section we study the relation between the δ_ρ-topology, the d_K-topology, and the operator norm topology on Harnack parts of $\mathcal{C}_\rho$. We prove that the hyperbolic metric δ_ρ is a complete metric on certain Harnack parts of $\mathcal{C}_\rho$, and that all the three topologies coincide on $[\mathcal{C}_\rho]_{<1}$. In particular, we prove that the hyperbolic metric δ_ρ is complete on the open unit unit ball $[\mathcal{C}_\rho]_{<1}$, while the other two metrics are not complete.

First, we mention another formula for the hyperbolic distance that will be used to prove the main result of this section. If $f \in \mathcal{A}_n \bar{\otimes}_{min} M_m$, $m \in \mathbb{N}$, then we call $\Re f$ strictly positive and denote $\Re f > 0$ if there exists a constant $a > 0$ such that $\Re f \ge aI$. We remark that, in this case, if $(T_1, \ldots, T_n) \in \mathcal{C}_\rho$, then, using the functional calculus for the class $\mathcal{C}_\rho$, we deduce that

$$\Re f(T_1, \ldots, T_n) + (\rho - 1)\Re f(0) \ge \rho a I.$$

The proof of the next result is similar to that of Proposition 3.5 from [49], but uses the functional calculus for the class $\mathcal{C}_\rho$ and Theorem 2.2 of the present paper. We shall omit it.

PROPOSITION 6.1. *Let $A := (A_1, \ldots, A_n)$ and $B := (B_1, \ldots, B_n)$ be in $\mathcal{C}_\rho$ such that $A \overset{H}{\sim} B$. Then*

$$\delta_\rho(A,B) = \frac{1}{2}\sup\left|\ln\frac{\langle[\Re f(A_1,\ldots,A_n) + (\rho-1)\Re f(0)]x,x\rangle}{\langle[\Re f(B_1,\ldots,B_n) + (\rho-1)\Re f(0)]x,x\rangle}\right|, \tag{6.1}$$

where the supremum is taken over all $f \in \mathcal{A}_n \otimes M_m$, $m \in \mathbb{N}$, with $\Re f > 0$ and $x \in \mathcal{H} \otimes \mathbb{C}^m$ with $x \neq 0$.

We remark that, under the conditions of Proposition 6.1, one can also prove that relation (6.1) holds if the supremum is taken over all noncommutative polynomials $f \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, with $\Re f > 0$, and $x \in \mathcal{H} \otimes \mathbb{C}^m$ with $x \neq 0$.

The main result of this section is the following.

THEOREM 6.2. *Let δ_ρ, $\rho > 0$, be the hyperbolic metric on a Harnack part Δ of $[\mathcal{C}_\rho]^{\prec 0}$. Then the following properties hold:*

(i) *δ_ρ is complete on Δ;*
(ii) *the δ_ρ-topology is stronger then the d_K-topology on Δ;*
(iii) *the δ_ρ-topology, the d_K-topology, and the operator norm topology coincide on $[\mathcal{C}_\rho]_{<1}$;*
(iv) *$[\mathcal{C}_\rho]_{<1}$ is complete relative the hyperbolic metric, but not complete with respect to the Carathéodory metric d_K and the operator metric.*

Proof. Let $A := (A_1, \ldots, A_n)$ and $B := (B_1, \ldots, B_n)$ be n-tuples in a Harnack part Δ of $[\mathcal{C}_\rho]^{\prec 0}$. Then A is Harnack equivalent to B and

$$\Re f(A_1,\ldots,A_n) + (\rho-1)\Re f(0) \le \Lambda_\rho(A,B)^2[\Re f(B_1,\ldots,B_n) + (\rho-1)\Re f(0)]$$

for any $f \in \mathcal{A}_n \bar{\otimes}_{min} M_m$ with $\Re f \geq 0$, where $\Lambda_\rho(A,B)$ is defined by (3.3). Hence, we deduce that

$$\Re f(A_1,\ldots,A_n) - \Re f(B_1,\ldots,B_n) \le [\Lambda_\rho(A,B)^2 - 1][\Re f(B_1,\ldots,B_n) + (\rho-1)\Re f(0)]. \tag{6.2}$$

Since $B \overset{H}{\prec} 0$, we have the joint spectral radius $r(B) < 1$, so the ρ-pluriharmonic kernel $P_\rho(B,R)$ makes sense. Due to the fact that the noncommutative Poisson transform $\mathrm{id} \otimes P_{rR}$ is completely positive, and $P_\rho(B,S) \le \|P_\rho(B,R)\|I$, one can easily see that

$$\begin{aligned} P_\rho(rB,R) &= (\mathrm{id} \otimes P_{rR})[P_\rho(B,S)] \le \|P_\rho(B,R)\|I \\ &= \frac{1}{\rho}\|P_\rho(B,R)\|P_\rho(0,R) \end{aligned}$$

for any $r \in [0,1)$. Using the equivalence $(ii) \leftrightarrow (iii)$ of Theorem 2.2, when $c^2 = \frac{1}{\rho}\|P_\rho(B,R)\|$, we obtain $\Re f(rB_1,\ldots,rB_n) + (\rho-1)\Re f(0) \le \|P_\rho(B,R)\|\Re f(0)$ for any $r \in [0,1)$. Letting $r \to 1$, in the operator norm topology, we deduce that

$$\Re f(B_1,\ldots,B_n) + (\rho-1)\Re f(0) \le \|P_\rho(B,R)\|\Re f(0).$$

Hence, and using relation (6.2), we obtain

$$\Re f(A_1, \ldots, A_n) - \Re f(B_1, \ldots, B_n) \leq [\Lambda_\rho(A, B)^2 - 1] \|P_\rho(B, R)\| \Re f(0).$$

We can obtain a similar inequality if we interchange A with B. If, in addition, we assume that $\Re f(0) = I$, then we obtain

$$-tI \leq \Re f(A_1, \ldots, A_n) - \Re f(B_1, \ldots, B_n) \leq tI,$$

where $t := [\Lambda_\rho(A, B)^2 - 1] \max\{\|P_\rho(A, R)\|, \|P_\rho(B, R)\|\}$. On the other hand, since $\Re f(A_1, \ldots, A_n) - \Re f(B_1, \ldots, B_n)$ is a self-adjoint operator, we get $\|\Re f(A_1, \ldots, A_n) - \Re f(B_1, \ldots, B_n)\| \leq t$. Hence, we deduce that $d_K(A, B) \leq s$. As a consequence, we obtain

$$d_K(A, B) \leq \max\{\|P_\rho(A, R)\|, \|P_\rho(B, R)\|\} \left(e^{2\delta_\rho(A,B)} - 1\right). \tag{6.3}$$

Let us prove that δ_ρ is a complete metric on Δ. To this end, let $\{A^{(k)} := (A_1^{(k)}, \ldots, A_n^{(k)})\}_{k=1}^\infty \subset \Delta$ be a δ_ρ-Cauchy sequence. First, we prove that the sequence $\{\|P_\rho(A^{(k)}, R)\|\}_{k=1}^\infty$ is bounded. Given $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that

$$\delta_\rho(A^{(k)}, A^{(p)}) < \epsilon \quad \text{for any} \quad k, p \geq k_0. \tag{6.4}$$

Let $f \in \mathcal{A}_n \bar{\otimes}_{min} M_m$ with $\operatorname{Re} f \geq 0$. Since $A^{(k_0)} \overset{H}{\prec} 0$ and

$$P_\rho(rA^{(k_0)}, R) \leq \frac{1}{\rho} \|P_\rho(rA^{(k_0)}, R)\| P_\rho(0, R),$$

Theorem 2.2 implies

$$\Re f(A^{(k_0)}) + (\rho - 1)\Re f(0) \leq \frac{1}{\rho} \|P_\rho(rA^{(k_0)}, R)\| [\Re f(0) + (\rho - 1)\Re f(0)].$$

On the other hand, since $A^{(k)} \overset{H}{\sim} A^{(k_0)}$, Theorem 2.2 implies

$$\Re f(A^{(k)}) + (\rho - 1)\Re f(0) \leq \Lambda_\rho(A^{(k)}, A^{(k_0)})^2 [\Re f(A^{(k_0)}) + (\rho - 1)\Re f(0)].$$

Combining these inequalities, we obtain

$$\Re f(A^{(k)}) + (\rho - 1)\Re f(0) \leq c^2 \frac{1}{\rho} [\Re f(0) + (\rho - 1)\Re f(0)], \tag{6.5}$$

where $c := \|P_\rho(A^{(k_0)}, R)\|^{1/2} \Lambda_\rho(A^{(k)}, A^{(k_0)})$, for any $f \in \mathcal{A}_n \otimes M_m$ with $\Re f \geq 0$. Consequently, due to Theorem 2.2, we have $\|P_\rho(A^{(k)}, R)\| \leq c^2$ for any $k \geq k_0$. Combining this with relation (6.4), we obtain

$$\|P_\rho(A^{(k)}, R)\| \leq \|P_\rho(A^{(k_0)}, R)\| e^{2\epsilon}$$

for any $k \geq k_0$. This shows that the sequence $\{\|P_\rho(A^{(k)}, R)\|\}_{k=1}^\infty$ is bounded. Consequently, inequality (6.3) implies that $\{A^{(k)}\}$ is a d_K-Cauchy sequence. Due to Theorem 5.3, there exists $A := (A_1, \ldots, A_n) \in [\mathcal{C}_\rho]^{\prec 0}$ such that

$$d_K(A^{(k)}, A) \to 0 \quad \text{as} \quad k \to \infty. \tag{6.6}$$

In what follows, we prove that $A \in \Delta$. Let $f \in \mathcal{A}_n \otimes M_m$ with $\Re f \geq 0$ and $\Re f(0) = I$. Taking into account relations (6.5) and (6.4), we have

$$\begin{aligned}\Re f(A^{(k)}) + (\rho - 1)\Re f(0) &\leq \Lambda_\rho(A^{(k)}, A^{(k_0)})^2[\Re f(A^{(k_0)}) + (\rho - 1)\Re f(0)] \\ &\leq e^{2\epsilon}[\Re f(A^{(k_0)}) + (\rho - 1)\Re f(0)]\end{aligned} \tag{6.7}$$

for $k \geq k_0$. According to relation (6.6) and the definition of d_K, $\Re f(A^{(k)}) \to \Re f(A)$, as $k \to \infty$, in the operator norm topology. Consequently, relation (6.7) implies

$$\Re f(A) + (\rho - 1)\Re f(0) \leq e^{2\epsilon}[\Re f(A^{(k_0)}) + (\rho - 1)\Re f(0)]. \tag{6.8}$$

Such an inequality can be deduced in the more general case when $f \in \mathcal{A}_n \otimes M_m$ with $\Re f \geq 0$. Indeed, for each $\epsilon' > 0$ let $g := \epsilon' I + f$, $Y := \Re g(0)$, and $\varphi := Y^{-1/2} g Y^{-1/2}$. Since $\Re\varphi \geq 0$ and $\Re\varphi(0) = I$, we can apply inequality (6.8) to φ and deduce that

$$\rho\epsilon' I + \Re f(A) + (\rho - 1)\Re f(0) \leq e^{2\epsilon}\left[\rho\epsilon' I + \Re f(A^{(k_0)}) + (\rho - 1)\Re f(0)\right]$$

for any $\epsilon' > 0$. Letting $\epsilon' \to 0$, we get

$$\Re f(A) + (\rho - 1)\Re f(0) \leq e^{2\epsilon}[\Re f(A^{(k_0)}) + (\rho - 1)\Re f(0)] \tag{6.9}$$

for any $f \in \mathcal{A}_n \otimes M_m$ with $\Re f \geq 0$. Therefore,

$$A \overset{H}{\prec} A^{(k_0)}. \tag{6.10}$$

On the other hand, since $A^{(k_0)} \overset{H}{\prec} A^{(k)}$ for any $k \geq k_0$, Theorem 2.2 and relation (6.4), imply

$$\begin{aligned}\Re p(A^{(k_0)}) + (\rho - 1)\Re p(0) &\leq \Lambda_\rho(A^{(k_0)}, A^{(k)})^2[\Re p(A^{(k)}) + (\rho - 1)\Re p(0)] \\ &\leq e^{2\epsilon}[\Re p(A^{(k)}) + (\rho - 1)\Re(0)]\end{aligned}$$

for $k \geq k_0$ and any polynomial $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$, $m \in \mathbb{N}$, with $\Re p \geq 0$. According to Theorem 5.3, the d_K-topology coincides with the norm topology on $[\mathcal{C}_\rho]^{\prec 0}$. Therefore, relation (6.6) implies $A^{(k)} \to A \in [\mathcal{C}_\rho]^{\prec 0}$ in the operator norm topology. Taking the limit, as $k \to \infty$, in the inequality above, we deduce that

$$\Re p(A^{(k_0)}) + (\rho - 1)\Re p(0) \leq e^{2\epsilon}[\Re p(A) + (\rho - 1)\Re p(0)] \tag{6.11}$$

for any $p \in \mathbb{C}[X_1, \ldots, X_n] \otimes M_m$ with $\Re p \geq 0$. Consequently, we get $A^{(k_0)} \overset{H}{\prec} A$. Hence, and using relation (6.10), we obtain $A \overset{H}{\sim} A^{(k_0)}$, which proves that $A \in \Delta$. The inequalities (6.9) and (6.11) imply $\Lambda_\rho(A^{(k_0)}, A) \leq e^{2\epsilon}$. This shows that $\delta_\rho(A^{(k_0)}, A) < \epsilon$, which together with relation (6.4) imply $\delta_\rho(A^{(k)}, A) < 2\epsilon$ for any $k \geq k_0$. Therefore, $\delta_\rho(A^{(k)}, A) \to 0$, as $k \to \infty$, which proves that δ_ρ is a complete metric on the Harnack part Δ. Note that we have also proved part (ii) of this theorem.

In what follows, we prove part (iii). To this end, assume that A and B are n-tuples of operators in $[\mathcal{C}_\rho]_{<1}$. Due to Theorem 3.2, $P_\rho(B,R)$ is a positive invertible operator. Since $P_\rho(B,R)^{-1} \leq \|P_\rho(B,R)^{-1}\|$, we have $I \leq \|P_\rho(B,R)^{-1}\| P_\rho(B,R)$, which, applying the noncommutative Poisson transform, implies $I \leq \|P_\rho(B,R)^{-1}\| P_\rho(rB,R)$ for any $r \in [0,1)$. By Theorem 2.2, we deduce that $0 \overset{H}{\prec} B$ and

$$\Re f(0) \leq \|P_\rho(B,R)^{-1}\| \left[\Re f(B) + (\rho-1)\Re f(0)\right]$$

for any $f \in \mathcal{A}_n \otimes M_m$ with $\Re f \geq 0$. If, in addition, $\Re f(0) = I$, then the latter inequality implies

$$\begin{aligned} \frac{\langle [\Re f(A) + (\rho-1)\Re f(0)]x, x\rangle}{\langle [\Re f(B) + (\rho-1)\Re f(0)]x, x\rangle} - 1 &\leq \frac{\|P_\rho(B,R)^{-1}\|}{\|x\|} \left\langle \left(\Re f(A) - \Re f(B)\right) x, x\right\rangle \\ &\leq \|P_\rho(B,R)^{-1}\| d_K(A,B) \end{aligned}$$

for any $x \in \mathcal{H} \otimes \mathbb{C}^m$, $x \neq 0$. Consequently, we have

$$\ln \frac{\langle [\Re f(A) + (\rho-1)\Re f(0)]x, x\rangle}{\langle [\Re f(B) + (\rho-1)\Re f(0)]x, x\rangle} \leq \ln\left(1 + \|P_\rho(B,R)^{-1}\| d_K(A,B)\right).$$

A similar inequality can be obtained interchanging A with B. Combining these two inequalities, we get

$$\begin{aligned} &\left| \ln \frac{\langle [\Re f(A) + (\rho-1)\Re f(0)]x, x\rangle}{\langle [\Re f(B) + (\rho-1)\Re f(0)]x, x\rangle} \right| \\ &\quad \leq \ln\left(1 + \max\{\|P_\rho(B,R)^{-1}\|, \|P_\rho(A,R)^{-1}\|\} d_K(A,B)\right). \end{aligned} \tag{6.12}$$

Now, we consider the general case when $g \in \mathcal{A}_n \otimes M_m$ with $\Re g > 0$. Note that $Y := \Re g(0)$ is a positive invertible operator on $\mathcal{H} \otimes \mathbb{C}^m$ and $f := Y^{-1/2} g Y^{-1/2}$ has the properties $\Re f \geq 0$ and $\Re f(0) = I$. Applying inequality (6.12) to f when $x := Y^{-1/2} y$, $y \in \mathcal{H} \otimes \mathbb{C}^m$, and $y \neq 0$, we obtain

$$2\delta_\rho(A,B) \leq \ln\left(1 + \max\{\|P_\rho(B,R)^{-1}\|, \|P_\rho(A,R)^{-1}\|\} d_K(A,B)\right). \tag{6.13}$$

Consider a sequence $\{A^{(k)}\}_{k=1}^\infty$ of elements in $[\mathcal{C}_\rho]_{<1}$ and let $A \in [\mathcal{C}_\rho]_{<1}$ be such that $d_K(A^{(k)}, A) \to 0$, as $k \to \infty$. By Proposition 5.2, we deduce that $P_\rho(A^{(k)}, R) \to P_\rho(A,R)$ in the operator norm topology. On the other hand, due to Theorem 3.2, the operators $P(A^{(k)}, R)$ and $P(A,R)$ are invertible. Hence, and using the well-known fact that the map $Z \mapsto Z^{-1}$ is continuous on the open set of all invertible operators, we deduce that $P_\rho(A^{(k)}, R)^{-1} \to P_\rho(A,R)^{-1}$ in the operator norm topology, as $k \to \infty$. Hence, we deduce that the sequence $\{\|P_\rho(A^{(k)}, R)^{-1}\|\}_{k=1}^\infty$ is bounded. Consequently, there exists $M > 0$ with $\|P_\rho(A^{(k)}, R)^{-1}\| \leq M$ for any $k \in \mathbb{N}$. Using inequality (6.13), we obtain

$$2\delta_\rho(A^{(k)}, A) \leq \ln\left(1 + M d_K(A^{(k)}, A)\right), \qquad k \in \mathbb{N}.$$

Since $d_K(A^{(k)}, A) \to 0$, as $k \to \infty$, the latter inequality implies that $\delta_\rho(A^{(k)}, A) \to 0$. Therefore, the d_K-topology on $[\mathcal{C}_\rho]_{<1}$ is stronger than the

δ_ρ-topology. Due to the first part of this theorem, the two topologies coincide on $[\mathcal{C}_\rho]_{<1}$. Using now Theorem 5.3, we complete the proof of part (iii).

Now, we prove item (iv). Since $[\mathcal{C}_\rho]_{<1}$ is the Harnack part of 0 (see Theorem 3.2), part (i) implies its completeness with respect to the hyperbolic metric. To prove that $[\mathcal{C}_\rho]_{<1}$ is not complete with respect to the Carathéodory metric d_K and the operator metric, we consider the following example. Let $(T_1,\dots,T_n)\in B(\mathcal{P}_1)^n$ be the n-tuple of operators defined by $T_i := P_{\mathcal{P}_1} S_i|_{\mathcal{P}_1}$, $i=1,\dots,n$, where $\mathcal{P}_1 := \text{span}\{e_\alpha : \ |\alpha|\leq 1\}$. Note that $\|[T_1,\dots,T_n]\|=1$ and $T_\alpha=0$ for any $\alpha\in\mathbb{F}_n^+$ with $|\alpha|\geq 2$. Set $X_i:=\rho T_i$, $i=1,\dots,n$, and note that

$$X_\beta=\rho T_\beta=\rho P_{\mathcal{P}_1} S_\beta|_{\mathcal{P}_1}, \qquad \beta\in\mathbb{F}_n^+\backslash\{g_0\}.$$

Therefore, $(X_1,\dots,X_n)\in\mathcal{C}_\rho$, i.e., $\omega_\rho(X_1,\dots,X_n)\leq 1$, which implies $\omega_\rho(T_1,\dots,T_n)\leq\frac{1}{\rho}$. The reverse inequality is due to the fact that $\|[T_1,\dots,T_n]\|\leq\rho\omega_\rho(T_1,\dots,T_n)$. Consequently, we have

$$\omega_\rho(T_1,\dots,T_n)=\frac{1}{\rho}, \qquad \text{for } \rho\in(0,\infty).$$

On other hand, the condition $T_\alpha=0$ if $|\alpha|\geq 2$ implies $r(T_1,\dots,T_n)=0$. Therefore, we have

$$\omega_\rho(X_1,\dots,X_n)=1 \quad \text{and} \quad r(X_1,\dots,X_n)=0.$$

Now, let $c\in(0,1)$ and define $Y^{(k)}:=c^{1/k}(X_1,\dots,X_n)$ for $k=1,2,\dots$. Since $\omega_\rho(Y^{(k)})=c^{1/n}<1$, Theorem 3.2 implies $Y^{(k)}\overset{H}{\sim}0$ in $\mathcal{C}_\rho$ and $Y^{(k)}\in[\mathcal{C}_\rho]_{<1}$. On the other hand, since $\omega_\rho(X_1,\dots,X_n)=1$, we have $X:=(X_1,\dots,X_n)\notin[\mathcal{C}_\rho]_{<1}$. Now, note that

$$\begin{aligned}d_K(Y^{(k)},X)&\leq 2\|(I-R_{Y^{(k)}})^{-1}-(I-R_X)^{-1}\|\\&=2\|R_{Y^{(k)}}-R_X\|=2\|Y^{(k)}-X\|=2\|X\|(1-c^{1/k}).\end{aligned}$$

Consequently, $Y^{(k)}\to X$ in the operator norm and $d_K(Y^{(k)},X)\to 0$, as $k\to\infty$. This shows that $[\mathcal{C}_\rho]_{<1}$ is not complete with respect to the Carathéodory metric d_K and the operator metric. The proof is complete. ◻

Corollary 6.3. *Let δ_ρ be the hyperbolic metric on a Harnack part Δ of $[\mathcal{C}_\rho]^{\prec 0}$. Then*

$$d_K(A,B)\leq\max\{\|P_\rho(A,R)\|,\|P_\rho(B,R)\|\}\left(e^{2\delta_\rho(A,B)}-1\right), \qquad A,B\in\Delta.$$

If, in addition $A,B\in[\mathcal{C}_\rho]_{<1}$, then

$$2\delta_\rho(A,B)\leq\ln\left(1+\max\{\|P_\rho(B,R)^{-1}\|,\|P_\rho(A,R)^{-1}\|\}d_K(A,B)\right).$$

Corollary 6.4. *Let $f:=(f_1,\dots,f_m)$ be a contractive free holomorphic function with $\|f(0)\|<1$ such that the boundary functions $\widetilde{f}_1,\dots,\widetilde{f}_m$ are in the noncommutative disc algebra $\mathcal{A}_n$. If Δ is a Harnack part of $[\mathcal{C}_\rho]^{\prec 0}$, then the map*

$$\Delta\ni(T_1,\dots,T_n)\mapsto f(T_1,\dots,T_n)\in[\mathcal{C}_{\rho_f}]^{\prec 0}$$

is continuous with respect to the hyperbolic metric δ_ρ on Δ and the Carathéodory metric d_K on $[\mathcal{C}_{\rho_f}]^{\prec 0}$, where ρ_f is defined by relation (4.1)*. In particular, tha map*

$$[\mathcal{C}_\rho]_{<1} \ni (T_1, \ldots, T_n) \mapsto f(T_1, \ldots, T_n) \in [\mathcal{C}_{\rho_f}]_{<1}$$

is continuous with respect to the hyperbolic metric.

7. Harnack domination and hyperbolic metric for ρ-contractions (case $n = 1$)

In this section, we consider the single variable case ($n = 1$) and show that our Harnack domination of ρ-contractions is equivalent to the one introduced and studied by Cassier and Suciu in [9]. We recover some of their results and obtain some results which seem to be new even in the single variable case.

In the particular case when $n = 1$, the free pluriharmonic Poisson kernel $P_\rho(rY, R)$, $r \in [0, 1)$, coincides with

$$Q_\rho(rY, U) := \sum_{k=1} r^k Y^{*k} \otimes U^k + \rho I \otimes I + \sum_{k=1}^{\infty} r^k Y^k \otimes U^{*k}, \qquad Y \in \mathcal{C}_\rho \subset B(\mathcal{H}),$$

where the convergence of the series is in the operator norm topology and U is the unilateral shift acting on the Hardy space $H^2(\mathbb{T})$. For each ρ-contraction $T \in B(\mathcal{H})$, consider the operator-valued Poisson kernel defined by

$$K_\rho(z, T) := \sum_{k=1}^{\infty} z^k T^{*k} + \rho I + \sum_{k=1}^{\infty} \bar{z}^k T^k, \qquad z \in \mathbb{D},$$

which was employed by Cassier and Fack in [8]. Using Theorem 2.2, in the particular case when $n = 1$, we can prove the following result.

Proposition 7.1. *Let T and T' be two ρ-contractions in $B(\mathcal{H})$ and let $c \geq 1$. Then the following statements are equivalent:*

(i) $T \overset{H}{\underset{c}{\prec}} T'$;
(ii) $Q_\rho(rT, U) \leq c^2 Q_\rho(rT', U)$ *for any* $r \in [0, 1)$;
(iii) $K_\rho(z, T) \leq c^2 K_\rho(z, T')$ *for any* $z \in \mathbb{D}$.

Proof. The equivalence $(i) \leftrightarrow (ii)$ follows from Theorem 2.2, when $n = 1$. To prove the implication $(ii) \implies (iii)$, we apply the noncommutative Poisson transform (when $n = 1$) at $e^{it}I$ to the inequality of part (ii). Consequently, we obtain

$$\begin{aligned} K_\rho(re^{it}, T) &= (\mathrm{id} \otimes P_{e^{it}I})[Q_\rho(rT, U)] \\ &\leq c^2 (\mathrm{id} \otimes P_{e^{it}I})[Q_\rho(rT', U)] = c^2 K_\rho(re^{it}, T') \end{aligned}$$

for any $r \in [0,1)$ and $t \in \mathbb{R}$. Now let us prove that $(iii) \implies (ii)$. Since

$$\left\langle (T^{*k} \otimes U^k)(h_m \otimes e^{imt}), h_p \otimes e^{ipt} \right\rangle_{\mathcal{H}\otimes H^2(\mathbb{T})} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left\langle e^{ikt} T^{*k}(e^{imt}h_m), e^{ipt}h_p \right\rangle_{\mathcal{H}} dt$$

for any $h_m, h_p \in \mathcal{H}$ and $k, m, p \in \mathbb{N}$, one can easily obtain

$$\left\langle \left(c^2 Q_\rho(rT', U) - Q_\rho(rT, U)\right) h(e^{it}), h(e^{it}) \right\rangle_{\mathcal{H}\otimes H^2(\mathbb{T})} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left\langle \left(c^2 K_\rho(re^{it}, T') - K_\rho(re^{it}, T)\right) h(e^{it}), h(e^{it}) \right\rangle_{\mathcal{H}}$$

for any function $e^{it} \mapsto h(e^{it})$ in $\mathcal{H}\otimes H^2(\mathbb{T})$. Now, the implication $(iii) \implies (ii)$ is clear. The proof is complete. □

Let $T, T' \in B(\mathcal{H})$ be ρ-contractions such that $T \overset{H}{\prec} T'$. Due to Proposition 7.1 and Corollary 2.3, we deduce that

$$\begin{aligned}
\|L_{T',T}\| &= \inf\{c > 1 : \ Q_\rho(rT, U) \le c^2 Q_\rho(rT', U) \ \text{ for any } \ r \in [0,1)\} \\
&= \inf\{c > 1 : \ K_\rho(z, T) \le c^2 K_\rho(z, T') \ \text{ for any } \ z \in \mathbb{D}\} \\
&= \inf\{c > 1 : \ K_\rho(z, T^*) \le c^2 K_\rho(z, T'^*) \ \text{ for any } \ z \in \mathbb{D}\} = \|L_{T'^*, T^*}\|.
\end{aligned}$$

Therefore $T \overset{H}{\prec} T'$ if and only if $T^* \overset{H}{\prec} T'^*$.

Theorem 7.2. *Let $T, T' \in B(\mathcal{H})$ be such that $T, T' \in [\mathcal{C}_\rho]_{<1}$. Then*

$$\|L_{T',T}\| = \sup_{z\in\mathbb{D}} \|\Delta_{\rho,T'^*}(z)^{-1}(I - \bar{z}T'^*)(I - \bar{z}T^*)^{-1}\Delta_{\rho,T^*}(z)\|,$$

where

$$\Delta_{\rho,T}(z) := [\rho I + (1-\rho)(zT^* + \bar{z}T) + (\rho - 2)TT^*]^{1/2}, \qquad z \in \mathbb{D}.$$

Moreover,

$$\delta_\rho(T, T') = \ln \max\left\{\|L_{T,T'}\|, \|L_{T',T}\|\right\}.$$

Proof. If $T, T' \in [\mathcal{C}_\rho]_{<1}$, Theorem 3.4 implies

$$\begin{aligned}
\|L_{T',T}\| = \|L_{T'^*,T^*}\| &= \sup_{z\in\mathbb{D}} \|\Delta_{\rho,T^*}(z)(I - zT)^{-1}(I - zT')\Delta_{\rho,T'^*}(z)^{-1}\| \\
&= \sup_{z\in\mathbb{D}} \|\Delta_{\rho,T'^*}(z)^{-1}(I - \bar{z}T'^*)(I - \bar{z}T^*)^{-1}\Delta_{\rho,T^*}(z)\|.
\end{aligned}$$

Using now Theorem 3.5, we complete the proof. □

We mention that when $\rho = 1$, we recover a result obtained by I. Suciu [53], using different methods. However, if $\rho > 0$ and $\rho \neq 1$, the result of Theorem 7.2 seems to be new. We also remark that Proposition 3.12 , Proposition 5.2, and part (i) of Theorem 5.3 are new even in the single variable case ($n = 1$).

The next result makes an interesting connection between the Harnack domination for n-tuples of operators in $\mathcal{C}_\rho$ and and the Harnack domination for ρ-contractions ($n = 1$), via the reconstruction operator.

Theorem 7.3. *Let $A := (A_1, \ldots, A_n)$ and $B := (B_1, \ldots, B_n)$ be in $\mathcal{C}_\rho$ and let $c > 0$. Then the following statements are equivalent:*

(i) $A \overset{H}{\underset{c}{\prec}} B$;

(ii) $R_A \overset{H}{\underset{c}{\prec}} R_B$, *where $R_X := X_1^* \otimes R_1 + \cdots + X_n^* \otimes R_n$ is the reconstruction operator associated with $X := (X_1, \ldots, X_n) \in \mathcal{C}_\rho$ and the right creation operators $R_1, \ldots, R_n$.*

(iii) $R_A^* \overset{H}{\underset{c}{\prec}} R_B^*$.

Proof. First, assume that item (i) holds. Due to Theorem 2.2, we have

$$P_\rho(rA, S) \leq c^2 P_\rho(rB, S) \tag{7.1}$$

for any $r \in [0, 1)$, where $S := (S_1, \ldots, S_n)$ is the n-tuple of left creation operators. Let U be the unilateral shift on the Hardy space $H^2(\mathbb{T})$. Since $R_i^* R_j = \delta_{ij} I$, the n-tuple $(R_1 \otimes U^*, \ldots, R_n \otimes U^*)$ is a row contraction acting from $[F^2(H_n) \otimes H^2(\mathbb{T})]^n$ to $F^2(H_n) \otimes H^2(\mathbb{T})$. Applying the noncommutative Poisson transform at $(R_1 \otimes U^*, \ldots, R_n \otimes U^*)$ to inequality (7.1), we obtain

$$\begin{aligned} Q_\rho(rR_A, U) &= \left(\mathrm{id} \otimes P_{(R_1 \otimes U^*, \ldots, R_n \otimes U^*)}\right) [P_\rho(rA, S)] \\ &\leq c^2 \left(\mathrm{id} \otimes P_{(R_1 \otimes U^*, \ldots, R_n \otimes U^*)}\right) [P_\rho(rB, S)] = c^2 Q_\rho(rR_B, U) \end{aligned}$$

for any $r \in [0, 1)$. Using Proposition 7.1, we obtain that $R_A \overset{H}{\underset{c}{\prec}} R_B$. Now, assume that (ii) holds. Proposition 7.1 implies

$$K_\rho(re^{it}, R_A) \leq c^2 K_\rho(re^{it}, R_B), \qquad r \in [0, 1) \text{ and } t \in \mathbb{R}. \tag{7.2}$$

Taking $t = 0$, we obtain $P_\rho(rA, R) \leq c^2 P_\rho(rB, R)$ for any $r \in [0, 1)$, which, due to Theorem 2.2, implies $A \overset{H}{\underset{c}{\prec}} B$. The equivalence $(ii) \leftrightarrow (iii)$ is a consequence of Proposition 7.1 and the fact that inequality (7.2) is equivalent to

$$K_\rho(re^{it}, R_A^*) \leq c^2 K_\rho(re^{it}, R_B^*), \qquad r \in [0, 1) \text{ and } t \in \mathbb{R}.$$

This completes the proof. □

We remark that, according to Theorem 3.4 and Corollary 2.3, we have

$$\|L_{B,A}\| = \|C_{\rho,A} C_{\rho,B}^{-1}\| = \inf\{c > 1 : \ P_\rho(A, R) \leq c^2 P_\rho(B, R)\}$$

for any $A, B \in [\mathcal{C}_\rho]_{<1}$, where $C_{\rho,A}$ is defined in Theorem 3.4.

Corollary 7.4. *If A, B are n-tuples of operators in $[\mathcal{C}_\rho]_{<1}$, then $\|L_{B,A}\| = \|L_{R_B,R_A}\| = \|L_{R_B^*,R_A^*}\|$. Moreover, $\delta_\rho(A, B) = \delta_\rho(R_A, R_B)$.*

References

[1] T. Ando, I. Suciu, D. Timotin, Characterization of some Harnack parts of contractions, *J. Operator Theory* 2 (1979), no. 2, 233–245.

[2] C. Badea and G. Cassier, Constrained von Neumann inequalities, *Adv. Math.* 166 (2002), 260–297.

[3] C.A. Berger, A strange dilation theorem, Abstract 625–152, *Notices Amer. Math. Soc.* 12 (1965), 590.

[4] C.A. Berger and J.G. Stampfli, Mappings theorems for numerical ranges, *Amer J. Math.* 89 (1967), 1047–1055.

[5] C.A. Berger and J.G. Stampfli, Norm relations and skew dilations, *Acta Sci. Math. (Szeged)* 28 (1967), 191–195.

[6] S. Bergman, *The kernel function and conformal mapping*, Mathematical Surveys, No. V. American Mathematical Society, Providence, R.I., 1970. x+257 pp.

[7] J. W. Bunce, Models for n-tuples of noncommuting operators, *J. Funct. Anal.* 57(1984), 21–30.

[8] G. Cassier and T. Fack, Un noyau pour divers calculs fonctionnels, *C. R. Acad. Sci. Paris Sr. I Math.* 317 (1993), no. 7, 683–688.

[9] G. Cassier and N. Suciu, Mapping theorems and Harnack ordering for ρ-contractions, *Indiana Univ. Math. J.* 55 (2006), no. 2, 483–523.

[10] G. Cassier and N. Suciu, Analytic functions of a uniformly stable ρ-contraction. *Operator theory 20*, 55–73, Theta Ser. Adv. Math., 6, Theta, Bucharest, 2006.

[11] J. Cuntz, Simple C^*–algebras generated by isometries, *Commun. Math. Phys.* 57 (1977), 173–185.

[12] K. R. Davidson Free semigroup algebras. A survey. *Systems, approximation, singular integral operators, and related topics* (Bordeaux, 2000), 209–240, Oper. Theory Adv. Appl., 129, Birkhuser, Basel, 2001.

[13] E.G. Effros and Z.J. Ruan, *Operator spaces*, London Mathematical Society Monographs. New Series, 23. The Clarendon Press, Oxford University Press, New York, 2000.

[14] K. Fan, Analytic functions of a proper contraction, *Math. Z* 160 (1978), 275–290.

[15] C. Foiaş, On Harnack parts of contractions. *Rev. Roumaine Math. Pures Appl.* 19 (1974), 315–318.

[16] A. E. Frazho, Models for noncommuting operators, *J. Funct. Anal.* 48 (1982), 1–11.

[17] P.R. Halmos, *A Hilbert space problem book*, D. Van Nostrand, Princeton, NJ, 1967.

[18] P.R. Halmos, Positive definite sequences and the miracle of w, 17 pages, mimeographed.

[19] J.A.R. Holbrook, On the power-bounded operators of Sz.-Nagy and Foias, *Acta. Sci. Math. (Szeged)* 29 (1968), 299-310.

[20] J.A.R. Holbrook, Multiplicative propreties of the numerical radius in operator theory, *J. Reine Angew. Math.* 237 (1969), 166–174.
[21] T. Kato, Some mapping theorems for the numerical range, *Proc. Japan Acad.* 41 (1965), 652-655.
[22] S. Kobayashi, *Hyperbolic complex spaces*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 318. Springer-Verlag, Berlin, 1998. xiv+471 pp.
[23] S. Kobayashi, *Hyperbolic manifolds and holomorphic mappings*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005. xii+148 pp.
[24] S.G. Krantz, *Geometric function theory*, Explorations in complex analysis. Cornerstones. Birkhuser Boston, Inc., Boston, MA, 2006. xiv+314 pp.
[25] V. Müller, On the joint spectral radius, *Ann. Polon. Math.* 66 (1997), 173–182.
[26] P.S. Muhly and B. Solel, Tensor algebras over C^*-correspondences: representations, dilations, and C^*-envelopes, *J. Funct. Anal.* 158 (1998), 389–457.
[27] P.S. Muhly and B. Solel, Hardy algebras, W^*-correspondences and interpolation theory, *Math. Ann.* 330 (2004), 353–415.
[28] P.S. Muhly and B. Solel, Schur Class Operator Functions and Automorphisms of Hardy Algebras, *Documenta Math.* 13 (2008), 365–411.
[29] V.I. Paulsen, *Completely Bounded Maps and Dilations*, Pitman Research Notes in Mathematics, Vol.146, New York, 1986.
[30] C. Pearcy, An elementary proof of the power inequality for the numerical range, *Michigan Math. J.* 13 (1966), 289–291.
[31] G. Pisier, *Similarity Problems and Completely Bounded Maps*, Springer Lect. Notes Math., Vol.1618, Springer-Verlag, New York, 1995.
[32] G. Popescu, Models for infinite sequences of noncommuting operators, *Acta. Sci. Math. (Szeged)* 53 (1989), 355–368.
[33] G. Popescu, Isometric dilations for infinite sequences of noncommuting operators, *Trans. Amer. Math. Soc.* 316 (1989), 523–536.
[34] G. Popescu, Characteristic functions for infinite sequences of noncommuting operators, *J. Operator Theory* 22 (1989), 51–71.
[35] G. Popescu, Multi-analytic operators and some factorization theorems, *Indiana Univ. Math. J.* 38 (1989), 693–710.
[36] G. Popescu, Von Neumann inequality for $(B(H)^n)_1$, *Math. Scand.* 68 (1991), 292–304.
[37] G. Popescu, Functional calculus for noncommuting operators, *Michigan Math. J.* 42 (1995), 345–356.
[38] G. Popescu, Multi-analytic operators on Fock spaces, *Math. Ann.* 303 (1995), 31–46.
[39] G. Popescu, Positive-definite functions on free semigroups, *Canad. J. Math.* 48 (1996), no. 4, 887–896.
[40] G. Popescu, Noncommutative disc algebras and their representations, *Proc. Amer. Math. Soc.* 124 (1996), 2137–2148.

[41] G. Popescu, Poisson transforms on some C^*-algebras generated by isometries, *J. Funct. Anal.* 161 (1999), 27–61.

[42] G. Popescu, Curvature invariant for Hilbert modules over free semigroup algebras, *Adv. Math.* 158 (2001), 264–309.

[43] G. Popescu, Similarity and ergodic theory of positive linear maps, *J. Reine Angew. Math.* 561 (2003), 87–129.

[44] G. Popescu, Free holomorphic functions on the unit ball of $B(\mathcal{H})^n$, *J. Funct. Anal.* 241 (2006), 268–333.

[45] G. Popescu, Characteristic functions and joint invariant suspaces, *J. Funct. Anal.* 237 (2006), 277–320.

[46] G. Popescu, Hyperbolic geometry on the unit ball of $B(\mathcal{H})^n$ and dilation theory, *Indiana Univ. Math. J.* 57 (2008), No.6, 2891-2930.

[47] G. Popescu, Noncommutative transforms and free pluriharmonic functions, *Adv. Math.* 220 (2009), 831-893.

[48] G. Popescu, Unitary invariants in multivariable operator theory, *Mem. Amer. Math. Soc.* 200 (2009), no. 941, vi+91 pp.

[49] G. Popescu, Noncommutative hyperbolic geometry on the unit ball of $B(\mathcal{H})^n$, *J. Funct. Anal.* 256 (2009), 4030-4070.

[50] G. Popescu, Free holomorphic automorphisms of the unit ball of $B(\mathcal{H})^n$, *J. Reine Angew. Math.*, to appear.

[51] G. Popescu, Free holomorphic functions on the unit ball of $B(\mathcal{H})^n$. II., *J. Funct. Anal.*, to appear.

[52] W. Rudin, *Function theory in the unit ball of* $\mathbb{C}^n$, Springer-verlag, New-York/Berlin, 1980.

[53] I. Suciu, Analytic formulas for the hyperbolic distance between two contractions, *Ann. Polon. Math.* 66 (1997), 239–252.

[54] B. Sz.-Nagy, Sur les contractions de l'espace de Hilbert, *Acta. Sci. Math. (Szeged)* 15 (1953), 87–92.

[55] B. Sz.-Nagy and C. Foiaş, On certain classes of power-bounded operators in Hilbert space, *Acta. Sci. Math. (Szeged)* 27 (1966), 17–25.

[56] B. Sz.-Nagy and C. Foiaş, *Harmonic Analysis of Operators on Hilbert Space*, North Holland, New York 1970.

[57] J. von Neumann, Eine Spectraltheorie für allgemeine Operatoren eines unitären Raumes, *Math. Nachr.* 4 (1951), 258–281.

[58] K. Zhu, *Spaces of holomorphic functions in the unit ball*, Graduate Texts in Mathematics, 226. Springer-Verlag, New York, 2005. x+271 pp.

[59] J.P. Williams, Schwarz norms for operators, *Pacific J. Math.* 24 (1968), 181–188.

Gelu Popescu
Department of Mathematics
The University of Texas at San Antonio
San Antonio
TX 78249
USA
gelu.popescu@utsa.edu

Vanishing of Hochschild Cohomology for Affine Group Schemes and Rigidity of Homomorphisms between Algebraic Groups

Benedictus Margaux

Received: April 2, 2009
Revised: November 17, 2009

Communicated by Wolfgang Soergel

Abstract. Let k be an algebraically closed field. If $\mathbf{G}$ is a linearly reductive k–group and $\mathbf{H}$ is a smooth algebraic k–group, we establish a rigidity property for the set of group homomorphisms $\mathbf{G} \to \mathbf{H}$ up to the natural action of $\mathbf{H}(k)$ by conjugation. Our main result states that this set remains constant under any base change K/k with K algebraically closed. This is proven as consequence of a vanishing result for Hochschild cohomology of affine group schemes.

2000 Mathematics Subject Classification: 20G05
Keywords and Phrases: Group schemes, representations, linearly reductive group, deformation theory.

Contents

1. Introduction.

Our original goal was to prove a strong version of a rigidity principle for homomorphisms between algebraic groups which is part of the area's folklore. The general philosophy is that if $\mathbf{G}$ and $\mathbf{H}$ are algebraic groups over an algebraically closed field k, then the set $\mathrm{Hom}_{k-gr}(\mathbf{G}, \mathbf{H})$ modulo the adjoint action of $\mathbf{H}$ should remain constant under any base change K/k with K algebraically closed. Our result is as follows.

Theorem 1.1. *Let k be an algebraic closed field. Let $\mathbf{G}$ be a linearly reductive (affine) algebraic k–group, and $\mathbf{H}$ a smooth algebraic k–group scheme. Then for every algebraically closed field extension K/k, the natural map*

$$\mathrm{Hom}_{k-gr}(\mathbf{G}, \mathbf{H})/\mathbf{H}(k) \to \mathrm{Hom}_{K-gr}(\mathbf{G}_K, \mathbf{H}_K)/\mathbf{H}(K)$$

is bijective.

When k is of characteristic 0 and $\mathbf{G}$ and $\mathbf{H}$ are both reductive this result has been established by Vinberg [19, prop. 10] by reducing to the case where $\mathbf{G} = \mathbf{GL}_N$ and $\mathbf{H}$ is connected. Our proof is very different in spirit than Vinberg's, and the main result more general. The proof we give is based on the deformation theory à la Demazure-Grothendieck described in [17], which is itself linked to the analytic viewpoint later taken by Richardson on similar problems [12] [13] [16]. The main auxiliary statement we use is case (i) of the following Theorem, a vanishing result for Hochschild cohomology of affine group schemes which is of its own interest.

Theorem 1.2. *Let R be a commutative ring. Let $\mathbf{G}$ be a flat affine group scheme over $\mathrm{Spec}(R)$. Assume that the fibers of $\mathbf{G}$ over all closed points of $\mathrm{Spec}(R)$ are linearly reductive groups (as affine groups over the corresponding residue fields. See §3.1 below for the relevant definitions and references). Let L be a $\mathbf{G}$-R–module (see §2.1 below). Assume that one of the following two conditions holds:*

(i) R is noetherian,

(ii) the group $\mathbf{G}$ is of finite presentation as an R-scheme, and L is a direct limit $\mathbf{G}$-R–modules which are finitely presented as R-modules.

Then

$$H^i(\mathbf{G}, L) = 0 \text{ for all } i > 0.$$

This result extends a theorem of Grothendieck for R–groups of multiplicative type [17, IX.3.1].

At this point we recall some standard notation that will be used throughout the paper. Let S be scheme, and $\mathbf{G}$ a group scheme over S. For all scheme morphism $S \to T$ we will denote as it is customary the T-group $\mathbf{G} \times_S T$ by $\mathbf{G}_T$. If $T = \mathrm{Spec}(R)$ we write $\mathbf{G}_R$ instead of $\mathbf{G}_T$, and $\mathbf{G}(R)$ instead of $\mathbf{G}(T)$. Group schemes over a given scheme S will for brevity and convenience sometimes be refereed to simply as S–groups, or R–groups in the case when $S = \mathrm{Spec}(R)$.

2. Generalities on Hochschild cohomology

In this paper, we deal with Hochschild cohomology of a flat affine group scheme $\mathbf{G}$ over an affine base $\mathfrak{X} = \mathrm{Spec}(R)$, and their corresponding $\mathbf{G}$-$\mathcal{O}_{\mathfrak{X}}$–modules [17, I 4.7]. This set up is equivalent to that of $\mathbf{G}$-R-modules as we now explain. Let $\mathbf{G} = \mathrm{Spec}(R[\mathbf{G}])$. The group structure of $\mathbf{G}$ gives the R-algebra $R[\mathbf{G}]$ a coassociative and counital Hopf algebra structure. We have thus a comultiplication $\Delta_{\mathbf{G}} : R[\mathbf{G}] \to R[\mathbf{G}] \otimes_R R[\mathbf{G}]$, a counit $\epsilon : R[\mathbf{G}] \to R$ and an antipode map $\iota : R[\mathbf{G}] \to R[\mathbf{G}]$.
For any ring homomorphism $R \to S$ recall that the S-group $\mathbf{G} \times_R S$ obtained by base change is denoted by $\mathbf{G}_S$. This is an affine S–group with $S[\mathbf{G}] = S \otimes_R R[\mathbf{G}]$ as its Hopf algebra. Similarly, for any R–module L we denote the S–module $L \otimes_R S$ by L_S.

2.1. Definition and basic properties. Let L be an R-module, and $\rho : \mathbf{G} \to \mathbf{GL}(L)$ a linear representation of $\mathbf{G}$. This amounts to give for each R–algebra S an S–linear representation ρ_S of the abstract group $\mathbf{G}(S)$ on the S–module L_S in such a way that the family (ρ_S) is "functorial on S." We also then say that L is a (left) $\mathbf{G}$-R–module. Because $\mathbf{G}$ is affine, to give L a $\mathbf{G}$-R–module structure is equivalent to give L a (right) $R[\mathbf{G}]$-comodule structure, that is an R–linear map

$$\Delta_L : L \to L \otimes_R R[\mathbf{G}]$$

satisfying the two following natural axioms:

(CM1) The following diagram is commutative

$$\begin{array}{ccc} L & \xrightarrow{\Delta_L} & L \otimes_R R[\mathbf{G}] \\ {\scriptstyle \Delta_L}\downarrow & & \downarrow{\scriptstyle \mathrm{id}_L \otimes \Delta_{\mathbf{G}}} \\ L \otimes_R R[\mathbf{G}] & \xrightarrow{\Delta_L \otimes \mathrm{id}_{R[\mathbf{G}]}} & L \otimes_R R[\mathbf{G}] \otimes_R R[\mathbf{G}] \end{array}$$

(CM2) The composite map

$$L \xrightarrow{\Delta_L} L \otimes_R R[\mathbf{G}] \xrightarrow{\mathrm{id}_L \otimes \epsilon} L$$

is the identity map id_L.

The flatness condition on $\mathbf{G}/R$ is natural within the present context since the category of $\mathbf{G}$-R–modules is then abelian. See [15, prop. 2].[1] Recall that the fixed points of L under $\mathbf{G}$ are defined by

$$L^{\mathbf{G}} := \big\{ f \in L \quad | \quad \Delta_L(f) = f \otimes 1 \big\}.$$

This is an R–submodule of L. Because of the assumption on flatness, the Hochschild cohomology groups $H^n(\mathbf{G}, L)$ are the derived functors of the "fixed point" functor $\mathbf{G}\text{-}R - mod \to R - mod$ given by $L \to L^{\mathbf{G}}$ [17, I 5.3.1]. The $H^n(\mathbf{G}, L)$ can thus be computed as the cohomology groups of the complex [4, II §3.3.1]

[1]The existence of the unit section of $\mathbf{G}$, more precisely of the counit ϵ, shows that $R[\mathbf{G}]$ is in fact a faithfully flat R-algebra.

$$L \overset{\partial_0}{\to} L \otimes_R R[\mathbf{G}] \overset{\partial_1}{\to} L \otimes_R R[\mathbf{G}^2] \overset{\partial_2}{\to} L \otimes_R R[\mathbf{G}^3] \to \cdots \tag{2.1}$$

where as usual $R[\mathbf{G}^n] = R[\mathbf{G} \times \cdots \times \mathbf{G}] \simeq R[\mathbf{G}] \otimes_R \cdots \otimes_R R[\mathbf{G}]$, and both products and tensor products are taken n-times. We denote as it is customary $\ker(\partial_i)$ by $Z^i(\mathbf{G}, L)$ (the cocycles), and $\mathrm{Im}(\partial_{i-1})$ by $B^i(\mathbf{G}, L)$ (the coboundaries). In particular we have the exact sequence

$$0 \to L^{\mathbf{G}} = H^0(\mathbf{G}, L) \to L \overset{\partial_0}{\to} Z^1(\mathbf{G}, L) \to H^1(\mathbf{G}, L) \to 0. \tag{2.2}$$

The following four properties easily follow from the resolution (2.1).

Lemma 2.1. *Let L be a $\mathbf{G}$-R–module.*

(1) Let I be an ideal of R which annihilates L. Then $L_{R/I} = L \otimes_R R/I$ is naturally a $\mathbf{G}_{R/I}$-R/I–module, and $H^n(\mathbf{G}, L) \xrightarrow{\sim} H^n(\mathbf{G}_{R/I}, L_{R/I})$ for all $n \geq 0$.

(2) If S/R is a flat extension of rings, then

$$H^n(\mathbf{G}, L) \otimes_R S \xrightarrow{\sim} H^n(\mathbf{G}_S, L_S) \text{ for all } n \geq 0.$$

(3) Let $L = \varinjlim_i L_i$ be the inductive limit of $\mathbf{G}$-R–modules. Then

$$\varinjlim_i H^n(\mathbf{G}, L_i) \xrightarrow{\sim} H^n(\mathbf{G}, L) \text{ for all } n \geq 0.$$

(4) Let $S = \varinjlim_\alpha S_\alpha$ be an inductive limit of R–rings. Then

$$\varinjlim_\alpha H^n(\mathbf{G}_{S_\alpha}, L_{S_\alpha}) \xrightarrow{\sim} H^n(\mathbf{G}_S, L_S) \text{ for all } n \geq 0.$$

Proof. (1) The natural map $L \to L \otimes_R R/I$ is an isomorphism of both R and R/I–modules. We have R and R/I–module isomorphisms

$$\begin{aligned} L \otimes_R R[\mathbf{G}^n] \simeq L \otimes_R R/I \otimes_R R[\mathbf{G}^n] \simeq L \otimes_R R/I[\mathbf{G}^n] \simeq \\ \simeq L \otimes_R R/I \otimes_{R/I} R/I[\mathbf{G}^n] \simeq L_{R/I} \otimes_{R/I} R/I[\mathbf{G}^n]. \end{aligned}$$

Now (1) follows from the fact that $H^n(\mathbf{G}, L)$ and $H^n(\mathbf{G}_{R/I}, L_{R/I})$ are computed by the cohomology of the same complex. This is also a special case of [17, III 1.1.2].

(2) See [10, I.4.13].

(3) See [10, I.4.17].

(4) The terms of the complex (2.1) for the $\mathbf{G}_S$-S–module L_S are

$$L_S \otimes_S S[\mathbf{G}^n] = (L \otimes_R S) \otimes_S (S \otimes_R R[\mathbf{G}^n]) \simeq \big(L \otimes_R R[\mathbf{G}^n]\big) \otimes_R S$$

So this complex reads

$$L \otimes_R S \to \big(L \otimes_R R[\mathbf{G}]\big) \otimes_R S \to \big(L \otimes_R R[\mathbf{G}^2]\big) \otimes_R S \to \big(L \otimes_R R[\mathbf{G}^3]\big) \otimes_R S \cdots$$

which is the inductive limit over the S_α of the complexes

$$L \otimes_R S_\alpha \to \big(L \otimes_R R[\mathbf{G}]\big) \otimes_R S_\alpha \to \big(L \otimes_R R[\mathbf{G}^2]\big) \otimes_R S_\alpha \to \big(L \otimes_R R[\mathbf{G}^3]\big) \otimes_R S_\alpha \cdots$$

whence the statement. □

The third property in the last Proposition is useful in view of the following fact.

Proposition 2.2. *(Serre) Assume that one of the following hypothesis holds.*

(i) R is noetherian,

(ii) $\mathbf{G}$ is essentially free over R (see §6).

Let L be a $\mathbf{G}$-R–module. Then L is the inductive limit of its $\mathbf{G}$-R submodules which are of finite type as R–modules.

Proof. (i) See [14, prop. 2].

(ii) See [17, VI_B 11.10].

□

We also recall the following application of erasing functors.

Lemma 2.3. *Let $d > 0$ be a positive integer such that $H^d(\mathbf{G}, L) = 0$ for all $\mathbf{G}$-R–modules L. Then $H^{d+i}(\mathbf{G}, L) = 0$ for all $\mathbf{G}$-R–modules L and for all $i \geq 0$.*

Proof. It is enough to prove the vanishing for $d+1$. Let $\mathbf{e}_R$ be the trivial R–group, and view L as a (necessarily trivial) $\mathbf{e}_R$-R–module. We also view L as a trivial $\mathbf{G}$-R–module which we denote by L^0 to avoid any possible confusion. Now we embed L into the induced $\mathbf{G}$-R–module $\mathrm{ind}_{\mathbf{e}_R}^{\mathbf{G}}(L) = L^0 \otimes_R R[\mathbf{G}]$ via the comodule map Δ_L, and denote by Q the resulting quotient. We know that the Shapiro lemma holds [10, I.4.6], namely that

$$H^i(\mathbf{G}, \mathrm{ind}_{\mathbf{e}_R}^{\mathbf{G}}(L)) \xrightarrow{\sim} H^i(\mathbf{e}_R, L) = 0 \ \ \forall \ \ i > 0.$$

The long exact sequence for cohomology for $0 \to L \to \mathrm{ind}_1^{\mathbf{G}}(L) \to Q \to 0$ yields an isomorphism $H^d(\mathbf{G}, Q) \xrightarrow{\sim} H^{d+1}(\mathbf{G}, L)$, whence the result. □

3. Vanishing of Hochschild cohomology

The proof of Theorem 1.2 proceeds by considering successively the cases of fields, artinian rings, complete noetherian rings and local rings. We begin by recalling and collecting a few facts about linearly reductive groups.

3.1. Linearly reductive groups. Let k be a field. A k–group $\mathbf{G}$ is *linearly reductive* if it is affine and its corresponding category $\mathrm{Rep}_k(\mathbf{G})$ of finite dimensional linear representations is semisimple. We recall the following criterion.

Proposition 3.1. *Let $\mathbf{G}$ be an affine k-group. Then the following are equivalent:*

1. *$\mathbf{G}$ is linearly reductive.*
2. *Every linear representation of $\mathbf{G}$ is semisimple.*
3. *$H^1(\mathbf{G}, V) = 0$ for any finitely dimensional $\mathbf{G}$-k–module V.*
4. *$H^1(\mathbf{G}, V) = 0$ for any $\mathbf{G}$-k–module V.*
5. *$H^i(\mathbf{G}, V) = 0$ for any $\mathbf{G}$-k–module V and all $i > 0$.*

(6) *k is a direct summand of the $\mathbf{G}$-k–module structure on $k[\mathbf{G}]$ corresponding to the right regular representation.*
(7) *k is an injective $\mathbf{G}$-k–module.*

Proof. For the equivalence of the first five assertions, see [4, II prop. 3.3.7].

(2) $\Longrightarrow$ (6): This follows from the fact that k is a submodule of $k[\mathbf{G}]$.

(6) $\Longrightarrow$ (7): See [10, I.3.10].

(7) $\Longrightarrow$ (3): If k is an injective $\mathbf{G}$-k–module, the group $H^1(\mathbf{G}, V) = \mathrm{Ext}^1_{\mathbf{G}}(k, V)$ vanishes for each finite dimensional $\mathbf{G}$-k–module V (by duality). □

The property of being linearly reductive behaves well with respect to base change.

PROPOSITION 3.2. *Let $\mathbf{G}$ be an affine algebraic k–group. Let K/k be a field extension. For the K-group $\mathbf{G}_K$ to be linearly reductive it is necessary and sufficient that $\mathbf{G}$ be linearly reductive. In particular, if $\overline{k}$ is an algebraic closure of k and $\mathbf{G}_{\overline{k}}$ is a linearly reductive $\overline{k}$-group, then $\mathbf{G}$ is linearly reductive.*

This result is certainly known. We give three different proofs for the sake of completeness.

Proof. (1) As observed by S. Donkin in §2 of [5], $\mathbf{G}$ is linearly reductive if and only if the injective envelope $E_{\mathbf{G}}(k)$ of the trivial $\mathbf{G}$-k–module k coincides with k. One also knows [*ibid.* eq. (1)] that $E_{\mathbf{G}_K}(K) = E_{\mathbf{G}}(k) \otimes_k K$. The proposition follows.

(2) Assume that the k–group $\mathbf{G}$ is linearly reductive. By the criterion (6) of Proposition 3.1, k is a direct summand of $k[\mathbf{G}]$. Hence K is a direct summand of $K[\mathbf{G}]$ and therefore $\mathbf{G}_K$ is linearly reductive. Conversely if $\mathbf{G}_K$ is linearly reductive and V is a $\mathbf{G}$-k–module, then by Lemma 2.1.2 we have $H^1(\mathbf{G}, V) \otimes_k K \simeq H^1(\mathbf{G}_K, V_K) = 0$.

(3) The argument depends on the characteristic of k. One uses [4] IV prop. 3.3 in characteristic 0, and Nagata's theorem (*ibid.* théorème 3.6) if the characteristic is positive.

□

REMARK 3.3. Let $\mathbf{G}$ be an affine algebraic group over a field k. Let S be a scheme over k, and consider the S-group scheme $\mathbf{G}_S = \mathbf{G} \times_k S$. The fibers of $\mathbf{G}_S$ are then affine algebraic groups over the corresponding residue fields. It follows from the previous proposition that if *any* of the fibers is linearly reductive, then *all* fibers are linearly reductive.

The following useful statement seems to have gone unnoticed in the literature.

PROPOSITION 3.4. *Let $1 \to \mathbf{G}_1 \to \mathbf{G}_2 \to \mathbf{G}_3 \to 1$ be an exact sequence of affine algebraic k–groups. Then the following are equivalent:*

(1) *$\mathbf{G}_2$ is linearly reductive ,*
(2) *$\mathbf{G}_1$ and $\mathbf{G}_3$ are linearly reductive.*

Proof. (1) $\Longrightarrow$ (2): Since $\mathbf{G}_2/\mathbf{G}_1$ is affine, we know that the induction functor $\mathrm{ind}_{\mathbf{G}_1}^{\mathbf{G}_2}$ is exact [10, I.5.13], and therefore Shapiro's lemma hence holds (*ibid.* I.4.6). Thus

$$H^*(\mathbf{G}_2, \mathrm{ind}_{\mathbf{G}_1}^{\mathbf{G}_2}(V_1)) \xrightarrow{\sim} H^*(\mathbf{G}_1, V_1)$$

for any $\mathbf{G}_1$-k–module V_1. Thus $H^i(\mathbf{G}_1, V_1) = 0$ for $i > 0$ and Proposition 3.1 shows that $\mathbf{G}_1$ is linearly reductive. Since the functor $\mathrm{ind}_{\mathbf{G}_1}^{\mathbf{G}_2}$ is exact we can use the Hochschild-Serre spectral sequence in this framework (*ibid.* I.6.6.) Given a finite dimensional representation V_3 of $\mathbf{G}_3$, this spectral sequence reads as follows

$$E_{p,q}^2 = H^p(\mathbf{G}_3, H^q(\mathbf{G}_1, V_3)) \Longrightarrow H^{p+q}(\mathbf{G}_2, V_3).$$

Since $\mathbf{G}_1$ is linearly reductive $H^q(\mathbf{G}_1, V_3)$ vanishes for all $q \geq 1$, hence $H^n(\mathbf{G}_3, V_3) \xrightarrow{\sim} H^n(\mathbf{G}_2, V_3)$ for all $n \geq 0$. Since $H^1(\mathbf{G}_2, V_3) = 0$, $H^1(\mathbf{G}_3, V_3) = 0$ and we conclude that $\mathbf{G}_3$ is linearly reductive by Proposition 3.1.

(2) $\Longrightarrow$ (1): Assume that $\mathbf{G}_1$ and $\mathbf{G}_3$ are linearly reductive. Let us check that $\mathbf{G}_2$ is linearly reductive by again appealing to Proposition 3.1. Let V_2 be a finitely dimensional representation of $\mathbf{G}_2$. Again we can use the Hochschild-Serre spectral sequence which now reads as follows

$$E_{p,q}^2 = H^p(\mathbf{G}_3, H^q(\mathbf{G}_1, V_2)) \Longrightarrow H^{p+q}(\mathbf{G}_2, V_2).$$

The only non zero E_2-term is $H^0(\mathbf{G}_3, H^0(\mathbf{G}_1, V_2)) = H^0(\mathbf{G}_2, V_2)$. Hence $H^i(\mathbf{G}_2, V_2) = 0$ for $i > 0$. Thus $\mathbf{G}_2$ is linearly reductive. □

Note that Proposition 3.4 agrees with Nagata's theorem characterizing linearly reductive groups over an algebraically closed field [11].

Proposition 3.5. *Let* $\mathbf{G}$ *be an affine algebraic* k*–group which admits a composition series where each of the factors is of one of the following types:*

(i) algebraic k*-groups of multiplicative type,*

(ii) finite étale k*-group whose order is invertible in* k*,*

(iii) reductive k*-group if* k *is of characteristic zero.*

Then $\mathbf{G}$ *is linearly reductive.*

Proof. By Proposition 3.4 we are reduced to verifying the result for each of the given types. Proposition 3.2 permits us to assume that the base field k is algebraically closed. Case (i) is then that of a diagonalizable k–group [17, th. I.5.3.3]. Case (ii) is the case of a finite constant group of invertible order (Maschke's theorem, see [11]). Case (iii) is a classical result due to H. Weyl (see [18, th. 27.3.3]). □

3.2. Finiteness considerations. Recall that for arbitrary groups schemes $\mathbf{G}$ and $\mathbf{H}$ over a scheme S the functor $\mathbf{Hom}_{S-\mathrm{gp}}(\mathbf{G},\mathbf{H}) : Sch/S \to Sets$ is defined by

$$T \mapsto \mathbf{Hom}_{S-gr}(\mathbf{G},\mathbf{H})(T) = \mathrm{Hom}_{T-gr}(\mathbf{G}_T,\mathbf{H}_T)$$

for all schemes T/S.

The following observations will be repeatedly used in the proofs of our main results. For the remainder of this section we assume that $\mathbf{G}$ and $\mathbf{H}$ are *finitely presented* group schemes over S.

Assume that $T = \mathrm{Spec}(B)$ is an affine scheme (in the absolute sense) over S. In what follows we will encounter ourselves several times in the situation where B is given to us as an inductive limit

$$B = \varinjlim_{\lambda\in\Lambda} B_\lambda \tag{3.1}$$

over some directed set Λ. Note that the $\mathrm{Spec}(B_\lambda)$ do not in general have any natural structure of schemes over S.

Under these assumptions the group schemes $\mathbf{G}_B$ and $\mathbf{H}_B$ are defined over some B_μ by [17, VI_B 10.10.3], i.e. there exists $\mu \in \Lambda$ and finitely presented B_μ–group schemes $\mathbf{G}_\mu$ and $\mathbf{H}_\mu$ such that

$$\mathbf{G}_B = \mathbf{G}_\mu \times_{B_\mu} B \text{ and } \mathbf{H}_B = \mathbf{H}_\mu \times_{B_\mu} B. \tag{3.2}$$

Furthermore if either $\mathbf{G}$ is affine (resp. flat, smooth), so is $\mathbf{G}_\mu$ by [9] 8.10.5 (resp. 11.2.6, 17.7.8). Similarly for $\mathbf{H}$.

It follows from the very definition that

(3.3)

$$\mathbf{Hom}_{S-gp}(\mathbf{G},\mathbf{H})(B) = \mathbf{Hom}_{B-gp}(\mathbf{G}_B,\mathbf{H}_B) = \mathbf{Hom}_{B_\mu-gp}(\mathbf{G}_\mu,\mathbf{H}_\mu)(B)$$

For all $\lambda \geq \mu$ define $\mathbf{G}_\lambda = \mathbf{G}_\mu \times_{B_\mu} B_\lambda$ and $\mathbf{H}_\lambda = \mathbf{H}_\mu \times_{B_\mu} B_\lambda$. Then the canonical map

$$\varinjlim_{\lambda\geq\mu} \mathrm{Hom}_{B_\lambda-gr}(\mathbf{G}_\lambda,\mathbf{H}_\lambda) \to \mathrm{Hom}_{B-gr}(\mathbf{G}_B,\mathbf{H}_B). \tag{3.4}$$

is bijective by [17, VI_B 10.10.2] (see also [9, théorème 8.8.2]).

Remark 3.6. From the foregoing it follows that if $u, v : \mathbf{G} \to \mathbf{H}$ are two homomorphisms of S-group schemes, then there exist $\mu \in \Lambda$ such that u_B and v_B are obtained by the base change $B_\mu \to B$ from group homomorphisms $u_\mu, v_\mu \in \mathrm{Hom}_{B_\mu-gp}(\mathbf{G}_\mu,\mathbf{H}_\mu)$.

Lemma 3.7. *Let L be a $\mathbf{G}_B$-B-module which is of finite presentation as a B-module. Then there exists an index μ and a $\mathbf{G}_\mu$-B_μ-module L_μ which is finitely presented as a B_μ–module such that $L = L_\mu \otimes_{B_\mu} B$.*

Proof. According to (3.2) and proposition 8.9.1 (ii) of [9] we can find an index α, a B_α–group $\mathbf{G}_\alpha$ and a finitely presented B_α-module L_α such that $\mathbf{G}_\alpha \times_{B_\alpha} B = \mathbf{G}_B$ and $L_\alpha \otimes_{B_\alpha} B = L$. For $\lambda \geq \alpha$, we set $\mathbf{G}_\lambda = \mathbf{G}_\alpha \times_{B_\alpha} B_\lambda$ and $L_\lambda = L \otimes_{B_\alpha} B_\lambda$.

By [9] 8.5.2.2, we have an isomorphism

$$\varinjlim_{\lambda\geq\alpha} \mathrm{Hom}_{B_\lambda}\big(L_\lambda, L_\lambda \otimes_{B_\lambda} B_\lambda[\mathbf{G}_\lambda]\big) \xrightarrow{\sim} \mathrm{Hom}_B\Big(\varinjlim_{\lambda\geq\alpha} L_\lambda, \varinjlim_{\lambda\geq\alpha} L_\lambda \otimes_{B_\lambda} B_\lambda[\mathbf{G}_\lambda]\Big) =$$
$$= \mathrm{Hom}_B(L, L\otimes_B B[\mathbf{G}]).$$

It follows that there exists $\lambda \geq \alpha$ such that the B–module homomorphisms $\Delta_L : L \to L\otimes_B B[\mathbf{G}]$ is obtained by the base change $B_\lambda \to B$ from a B_λ–module homomorphism $\Delta_{L_\lambda} : L_\lambda \to L_\lambda \otimes_{B_\lambda} B_\lambda[\mathbf{G}_\lambda]$. The same reasoning applied to $\mathrm{Hom}_B(L, L\otimes_B B[\mathbf{G}] \otimes_B B[\mathbf{G}])$ and $\mathrm{Hom}_B(L, L)$ show that there exists $\mu \geq \lambda$ such that Δ_{L_λ} satisfies conditions (CM1) and (CM2) after applying the base change $B_\lambda \to B_\mu$. □

3.3. Proof of Theorem 1.2. We assume throughout that $i > 0$.
Case (i) *R a noetherian ring*: The proof is a classical dévissage argument [8, §7.2.7)].

Case of R a field: The result follows from Proposition 3.1.

Case of R local artinian: Let $\mathfrak{m}$ be the maximal ideal of R, and k the corresponding residue field. By our assumption on the closed fibers of $\mathbf{G}$ the k–group $\mathbf{G}_k = \mathbf{G}\times_R k$ is linearly reductive.
Fix an integer $e \geq 2$ such that $\mathfrak{m}^e = 0$. Thus there exists a smallest integer $j = j(L)$ such that $0 < j \leq e$ and $\mathfrak{m}^j L = 0$. We reason by induction on j.
If $j = 1$ then $\mathfrak{m}L = 0$. By Lemma 2.1.1, we have $H^i(\mathbf{G}, L) \cong H^i(\mathbf{G}_k, L_k)$ for all i, and $H^i(\mathbf{G}_k, L_k)$ vanishes since $\mathbf{G}_k$ is linearly reductive. Assume now that $H^i(\mathbf{G}, M) = 0$ for all $\mathbf{G}$-R–modules M satisfying $\mathfrak{m}^j M = 0$. If $\mathfrak{m}^{j+1}L = 0$, we consider the exact sequence

$$0 \to \mathfrak{m}L \to L \to L' \to 0$$

of $\mathbf{G}$-R–modules. Observe that $\mathfrak{m}^j(\mathfrak{m}L) = 0$ and that $\mathfrak{m}L' = 0$. This sequence gives rise to the long exact sequence of cohomology [10, I.4.2]

$$\cdots \to H^i(\mathbf{G}, \mathfrak{m}L) \to H^i(\mathbf{G}, L) \to H^i(\mathbf{G}, L') \to \cdots$$

We have $H^i(\mathbf{G}, L') = 0$ by the case $j = 1$ and $H^i(\mathbf{G}, \mathfrak{m}L) = 0$ by the induction hypothesis. Thus $H^i(\mathbf{G}, L) = 0$ as desired.

Case of R local and complete: We denote by $\mathfrak{m}$ the maximal ideal of R, and set $R_n = R/\mathfrak{m}^{n+1}$ for all $n \geq 0$.
By Lemma 2.3 it will suffice to establish the case $i = 1$. Furthermore, Proposition 2.2 together with Lemma 2.1.3 allows us to assume that L is finitely generated over R. By the Artin-Rees lemma [7, cor. 0.7.3.3] we have a natural isomorphism

$$L \xrightarrow{\sim} \varprojlim_n L_n$$

where $L_n = L\otimes_R R_n$. We are given a cocycle $z \in Z^1(\mathbf{G}, L)$ and our goal is to show by using approximation that z is a coboundary. Since R_n is a local artinian ring we have $H^1(\mathbf{G}_n, L_n) = 0$ where $\mathbf{G}_n = \mathbf{G} \times_R R_n$.[2] We consider the exact commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & H^0(\mathbf{G}, L) & \longrightarrow & L & \xrightarrow{\partial_0} & Z^1(\mathbf{G}, L) & \longrightarrow & H^1(\mathbf{G}, L) \to 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & H^0(\mathbf{G}_n, L_n) & \longrightarrow & L_n & \xrightarrow{\partial_{0,n}} & Z^1(\mathbf{G}_n, L_n) & \longrightarrow & H^1(\mathbf{G}_n, L_n) = 0. \end{array}$$

Then the images z_n of z in the $Z^1(\mathbf{G}_n, L_n)$ define elements $b_n \in B^1(\mathbf{G}_n, L_n) \subset Z^1(\mathbf{G}, L_n)$. We look now

$$0 \longrightarrow H^0(\mathbf{G}_n, L_n) \longrightarrow L_n \xrightarrow{\partial_{0,n}} B^1(\mathbf{G}_n, L_n) \longrightarrow 0.$$

Since $H^0(\mathbf{G}_n, L_n)$ is a finitely generated R_n-module, it is artinian. Hence the system $\big(H^0(\mathbf{G}, L_n)\big)_{n \geq 0}$ satisfies the Mittag-Leffler condition [7, cor. 0.13.2.2]. We get then an exact sequence (*ibid.* prop. 13.2.2)

$$0 \longrightarrow \varprojlim_n H^0(\mathbf{G}_n, L_n) \longrightarrow \varprojlim_n L_n \longrightarrow \varprojlim_n B^1(\mathbf{G}_n, L_n) \longrightarrow 0.$$

It follows that there exists $l \in L$ such that $z = \partial_0(l)$ modulo $\mathfrak{m}^{n+1}$ for all $n \geq 0$. Thus $z = \partial_0(l)$ and therefore the image of z in $H^1(\mathbf{G}, L)$ vanishes.

Case of R local: We know that the completion $\widehat{R}$ of R is local noetherian and faithfully flat over R [7, cor. 0.7.3.5]). By Lemma 2.1.2, we have

$$H^i(\mathbf{G}, L) \otimes_R \widehat{R} \xrightarrow{\sim} H^i(\mathbf{G}_{\widehat{R}}, L_{\widehat{R}}).$$

The right hand side vanishes by the local complete case, hence $H^i(\mathbf{G}, L) = 0$ by faithfully flat descent.

Case of R arbitrary noetherian: By the same reasoning used in the previous case we have $H^i(\mathbf{G}, L) \otimes_R R_{\mathfrak{m}} = 0$ for any maximal ideal $\mathfrak{m}$ of R. Thus $H^i(\mathbf{G}, L) = 0$.

Case (ii) *The group $\mathbf{G}$ is finitely presented as an R-scheme and L is a direct limit of $\mathbf{G}$-R-modules which are finitely presented as R–modules*: By Lemma 2.1.3 we may assume that L is a finitely presented R–module. The same reasoning used in the final step of the noetherian case allows us to assume that R is a local ring. Let $\mathfrak{m}$ be the maximal ideal of R and k its residue field. We consider the standard filtration $R = \varinjlim_\lambda R_\lambda$ of R by its finitely generated (hence noetherian) $\mathbb{Z}$–subalgebras. For each λ, we consider the prime ideal $\mathfrak{p}_\lambda := \mathfrak{p} \cap R_\lambda$ of R_λ, and the corresponding local ring $R'_\lambda := (R_\lambda)_{\mathfrak{p}_\lambda}$ whose maximal ideal $\mathfrak{p}_\lambda R'_\lambda$ we denote by $\mathfrak{m}_\lambda$. Note that the residue field k_λ of R'_λ is a subfield of k. We have $R = \varinjlim_\lambda R'_\lambda$ and the following commutative diagram

[2] One of course verifies that the R_n-groups $\mathbf{G}_n$ satisfy the assumptions of the theorem. Similar considerations apply to the reductions that follow.

$$\begin{array}{ccc} R'_\lambda & \longrightarrow & R \\ \Big\downarrow & & \Big\downarrow \\ k_\lambda = R'_\lambda/\mathfrak{m}_\lambda & \longrightarrow & R/\mathfrak{m} = k. \end{array}$$

We now apply the considerations of §3.2 to the case when $S = \mathrm{Spec}(R)$, $B = R$ and $B_\lambda = R'_\lambda$. This yield the existence of an R'_μ, an affine, flat and finitely presented R'_μ– group scheme $\mathbf{G}_\mu$ and a $\mathbf{G}_\mu$-R'_μ-module L_μ such that $\mathbf{G} = \mathbf{G}_\mu \times_{R'_\mu} R$ and $L = L_\mu \otimes_{R'_\mu} R$. By Lemma 2.1.4, we have

$$H^i(\mathbf{G}, L) = \varinjlim_{\lambda \geq \mu} H^i(\mathbf{G}_\mu \times_{R'_\mu} R'_\lambda, L_\mu \otimes_{R'_\mu} R'_\lambda). \tag{3.5}$$

We also have by the transitivity of base change that

$$(\mathbf{G}_\mu \times_{R'_\mu} k_\mu) \times_{k_\mu} k \simeq \mathbf{G}_\mu \times_{R'_\mu} k \simeq (\mathbf{G}_\mu \times_{R'_\mu} R) \times_R k = \mathbf{G} \times_R k. \tag{3.6}$$

From our assumptions on the R–group $\mathbf{G}$ it follows that the k–group $\mathbf{G} \times_R k$ is affine algebraic and linearly reductive. It then follows from (3.6) and Proposition 3.2 that the k_μ–algebraic group $\mathbf{G}_\mu \times_{R'_\mu} k_\mu$ is linearly reductive as well. This shows that the R'_μ–group $\mathbf{G}_\mu$ satisfies the assumption of the first part of the theorem. Similar considerations apply to the R'_λ–group $\mathbf{G}_\mu \times_{R'_\mu} R'_\lambda$ for all $\lambda \geq \mu$. Thus the noetherian case that we have already established shows, with the aid of (3.5), that $H^i(\mathbf{G}, L) = 0$. $\square$

4. Rigidity and deformation theory

4.1. Locally finitely presented S–functors. Let S be a scheme and $F : Sch/S \to Sets$ a contravariant functor. We recall the following definitions:

- F is locally of finite presentation over S if for every filtered inverse system of affine S-schemes $\mathrm{Spec}(B_i)$, the canonical morphism

$$\varinjlim F(B_i) \to F(\varinjlim B_i)$$

is an isomorphism [3, §8.3].[3]

- F is formally smooth (resp. formally unramified, formally étale) if for any affine scheme $\mathrm{Spec}(B)$ over S and any subscheme $\mathrm{Spec}(B_0)$ of $\mathrm{Spec}(B)$ defined by a nilpotent ideal I of B, the map

$$F(B) \to F(B_0)$$

is surjective (resp. injective, bijective) [17, XI.1.1].

Note that all these definitions are stable by an arbitrary base change $T \to S$. In the second definition, we can require furthermore that $I^2 = 0$. The following lemma is elementary.

[3]This reference has assumptions on the nature of S related to Artin's approximation theorem which are relevant to their work, but not to ours. As it is customary, given an affine scheme $\mathrm{Spec}(B)$ over S, we write $F(B)$ instead of $F\big(\mathrm{Spec}(B)\big)$.

Lemma 4.1. *Assume that F is locally of finite presentation over S. Consider a field extension K/k over S, that is morphisms* $\mathrm{Spec}(K) \to \mathrm{Spec}(k) \to S$. *Assume that k is separably closed and K is a separable field extension of k. Then the map $F(k) \to F(K)$ is injective.*

Remark 4.2. *If k is algebraically closed, any field extension K/k is separable, hence the Lemma applies.*

Proof. We may assume without loss of generality that $S = \mathrm{Spec}(k)$. We are given two elements $\alpha, \beta \in F(k)$ with same image in $F(K)$. Since K is the inductive limit of its finitely generated subalgebras, there exists a finitely generated k–algebra A such that α and β have same image in $F(A)$. Since K/k is separable, the finitely generated k–subalgebras of K are separable over k. Hence A is integral and absolutely reduced (i.e. $A \otimes_k \overline{k}$ is reduced), and therefore the affine variety $\mathrm{Spec}(A)$ admits a k–point [1, AG.13.3]. In other words, the ring homomorphim $k \to A$ admits a section. This, in turn, induces a section of the group homomorphism $F(k) \to F(A)$, hence $\alpha = \beta$ in $F(k)$. □

4.2. Formal étalness. We recall the following crucial statement of deformation theory for group scheme homomorphisms due to Demazure.

Theorem 4.3. *(*[17, cor. III.2.6]*) Let $\mathbf{G}$ and $\mathbf{H}$ be group schemes over a scheme S. Assume that $\mathbf{G}$ is affine (in the absolute sense) and flat, and that $\mathbf{H}$ is smooth. Let S_0 be a closed subscheme of S defined by an ideal I of $\mathcal{O}_S$ such that $I^2 = 0$. We set $\mathbf{G}_0 = \mathbf{G} \times_S S_0$ and $\mathbf{H}_0 = \mathbf{H} \times_S S_0$. Let $f_0 : \mathbf{G}_0 \to \mathbf{H}_0$ be a homomorphism of S_0-groups, and let $\mathbf{G}_0$ act on* $\mathrm{Lie}(\mathbf{H}_0)$ *via f_0 and the adjoint representation of $\mathbf{H}_0$. Then*

(1) *If $H^2\big(\mathbf{G}_0, \mathrm{Lie}(\mathbf{H}_0) \otimes_{\mathcal{O}_{S_0}} I\big) = 0$ the homomorphism f_0 lifts to an S–group homomorphism $f : \mathbf{G} \to \mathbf{H}$.*

(2) *If $H^1\big(\mathbf{G}_0, \mathrm{Lie}(\mathbf{H}_0) \otimes_{\mathcal{O}_{S_0}} I\big) = 0$, then any two liftings f and f' of f_0 as in (1) are conjugate under an element of* $\ker\big(\mathbf{H}(S) \to \mathbf{H}(S_0)\big)$. *More precisely $f' = \mathrm{int}(h) f$ for some $h \in \ker\big(\mathbf{H}(S) \to \mathbf{H}(S_0)\big)$.* □

Combined with the vanishing result given by Theorem 1.2 we are very close to the completion of the proof of our main result. The missing ingredient is some detailed information pertaining to the nature of certain functors related to homomorphisms between group schemes.

Let $\mathbf{G}$ and $\mathbf{H}$ be group schemes over a scheme S. The functor $\mathbf{Hom}_{S-\mathrm{gp}}(\mathbf{G}, \mathbf{H})$ was already defined in §3.2. Any element $h \in \mathbf{H}(T)$ defines an inner automorphism $\mathrm{int}(t) \in \mathrm{Aut}_{T-gr}(\mathbf{H}_T)$, and this last group acts naturally on the set $\mathbf{Hom}_{S-gr}(\mathbf{G}, \mathbf{H})(T)$. This allows us to define a new functor $\overline{\mathbf{Hom}}_{S-gr}(\mathbf{G}, \mathbf{H}) :$ $Sch/S \to Sets$ by

$$T \mapsto \overline{\mathbf{Hom}}_{S-gr}(\mathbf{G}, \mathbf{H})(T) = \mathrm{Hom}_{T-gr}(\mathbf{G}_T, \mathbf{H}_T)/\mathbf{H}(T).$$

The final functor which is relevant to us is the transporter of two elements of $\mathrm{Hom}_{S-gr}(\mathbf{G}, \mathbf{H})$. Let $u, v : \mathbf{G} \to \mathbf{H}$ two homomorphisms of S-group schemes.

Recall the subfunctor $\mathbf{Transp}(u,v)$ of $\mathbf{H}$ defined by

$$T \to \mathbf{Transp}(u,v)(T) = \big\{ h \in \mathbf{H}(T) \mid u_T = \mathrm{int}(h)\, v_T \big\}.$$

We begin with an easy observation.

Lemma 4.4. *Let* $\mathbf{G}$ *and* $\mathbf{H}$ *be finitely presented group schemes over* S*, and let* $u,v \in \mathrm{Hom}_{S-gr}(\mathbf{G},\mathbf{H})$*. The* S*-functors* $\mathbf{Hom}_{S-gr}(\mathbf{G},\mathbf{H})$*,* $\overline{\mathbf{Hom}}_{S-gr}(\mathbf{G},\mathbf{H})$ *and* $\mathbf{Transp}(u,v)$ *are locally of finite presentation.*

Proof. For every filtered inverse system of affine schemes $\mathrm{Spec}(B_\lambda)$ over S based on some directed set Λ, and we have to show that the canonical morphisms

$$\varinjlim \mathbf{Hom}_{S-gr}(\mathbf{G},\mathbf{H})(B_\lambda) \to \mathbf{Hom}_{S-gr}(\mathbf{G},\mathbf{H})(\varinjlim B_\lambda) = \mathbf{Hom}_{S-gr}(\mathbf{G},\mathbf{H})(B),$$

$$\varinjlim \overline{\mathbf{Hom}}_{S-gr}(\mathbf{G},\mathbf{H})(B_\lambda) \to \overline{\mathbf{Hom}}_{S-gr}(\mathbf{G},\mathbf{H})(\varinjlim B_\lambda) = \overline{\mathbf{Hom}}_{S-gr}(\mathbf{G},\mathbf{H})(B),$$

and

$$\varinjlim \mathbf{Transp}(u,v)(B_\lambda) \to \mathbf{Transp}(u,v)(\varinjlim B_\lambda) = \mathbf{Transp}(u,v)(B)$$

are bijective. Taking into account (3.2), (3.3) and (4.1) we may replace S by $\mathrm{Spec}(B_\mu)$ for some suitable index $\mu \in \Lambda$, and replace Λ by the subset of Λ consisting of all indices $\lambda \geq \mu$. Denote $\mathbf{G} \times_{B_\mu} B_\lambda$ and $\mathbf{H} \times_{B_\mu} B_\lambda$ by $\mathbf{G}_\lambda$ and $\mathbf{H}_\lambda$ respectively, just as we did in §3.2. Then (3.4) shows that $\mathbf{Hom}_{S-gr}(\mathbf{G},\mathbf{H})$ is locally of finite presentation.

As for the second assertion, we look in view of (3.3) at the map

$$\varinjlim \mathrm{Hom}_{B_\lambda-gr}(\mathbf{G}_\lambda,\mathbf{H}_\lambda)/\mathbf{H}_\lambda(B_\lambda) \to \mathrm{Hom}_{B-gr}(\mathbf{G}_B,\mathbf{H}_B)/\mathbf{H}_B(B)$$

which is already known to be surjective. For the injectivity, we are given $\phi_\alpha, \phi'_\alpha \in \mathrm{Hom}_{B_\alpha-gr}(\mathbf{G}_\alpha,\mathbf{H}_\alpha)$ for some $\alpha \geq \mu$ whose images ϕ, ϕ' in $\mathrm{Hom}_{B-gr}(\mathbf{G}_B,\mathbf{H}_B)$ are conjugated under $\mathbf{H}_B(B) = \mathbf{H}(B)$. Since $\mathbf{H}$ is of finite presentation $\varinjlim \mathbf{H}(B_\lambda) \xrightarrow{\sim} \mathbf{H}(B)$. So there exists $\beta \geq \alpha$ and $h_\beta \in \mathbf{H}(B_\beta)$ such that $\phi = \mathrm{int}(h)\phi'$ where h stands for the image of h_β in $\mathbf{H}(B)$. By (3.4) there exists $\gamma \geq \beta$ such that $\phi_\alpha \times_{B_\alpha} \mathrm{id}_{B_\gamma} = \mathrm{int}(h_\gamma)(\phi'_\alpha \times_{B_\alpha} \mathrm{id}_{B_\gamma})$, where h_γ is the image of h_β in $\mathbf{H}(B_\gamma)$. In other words, $\phi_\alpha, \phi'_\alpha$ map to the same element of $\mathrm{Hom}_{B_\gamma-gr}(\mathbf{G}_\gamma,\mathbf{H}_\gamma)/\mathbf{H}_\gamma(B_\gamma)$, hence define the same element in the inductive limit $\varinjlim \mathrm{Hom}_{B_\lambda-gr}(\mathbf{G}_\lambda,\mathbf{H}_\lambda)/\mathbf{H}_\lambda(B_\lambda)$. We conclude that $\overline{\mathbf{Hom}}_{S-gr}(\mathbf{G},\mathbf{H})$ is locally of finite presentation.

Finally we look at the case of the transporter. Assume that $h \in \mathbf{H}(B)$ is such that $u_B = \mathrm{int}(h)v_B$. Since $\mathbf{H}$ is finitely presented there exists $\alpha \geq \mu$ and an element $h_\alpha \in \mathbf{H}(B_\alpha)$ whose image in $\mathbf{H}(B)$ is h. Then the two elements u_α and $\mathrm{int}(h_\alpha)v_\alpha$ of $\mathrm{Hom}_{B_\alpha}(\mathbf{G}_\alpha,\mathbf{H}_\alpha)$ map to the same element of $\mathrm{Hom}_B(\mathbf{G}_B,\mathbf{H}_B)$. By (3.4) there exists $\beta \geq \alpha$ such that $u_\beta = \mathrm{int}(h_\beta)v_\beta$ (where the subindex β denotes the image of the element in question under the map $B_\alpha \to B_\beta$). This shows that our map is surjective. Note that from the definition of the transporter it follows that

$$\mathbf{Transp}(u,v)(B) = \mathbf{Transp}(u_B,v_B)(B) = \mathbf{Transp}(u_\mu,v_\mu)(B) \tag{4.1}$$

Injectivity is clear since for all $\lambda \geq \mu$ we have $\mathbf{Transp}(u_\mu, v_\mu)(B_\lambda) \subset \mathbf{H}_\mu(B_\lambda)$ and $\mathbf{H}_\mu$ is of finite presentation. $\square$

Theorem 4.5. *Let S be a scheme and let $\mathbf{G}$ and $\mathbf{H}$ be finitely presented group schemes over S. Assume that $\mathbf{G}$ is affine (in the absolute sense) and flat, and that $\mathbf{H}$ is smooth. Assume that for all $s \in S$ the fiber $\mathbf{G}_s$ is linearly reductive (as an affine algebraic group over the residue field $\kappa(s)$ of s). Then*

(1) The functor $\mathbf{Hom}(\mathbf{G}, \mathbf{H})$ is formally smooth.

(2) The functor $\overline{\mathbf{Hom}}(\mathbf{G}, \mathbf{H})$ is formally étale.

(3) If $u, v : \mathbf{G} \to \mathbf{H}$ are two homomorphisms of S-group schemes, the subfunctor $\mathbf{Transp}(u, v)$ of $\mathbf{H}$ is formally smooth.

The case when $\mathbf{G}$ is of multiplicative type is an important result of Grothendieck [17, XI prop. 2.1]. If S is of characteristic zero and $\mathbf{G}$ is reductive, the first statement is due to Demazure [17, XXIV prop. 7.3.1.a].

Proof. We note that if $\mathbf{Hom}_{S-gr}(\mathbf{G}, \mathbf{H})$ is formally smooth, then $\overline{\mathbf{Hom}}_{S-gr}(\mathbf{G}, \mathbf{H})$ is formally smooth as well. As a consequence, to establish (1) and (2) it will suffice to prove that $\mathbf{Hom}_{S-gr}(\mathbf{G}, \mathbf{H})$ is formally smooth and that $\overline{\mathbf{Hom}}_{S-gr}(\mathbf{G}, \mathbf{H})$ is formally unramified. We are given an affine scheme $\mathrm{Spec}(B)$ over S, and a closed subscheme $\mathrm{Spec}(B_0)$ defined by an ideal I of B satisfying $I^2 = 0$, and we need to show that

(I) $\mathbf{Hom}_{S-gr}(\mathbf{G}, \mathbf{H})(B) \to \mathbf{Hom}_{S-gr}(\mathbf{G}, \mathbf{H})(B_0)$ is surjective,

(II) $\overline{\mathbf{Hom}}_{S-gr}(\mathbf{G}, \mathbf{H})(B) \to \overline{\mathbf{Hom}}_{S-gr}(\mathbf{G}, \mathbf{H})(B_0)$ is injective, and

(III) $\mathbf{Transp}(u, v)(B) \to \mathbf{Transp}(u, v)(B_0)$ is surjective.

Proof of (I) *and* (II): By the first equality of (3.3) we may assume with no loss of generality that $S = \mathrm{Spec}(B)$. We claim that, with the obvious adaptations to the notation of Theorem 4.3,

$$H^i\big(\mathbf{G}_0, \mathrm{Lie}(\mathbf{H}_0) \otimes_{B_0} I\big) = 0 \text{ for all } i > 0. \tag{4.2}$$

Write $B = \varinjlim B_\lambda$ where the limit is taken over all finitely generated $\mathbb{Z}$–subalgebras (hence noetherian) B_λ of B. Then $J_\lambda := I \cap B_\lambda$ is an ideal of B_λ such that $J_\lambda^2 = 0$ and $I = \varinjlim J_\lambda$. Consider the trivial $\mathbf{G}_0$-B_0 module

$$I_\lambda := J_\lambda \otimes_{B_\lambda} B_0.$$

Since J_λ is a B_λ-module of finite presentation, I_λ is a B_0-module of finite presentation. We have an isomorphism of B_0-modules

$$\varinjlim I_\lambda \xrightarrow{\sim} I$$

hence an isomorphism of $\mathbf{G}_0$-B_0–modules

$$\varinjlim \big(\mathrm{Lie}(\mathbf{H}_0) \otimes_{B_0} I_\lambda\big) \xrightarrow{\sim} \mathrm{Lie}(\mathbf{H}_0) \otimes_{B_0} I.$$

Because $\mathbf{H}_0$ is a smooth B_0–group $\mathrm{Lie}(\mathbf{H}_0)$ is a finitely presented B_0–module (see [4] II §4.8). Since the tensor product of finitely presented modules is finitely presented, $\mathrm{Lie}(\mathbf{H}_0) \otimes_{B_0} I$ is a direct limit of $\mathbf{G}_0$-B_0–modules which are finitely presented as B_0-modules. It is clear that the B_0-groups $\mathbf{G}_0$ and $\mathbf{H}_0$ satisfy the assumptions of Theorem 1.2.2. This shows that (4.2) holds, and we can now apply Theorem 4.3 to obtain (I) and (II)

Proof of (III): For convenience we denote $\mathbf{Transp}(u,v)$ by $\mathbf{T}(u,v)$. Consider the B-group homomorphisms $u_B, v_B \in \mathrm{Hom}_{B-gr}(\mathbf{G}_B, \mathbf{H}_B)$ induced by the base change $\mathrm{Spec}(B) \to S$. By the definition of the transporter we see that $\mathbf{T}(u,v)(B) = \mathbf{T}(u_B,v_B)(B)$ and $\mathbf{T}(u,v)(B_0) = \mathbf{T}(u_B,v_B)(B_0)$ where $\mathrm{Spec}(B_0) \to \mathrm{Spec}(B)$ is the natural map. From this it follows that to establish (III) we may assume without loss of generality that $S = \mathrm{Spec}(B)$.
Let u_0 and v_0 be the elements of $\mathrm{Hom}_{B_0-gr}(\mathbf{G}_0, \mathbf{H}_0)$ induced by the base change $B \to B_0$. Let $h_0 \in \mathbf{Transp}(u_0,v_0)(B_0)$, so that $u_0 = \mathrm{int}(h_0)v_0$. Lift h_0 to an element $h' \in \mathbf{H}(B)$ (which is possible since $\mathbf{H}$ is smooth), and set $u' = \mathrm{int}(h')v_B$. Then u' and u_B map to the same element of $\mathrm{Hom}_{B_0-gr}(\mathbf{G}_0, \mathbf{H}_0)$, namely u_0. By II there exists $h'' \in \mathbf{H}(B)$ such that $u_B = \mathrm{int}(h'')u'$. Furthermore, because of (4.2) we may assume that $h'' \in \ker\big(\mathbf{H}(B) \to \mathbf{H}(B_0)\big)$. Then $h = h''h' \in \mathbf{H}(B)$ maps to h_0 and satisfies $u_B = \mathrm{int}(h)v_B$. □

Remark 4.6. The assumption on the fibers of $\mathbf{G}$ is not superfluous. Let B $= \mathbb{C}[\epsilon]$ be the ring of dual numbers over $\mathbb{C}$, and let $S = \mathrm{Spec}(B)$. If $I = \mathbb{C}\epsilon$, then $B_0 = \mathbb{C}$. Consider now the case when $\mathbf{G} = \mathbf{G}_a$ and $\mathbf{H} = \mathbf{G}_m$ (the additive and multiplicative groups over B.)
It is well-known that $\mathbf{Hom}_{B-gr}(\mathbf{G},\mathbf{H})(B_0) = \mathrm{Hom}_{\mathbb{C}-gr}(\mathbf{G}_{a,\mathbb{C}}, \mathbf{G}_{m,\mathbb{C}})$ is trivial. On the other hand $\mathbf{Hom}_{B-gr}(\mathbf{G},\mathbf{H})(B)$ is infinite; it consists of the homomorphisms $\{\phi_z : z \in \mathbb{C}\}$ which under Yoneda correspond to the B–Hopf algebra homomorphisms $\phi_z^* : B[t^{\pm 1}] \to B[x]$ given by $\phi_z^* : t \mapsto 1 + z\epsilon x$. Since $\mathbf{H}$ is abelian the functors $\overline{\mathbf{Hom}}_{B-gr}(\mathbf{G},\mathbf{H})$ and $\mathbf{Hom}_{B-gr}(\mathbf{G},\mathbf{H})$ coincide. The above discussion shows that $\overline{\mathbf{Hom}}_{B-gr}(\mathbf{G},\mathbf{H})$ is not formally étale.

Lemma 4.7. *If* $\mathbf{G}$ *is essentially free over* S *(see §6), the functor* $\mathbf{Transp}(u,v)$ *is representable by a closed* S*–subscheme of* $\mathbf{H}$

Proof. Consider the two morphisms $q_1, q_2 : \mathbf{H} \to \mathbf{Hom}(\mathbf{G},\mathbf{H})$ which for all schemes T/S and $h \in \mathbf{H}(T)$ are given by and $q_1(h) = u_T$ and $q_2(h) = \mathrm{int}(h)v_T$. Since $\mathbf{G}$ is assumed essentially free over S and $\mathbf{H}$ is separated over S, Grothendieck's criterion [17, VIII.6.5.b] applied to $\mathbf{X} = \mathbf{H}$, $\mathbf{Y} = \mathbf{G}$, $\mathbf{Z} = \mathbf{H}$ shows the representability of $\mathbf{Transp}(u_1,u_2)$ by a closed S–subscheme of $\mathbf{H}$. □

Corollary 4.8. *Under the assumptions of Theorem 4.5, assume furthermore that* $\mathbf{G}$ *is essentially free over* S*. Let* $u,v : \mathbf{G} \to \mathbf{H}$ *be two homomorphisms of* S*-group schemes. Then the* S*-functor* $\mathbf{Transp}(u,v)$ *is representable by a smooth closed* S*-subscheme of* $\mathbf{H}$*. In particular, if* $u = v$*, then the centralizer subfunctor* $\mathbf{Centr}(u)$ *of* $\mathbf{H}$ *is representable by a smooth closed subscheme of* $\mathbf{H}$*.*

Proof. By the last Lemma the S-functor $\mathbf{Transp}(u,v)$ is representable by a closed subscheme of $\mathbf{H}$, which is locally of finite presentation by Lemma 4.4

and [9] 8.14.2, and formally smooth by Theorem 4.5. Thus $\mathbf{Transp}(u,v)$ is a smooth scheme over S. □

Corollary 4.9. *Under the assumptions of Theorem 4.5, assume furthermore that $\mathbf{G}$ is essentially free over S and that $S = \mathrm{Spec}(B)$ where B is a henselian local ring of residue field k. Then the map $\overline{\mathbf{Hom}}(\mathbf{G},\mathbf{H})(B) \to \overline{\mathbf{Hom}}(\mathbf{G},\mathbf{H})(k)$ is injective. Thus two homomorphisms $u, v : \mathbf{G} \to \mathbf{H}$ of S-group schemes are conjugate under $\mathbf{H}(B)$ if and only if $u_k, v_k : \mathbf{G} \times_B k \to \mathbf{H} \times_B k$ are conjugate under $\mathbf{H}(k)$.*

Proof. By Corollary 4.8, the B–functor $\mathbf{Transp}(u,v)$ is representable by a smooth B–scheme. By Hensel's lemma [3, §2.3] the map

$$\mathbf{Transp}(u,v)(B) \to \mathbf{Transp}(u,v)(k)$$

is surjective. Thus if $u_k, v_k : \mathbf{G} \times_B k \to \mathbf{H} \times_B k$ are such that $u_k = \mathrm{int}(h_0)v_k$ for some $h_0 \in \mathbf{H}(k)$, then there exists $h \in \mathbf{H}(B)$ such that $u = \mathrm{int}(h)v$. □

5. Applications

5.1. Rigidity. Our first result establishes Theorem 1.1

Theorem 5.1. *Let k be a field. Let $\mathbf{G}$ be a linearly reductive algebraic k–group and let $\mathbf{H}$ be a smooth algebraic k–group. Let K/k be a field extension such that k and K are both separably closed and K is separable over k (for example if both k and K are algebraically closed).*
Then the map

$$\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G},\mathbf{H})(k) \to \overline{\mathbf{Hom}}_{k-gr}(\mathbf{G},\mathbf{H})(K) = \overline{\mathbf{Hom}}_{K-gr}(\mathbf{G}_K,\mathbf{H}_K)(K)$$

is bijective.

Proof. By Lemma 4.1.1 and Lemma 4.4 the map $\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G},\mathbf{H})(k) \to$ $\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G},\mathbf{H})(K)$ is injective. Conversely we are given an element $u \in \mathrm{Hom}_{K-gr}(\mathbf{G}_K,\mathbf{H}_K)$ and we want to show that there exists $v_0 \in \mathrm{Hom}_{k-gr}(\mathbf{G},\mathbf{H})$ and $h \in \mathbf{H}(K)$ such that $v_0 \times_k \mathrm{id}_K = \mathrm{int}(h)u$.
The homomorphism $u : \mathbf{G}_K \to \mathbf{H}_K$ arises by base change from some A-group scheme homomorphism $v \in \mathrm{Hom}_{A-gr}(\mathbf{G}_A,\mathbf{H}_A)$, i.e. $u = v_K$, where $A \subset K$ is a finitely generated k–algebra. Under our assumption on k we may assume, by considering a basic open affine subsheme of $\mathrm{Spec}(A)$ if needed, that A is smooth over k. In particular, A is normal.
Since A is separable over k and k is separably closed, there exists a maximal ideal $\mathfrak{m}$ of A such that $A/\mathfrak{m} = k$. Then v gives rise to a k–homomorphism $v_0 : \mathbf{G} \to \mathbf{H}$. Denote by B the (strict) henselization of the local ring $A_{\mathfrak{m}}$. Then B is noetherian and may be identified with a subring of a separable closure of the fraction field of A [M, I.4.10, 11]. In particular B can be assumed to embed into K. By Proposition 3.2 and Remark 3.3 the group $\mathbf{G}_B$ (which is clearly affine and free of finite rank over B) satisfies the assumption on the fibers of Theorem 4.5. By Corollary 4.9, $v_0 \times_k \mathrm{id}_B = \mathrm{int}(h)(v \times_A \mathrm{id}_B)$ for some $h \in \mathbf{H}(B)$. Thus $v_0 \times_k \mathrm{id}_K = \mathrm{int}(h_K)u$ as desired. □

Remark 5.2. The assumption that $\mathbf{G}$ be linearly reductive is not superfluous. Recall (see §3.1) that $H^1(\mathbf{G}, V) = \mathrm{Ext}^1_{\mathbf{G}}(k, V)$. Assume that k is algebraically closed of positive characteristic, and let K/k be an arbitrary field extension. One knows from Nagata's work that for each non-trivial semisimple k-group $\mathbf{G}$ there exists a non-trivial finite dimensional irreducible $\mathbf{G}$-k–module V such that $H^1(\mathbf{G}, V) \neq 0$. This implies that Theorem 5.1 fails for $\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G}, \mathbf{GL}_n)$ if $n = \dim(V) + 1$. Indeed $\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G}, \mathbf{GL}_n)(k)$ measures the equivalence classes of n–dimensional linear representations of $\mathbf{G}$. We know that $\mathrm{Ext}^1_{\mathbf{G}}(k, V)$, which by assumption is a non-trivial k-space, can be identified with the subset of $\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G}, \mathbf{GL}_n)(k)$ that corresponds to those representations of $\mathbf{G}$ that are extensions of k by V. Similar considerations apply to $\overline{\mathbf{Hom}}_{K-gr}(\mathbf{G}_K, \mathbf{GL}_{n,K})(K)$. Since $H^1(\mathbf{G}, V) \otimes_k K = H^1(\mathbf{G}_K, V_K)$ the foregoing discussion shows that the natural map $\overline{\mathbf{Hom}}_{k-gr}(\mathbf{G}, \mathbf{H})(k) \to \overline{\mathbf{Hom}}_{k-gr}(\mathbf{G}, \mathbf{H})(K)$ is not surjective whenever $k \neq K$.

Remark 5.3. Let $\mathbf{H}$ be a simple Chevalley $\mathbb{Z}$–group of adjoint type. In [2] Borel, Friedman and Morgan provide a considerable amount of information about the set of conjugacy classes of n-tuples $\mathbf{x} = (x_1, \cdots, x_n)$ of commuting elements of finite order of $\mathbf{H}(\mathbb{C})$.[4] The methods used in [2] are topological and analytic in nature, and do not immediately translate to other algebraically closed fields of characteristic 0. One of the reasons why this problem is relevant is because of its applications to infinite dimensional Lie theory. The interested reader can consult [6] for details and further references about this topic.
Fix an n-tuple $\mathbf{m} = (m_1, \cdots, m_n)$ of positive integers, and let $\mathbf{F_m}$ be the finite constant $\overline{\mathbb{Q}}$–group corresponding to the finite group $\mathbb{Z}/m_1\mathbb{Z} \times \cdots \times \mathbb{Z}/m_n\mathbb{Z}$. Because of the nature of our base field the group $\mathbf{F_m}$ is diagonalizable, hence linearly reductive. Let K be an algebraically closed field of characteristic 0. The conjugacy classes of n-tuples $\mathbf{x} = (x_1, \cdots, x_n)$ of commuting elements of $\mathbf{H}(K)$ where the x_i satisfy $x_i^{m_i} = 1$ are parametrized by $\overline{\mathbf{Hom}}_{K-gr}(\mathbf{F_{m,K}}, \mathbf{H}_K)(K)$. By Theorem 1.1 we have natural bijections

$$\overline{\mathbf{Hom}}_{K-gr}(\mathbf{F_{m,K}}, \mathbf{H}_K)(K) \simeq \overline{\mathbf{Hom}}_{\overline{\mathbb{Q}}-gr}(\mathbf{F_m}, \mathbf{H}_{\overline{\mathbb{Q}}})(\overline{\mathbb{Q}})) \simeq \\ \simeq \overline{\mathbf{Hom}}_{\mathbb{C}-gr}(\mathbf{F}_{\mathbf{m},\mathbb{C}}, \mathbf{H}_{\mathbb{C}})(\mathbb{C}).$$

This allows us to translate the relevant information within [2] to the group $\mathbf{H}(K)$.

5.2. Lie algebras. Assume henceforth that the base scheme S is of "characteristic zero", i.e. that S is a scheme over $\mathbb{Q}$. Let $\mathbf{G}/S$ be a semisimple group scheme and let $\mathbf{H}/S$ be an affine smooth group scheme. In this case, we already know that the functor $\mathbf{Hom}_{S-gp}(\mathbf{G}, \mathbf{H})$ is representable by a smooth affine S-scheme of finite presentation [17, XXIV.7.3.1]. Furthermore, if $\mathbf{G}/S$ is simply connected, we have an S-scheme isomorphism

$$\mathbf{Hom}_{S-gr}(\mathbf{G}, \mathbf{H}) \xrightarrow{\sim} \mathbf{Hom}_{S-Lie}\big(Lie(\mathbf{G}), Lie(\mathbf{H})\big).$$

[4]When lifted to the simply connected cover of $\mathbf{H}(\mathbb{C})$ the n-tuples will be comprised of "almost commuting" elements.

From this and Theorem 4.5 it follows that the functor

$$T \mapsto \mathrm{Hom}_{T-Lie}\big(\mathrm{Lie}(\mathbf{G}_T), \mathrm{Lie}(\mathbf{H}_T)\big)/\mathbf{H}(T)$$

is formally étale.

COROLLARY 5.4. *Let k be an algebraically closed field of characteristic zero. Let $\mathfrak{g}$ be a finite dimensional semisimple Lie algebra over k. Let $\mathbf{H}$ be a smooth algebraic k–group. If K is an algebraically closed field extension of k, then the map*

$$\mathrm{Hom}_{k-\mathrm{Lie}}(\mathfrak{g}, \mathrm{Lie}(\mathbf{H}))/\mathbf{H}(k) \to \mathrm{Hom}_{K-\mathrm{Lie}}(\mathfrak{g} \otimes_k K, \mathrm{Lie}(\mathbf{H}) \otimes_k K)/\mathbf{H}(K)$$

is bijective. □

6. APPENDIX: AFFINE GROUP SCHEMES WHICH ARE ESSENTIALLY FREE

DEFINITION 6.1. [17, VIII 6] *A morphism of schemes X/S is essentially free if there exists an open covering of S by affine schemes $S_i = \mathrm{Spec}(A_i)$, and for all i an faithfully flat extension $S_i' = \mathrm{Spec}(A_i') \to S_i$ such that each $X \times_S S_i'$ admits an open covering by affine schemes $(\mathrm{Spec}(B_{i,j}'))$ such that $B_{i,j}'$ is a free A_i'–module for all j.*

Note that an essentially free morphism is flat. Furthermore this property is stable by arbitrary base change and is local with respect to the Zariski and the *fpqc* topology. Recall that that a sequence

$$1 \to \mathbf{G}_1 \to \mathbf{G}_2 \to \mathbf{G}_3 \to 1$$

of S–group schemes is said to be exact if it is exact as a sequence of *fpqc*-sheaves over S [17, VI_B 9].

LEMMA 6.2.
(1) *Let $\mathbf{G}/S$ be a flat group scheme which is essentially free over S. Let $X \to S$ be a $\mathbf{G}$–torsor which is locally trivial for the fpqc–topology. Then X is essentially free over S.*
(2) *Let $1 \to \mathbf{G}_1 \to \mathbf{G}_2 \to \mathbf{G}_3 \to 1$ be an exact sequence of S–group schemes. If $\mathbf{G}_1$ and $\mathbf{G}_3$ are essentially free over S, then $\mathbf{G}_2$ is essentialy free over S.*

Proof. (1) Since we can reason locally for the Zariski and for the *fpqc* topology, we can assume that X is the trivial torsor, namely $X = \mathbf{G}$.

(2) Similarly, we can assume that $S = \mathrm{Spec}(A)$ and that $\mathbf{G}_1$ (resp. $\mathbf{G}_3$) is covered by open affines subschemes $\mathrm{Spec}(B_j)$ (resp. $\mathrm{Spec}(C_l)$) for the *fpqc* topology such that the B_j and the C_l are free A–modules. Up to refining the second *fpqc* covering, we can furthermore assume that

$$\mathbf{G}_2 \times_{\mathbf{G}_3} \mathrm{Spec}(C_l) \xrightarrow{\sim} \mathbf{G}_1 \times_S \mathrm{Spec}(C_l).$$

It follows that the $\mathrm{Spec}(B_l \otimes_A C_l)$'s form a *fpqc*–cover of $\mathbf{G}_2$ where the $B_l \otimes_A C_l$ are free A–modules. Thus $\mathbf{G}_2$ is essentially free over S as desired. □

Several well-known affine group schemes are essentially free.

Proposition 6.3. *Let* $\mathbf{G}/S$ *be an affine* S*–group scheme which admits locally for the* $fpqc$ *topology a composition series with factors of the following kind:*

(i) S*-group schemes of multiplicative type,*

(ii) twisted finite constant S*–group schemes,*

(iii) smooth S*-group schemes with connected geometric fibers.*

Then $\mathbf{G}$ *is essentially free over* S*.*

Note that the last case includes reductive group schemes over S and their parabolic subgroups.

Proof. By Lemma 6.2 it suffices to verify each of the three cases locally for the $fpqc$ topology. Case (i) is then the case of diagonalizable groups which are essentially free over S by definition. Case (ii) is that of finite constant S–group schemes which are also essentially free over S by definition. Case (iii) has been noticed by Seshadri [15, Lemma 1 p. 230] using a result of Raynaud. □

Acknowledgements: The author would like to thank the referee for his/her thorough reading of the original submission. He would also like to express his gratitude to BIRS for their hospitality.

References

[1] A. Borel, *Linear Algebraic Groups (Second enlarged edition)*, Graduate text in Mathematics **126** (1991), Springer.

[2] A. Borel, R. Friedman and J. W. Morgan, *Almost commuting elements in compact Lie groups* Mem. Amer. Math. Soc. 747 (2002).

[3] S. Bosch, W. Lütkebohmert and M. Raynaud, *Néron Models*, Ergebnisse der Mathematik und ihrer Grenzgebiete **21** (1990) Springer-Verlag.

[4] M. Demazure, P. Gabriel, *Groupes algébriques. Tome I: Géométrie algébrique, généralités, groupes commutatifs*, Masson (1970).

[5] S. Donkin, *On the existence of Auslander-Reiten sequences of group representations I*, Algebras and Representation Theory **1** (1998), 97–127.

[6] P. Gille and A. Pianzola, *Galois cohomology and forms of algebras over Laurent polynomial rings*, Math. Annalen **338** 497–543 (2007)

[7] A. Grothendieck (avec la collaboration de J. Dieudonné), *Eléments de Géométrie Algébrique I*, Publications mathématiques de l'I.H.É.S. no 4 (1960).

[8] A. Grothendieck (avec la collaboration de J. Dieudonné), *Eléments de Géométrie Algébrique III*, Publications mathématiques de l'I.H.É.S. no 11, 17 (1962 - 1963).

[9] A. Grothendieck (avec la collaboration de J. Dieudonné), *Eléments de Géométrie Algébrique IV*, Publications mathématiques de l'I.H.É.S. no 20, 24, 28 and 32 (1964 - 1967).

[10] J.-C. Jantzen, *Representations of algebraic groups*, Second edition, Mathematical Surveys and Monographs **107** (2003), American Mathematical Society.

[M] J.S. Milne, *Étale cohomology*, Princeton University Press (1980).

[11] M. Nagata, *Complete reducibility of rational representations of a matrix group*, J. Math. Kyoto Univ. **1** (1961/1962), 87–99.

[12] R. W. Richardson, *Deformations of Lie subgroups*, Bull. Amer. Math. soc. **77** (1971), 92–96.

[13] R. W. Richardson, *Deformations of Lie subgroups and the variation of isotropy subgroups*, Acta Math. **129** (1972), 35–73.

[14] J.-P. Serre, *Groupe de Grothendieck des schémas en groupes réductifs déployés*, Publications Mathématiques de l'IHES **34** (1968), 37-52.

[15] C. S. Seshadri, *Geometric reductivity over arbitrary base*, Advances in Math. **26** (1977), 225–274.

[16] P. Slodowy, *Two notes on a finiteness problem in the representation theory of finite groups*, With an appendix by G.-Martin Cram. Austral. Math. Soc. Lect. Ser. **9** (1997), Algebraic groups and Lie groups, 331–348, Cambridge Univ. Press, Cambridge.

[17] *Séminaire de Géométrie algébrique de l'I.H.E.S., 1963-1964, schémas en groupes, dirigé par M. Demazure et A. Grothendieck*, Lecture Notes in Math. 151-153. Springer (1970).

[18] P. Tauvel, R. W. T. Yu., *Lie algebras and algebraic groups*, Springer Monographs in Mathematics, Springer-Verlag (2006).

[19] E. B. Vinberg, *On invariants of a set of matrices*, J. Lie Theory **6** (1996), 249–269.

Benedictus Margaux
Laboratoire de Recherche
"Princess Stephanie"
Monte Carlo 51840, Monaco.
benedictus.margaux@gmail.com

On Proper $\mathbb{R}$-Actions on Hyperbolic Stein Surfaces

Christian Miebach, Karl Oeljeklaus[(1)]

Received: April 18, 2009
Revised: August 3, 2009

Communicated by Thomas Peternell

Abstract. In this paper we investigate proper $\mathbb{R}$–actions on hyperbolic Stein surfaces and prove in particular the following result: Let $D \subset \mathbb{C}^2$ be a simply-connected bounded domain of holomorphy which admits a proper $\mathbb{R}$–action by holomorphic transformations. The quotient $D/\mathbb{Z}$ with respect to the induced proper $\mathbb{Z}$–action is a Stein manifold. A normal form for the domain D is deduced.

2000 Mathematics Subject Classification: 32E10; 32M05; 32Q45; 32T05
Keywords and Phrases: Stein manifolds; bounded domains of holomorphy; proper actions; quotient by a discrete group

1. Introduction

Let X be a Stein manifold endowed with a real Lie transformation group G of holomorphic automorphisms. In this situation it is natural to ask whether there exists a G–invariant holomorphic map $\pi\colon X \to X/\!\!/G$ onto a complex space $X/\!\!/G$ such that $\mathcal{O}_{X/\!\!/G} = (\pi_*\mathcal{O}_X)^G$ and, if yes, whether this quotient $X/\!\!/G$ is again Stein. If the group G is compact, both questions have a positive answer as is shown in [Hei91].

For non-compact G even the existence of a complex quotient in the above sense of X by G cannot be guaranteed. In this paper we concentrate on the most basic and already non-trivial case $G = \mathbb{R}$. We suppose that G acts properly on X. Let $\Gamma = \mathbb{Z}$. Then X/Γ is a complex manifold and if, moreover, it is Stein, we can define $X/\!\!/G := (X/\Gamma)/\!\!/(G/\Gamma)$. The following was conjectured by Alan Huckleberry.

Let X be a contractible bounded domain of holomorphy in $\mathbb{C}^n$ with a proper action of $G = \mathbb{R}$. Then the complex manifold $X/\mathbb{Z}$ is Stein.

(1) The authors would like to thank Peter Heinzner and Jean-Jacques Loeb for numerous discussions on the subject.

In [FI01] this conjecture is proven for the unit ball and in [Mie08] for arbitrary bounded homogeneous domains in $\mathbb{C}^n$. In this paper we make a first step towards a proof in the general case by showing

Theorem 1.1. — *Let D be a simply-connected bounded domain of holomorphy in $\mathbb{C}^2$. Suppose that the group $\mathbb{R}$ acts properly by holomorphic transformations on D. Then the complex manifold $D/\mathbb{Z}$ is Stein. Moreover, $D/\mathbb{Z}$ is biholomorphically equivalent to a domain of holomorphy in $\mathbb{C}^2$.*

As an application of this theorem we deduce a normal form for domains of holomorphy whose identity component of the automorphism group is non-compact as well as for proper $\mathbb{R}$–actions on them. Notice that we make no assumption on smoothness of their boundaries.
We first discuss the following more general situation. Let X be a hyperbolic Stein manifold with a proper $\mathbb{R}$–action. Then there is an induced local holomorphic $\mathbb{C}$–action on X which can be globalized in the sense of [HI97]. The following result is central for the proof of the above theorem.

Theorem 1.2. — *Let X be a hyperbolic Stein surface with a proper $\mathbb{R}$–action. Suppose that either X is taut or that it admits the Bergman metric and $H^1(X,\mathbb{R}) = 0$. Then the universal globalization X^* of the induced local $\mathbb{C}$–action is Hausdorff and $\mathbb{C}$ acts properly on X^*. Furthermore, for simply-connected X one has that $X^* \to X^*/\mathbb{C}$ is a holomorphically trivial $\mathbb{C}$–principal bundle over a simply-connected Riemann surface.*

Finally, we discuss several examples of hyperbolic Stein manifolds X with proper $\mathbb{R}$–actions such that $X/\mathbb{Z}$ is not Stein. If one does not require the existence of an $\mathbb{R}$–action, there are bounded Reinhardt domains in $\mathbb{C}^2$ with proper $\mathbb{Z}$–actions for which the quotients are not Stein.

2. Hyperbolic Stein $\mathbb{R}$–manifolds

In this section we present the general set-up.

2.1. The induced local $\mathbb{C}$–action and its globalization. — Let X be a hyperbolic Stein manifold. It is known that the group $\mathrm{Aut}(X)$ of holomorphic automorphisms of X is a real Lie group with respect to the compact-open topology which acts properly on X (see [Kob98]). Let $\{\varphi_t\}_{t\in\mathbb{R}}$ be a closed one parameter subgroup of $\mathrm{Aut}(D)$. Consequently, the action $\mathbb{R}\times X \to X$, $t\cdot x := \varphi_t(x)$, is proper. By restriction, we obtain also a proper $\mathbb{Z}$–action on X. Since every such action must be free, the quotient $X/\mathbb{Z}$ is a complex manifold. This complex manifold $X/\mathbb{Z}$ carries an action of $S^1 \cong \mathbb{R}/\mathbb{Z}$ which is induced by the $\mathbb{R}$–action on X.
Integrating the holomorphic vector field on X which corresponds to this $\mathbb{R}$–action we obtain a local $\mathbb{C}$–action on X in the following sense. There are an open neighborhood $\Omega \subset \mathbb{C}\times X$ of $\{0\}\times X$ and a holomorphic map $\Phi\colon \Omega \to X$, $\Phi(t,x) =: t\cdot x$, such that the following holds:

(1) For every $x \in X$ the set $\Omega(x) := \{t \in \mathbb{C};\ (t,x) \in \Omega\} \subset \mathbb{C}$ is connected;
(2) for all $x \in X$ we have $0 \cdot x = x$;
(3) we have $(t + t') \cdot x = t \cdot (t' \cdot x)$ whenever both sides are defined.

Following [Pal57] (compare [HI97] for the holomorphic setting) we say that a globalization of the local $\mathbb{C}$–action on X is an open $\mathbb{R}$–equivariant holomorphic embedding $\iota\colon X \hookrightarrow X^*$ into a (not necessarily Hausdorff) complex manifold X^* endowed with a holomorphic $\mathbb{C}$–action such that $\mathbb{C}\cdot\iota(X) = X^*$. A globalization $\iota\colon X \hookrightarrow X^*$ is called *universal* if for every $\mathbb{R}$–equivariant holomorphic map $f\colon X \to Y$ into a holomorphic $\mathbb{C}$–manifold Y there exists a holomorphic $\mathbb{C}$–equivariant map $F\colon X^* \to Y$ such that the diagram

$$\begin{tikzcd} X \arrow[rr, "\iota"] \arrow[dr, "f"'] & & X^* \arrow[dl, "F"] \\ & Y & \end{tikzcd}$$

commutes. It follows that a universal globalization is unique up to isomorphism if it exists.
Since X is Stein, the universal globalization X^* of the induced local $\mathbb{C}$–action exists as is proven in [HI97]. We will always identify X with its image $\iota(X) \subset X^*$. Then the local $\mathbb{C}$–action on X coincides with the restriction of the global $\mathbb{C}$–action on X^* to X.
Recall that X is said to be orbit-connected in X^* if for every $x \in X^*$ the set $\Sigma(x) := \{t \in \mathbb{C};\ t \cdot x \in X\}$ is connected. The following criterion for a globalization to be universal is proven in [CTIT00].

Lemma 2.1. — Let X^* be any globalization of the induced local $\mathbb{C}$–action on X. Then X^* is universal if and only if X is orbit-connected in X^*.

Remark. — The results about (universal) globalizations hold for a bigger class of groups ([CTIT00]). However, we will need it only for the groups $\mathbb{C}$ and $\mathbb{C}^*$ and thus will not give the most general formulation.

For later use we also note the following

Lemma 2.2. — The $\mathbb{C}$–action on X^* is free.

Proof. — Suppose that there exists a point $x \in X^*$ such that $\mathbb{C}_x$ is non-trivial. Because of $\mathbb{C} \cdot X = X^*$ we can assume that $x \in X$ holds. Since $\mathbb{C}_x$ is a non-trivial closed subgroup of $\mathbb{C}$, it is either a lattice of rank 1 or 2, or $\mathbb{C}$. The last possibility means that x is a fixed point under $\mathbb{C}$ which is not possible since $\mathbb{R}$ acts freely on X.
We observe that the lattice $\mathbb{C}_x$ is contained in the connected $\mathbb{R}$–invariant set $\Sigma(x) = \{t \in \mathbb{C};\ t \cdot x \in X\}$. By $\mathbb{R}$–invariance $\Sigma(x)$ is a strip. Since X is hyperbolic, this strip cannot coincide with $\mathbb{C}$. The only lattice in $\mathbb{C}$ which can possibly be contained in such a strip is of the form $\mathbb{Z}r$ for some $r \in \mathbb{R}$. Since this contradicts the fact that $\mathbb{R}$ acts freely on X, the lemma is proven. $\square$

Note that we do not know whether X^* is Hausdorff. In order to guarantee the Hausdorff property of X^*, we make further assumptions on X. The following result is proven in [Ian03] and [IST04].

Theorem 2.3. — Let X be a hyperbolic Stein manifold with a proper $\mathbb{R}$–action. Suppose in addition that X is taut or admits the Bergman metric. Then X^* is Hausdorff. If X is simply-connected, then the same is true for X^*.

We refer the reader to Chapter 5 in [Kob98] for the definition and examples of tautness. For the reader's convenience we describe here the construction of the Bergman metric for an arbitrary n–dimensional complex manifold X. For more details see Chapter 4.10 in [Kob98]. The space $\mathcal{A}^2(X)$ of square integrable holomorphic n–forms on X is a separable complex Hilbert space with respect to the inner product $\langle \omega_1, \omega_2 \rangle := i^{n^2} \int_X \omega_1 \wedge \overline{\omega_2}$. Let $\omega_1, \omega_2, \ldots$ be an orthonormal basis of $\mathcal{A}^2(X)$ and define $B_X := \sum_{j \geq 1} i^{n^2} \omega_j \wedge \overline{\omega_j}$. The non-negative (n,n)–form B_X is independent of the chosen basis and is called the Bergman kernel form of X. Suppose that B_X is positive, i. e. that for every $x \in X$ there exists $\omega \in \mathcal{A}^2(X)$ with $\omega_x \neq 0$. Then we may define the map $\iota \colon X \to \mathbb{P}\big(\mathcal{A}^2(X)^*\big)$ which associates to each $x \in X$ the hyperplane consisting of forms in $\mathcal{A}^2(X)$ which vanish at x. By definition one says that X *admits the Bergman metric* if this map ι is an immersion. The Bergman metric of X is then defined as the pull-back of the Fubini-Study metric of $\mathbb{P}\big(\mathcal{A}^2(X)^*\big)$.

Remark. — Every bounded domain in $\mathbb{C}^n$ admits the Bergman metric.

2.2. The quotient $X/\mathbb{Z}$. — We assume from now on that X fulfills the hypothesis of Theorem 2.3. Since X^* is covered by the translates $t \cdot X$ for $t \in \mathbb{C}$ and since the action of $\mathbb{Z}$ on each domain $t \cdot X$ is proper, we conclude that the quotient $X^*/\mathbb{Z}$ fulfills all axioms of a complex manifold except for possibly not being Hausdorff.

We have the following commutative diagram:

$$\begin{tikzcd} X \arrow[r] \arrow[d] & X^* \arrow[d] \\ X/\mathbb{Z} \arrow[r] & X^*/\mathbb{Z}. \end{tikzcd}$$

Note that the group $\mathbb{C}^* = (S^1)^{\mathbb{C}} \cong \mathbb{C}/\mathbb{Z}$ acts on $X^*/\mathbb{Z}$. Concretely, if we identify $\mathbb{C}/\mathbb{Z}$ with $\mathbb{C}^*$ via $\mathbb{C} \to \mathbb{C}^*$, $t \mapsto e^{2\pi i t}$, the quotient map $p \colon X^* \to X^*/\mathbb{Z}$ fulfills $p(t \cdot x) = e^{2\pi i t} \cdot p(x)$.

Lemma 2.4. — The induced map $X/\mathbb{Z} \hookrightarrow X^*/\mathbb{Z}$ is the universal globalization of the local $\mathbb{C}^*$–action on $X/\mathbb{Z}$.

Proof. — The open embedding $X \hookrightarrow X^*$ induces an open embedding $X/\mathbb{Z} \hookrightarrow X^*/\mathbb{Z}$. This embedding is S^1–equivariant and we have $\mathbb{C}^* \cdot X/\mathbb{Z} = X^*/\mathbb{Z}$. This implies that $X^*/\mathbb{Z}$ is a globalization of the local $\mathbb{C}^*$–action on $X/\mathbb{Z}$.

In order to prove that this globalization is universal, by the globalization theorem in [CTIT00] it is enough to show that $X/\mathbb{Z}$ is orbit-connected in $X^*/\mathbb{Z}$. Hence, we must show that for every $[x] \in X/\mathbb{Z}$ the set $\Sigma\big([x]\big) := \{t \in \mathbb{C}^*;\ t\cdot[x] \in X/\mathbb{Z}\}$ is connected in $\mathbb{C}^*$. For this we consider the set $\Sigma(x) = \{t \in \mathbb{C};\ t\cdot x \in X\}$. Since the map $X \to X/\mathbb{Z}$ intertwines the local $\mathbb{C}$– and $\mathbb{C}^*$–actions, we conclude that $t \in \Sigma(x)$ holds if and only if $e^{2\pi i t} \in \Sigma\big([x]\big)$ holds. Since X^* is universal, $\Sigma(x)$ is connected which implies that $\Sigma\big([x]\big)$ is likewise connected. Thus $X^*/\mathbb{Z}$ is universal. $\square$

Remark. — The globalization $X^*/\mathbb{Z}$ is Hausdorff if and only if $\mathbb{Z}$ or, equivalently, $\mathbb{R}$ act properly on X^*. As we shall see in Lemma 3.3, this is the case if X is taut.

2.3. A sufficient condition for $X/\mathbb{Z}$ to be Stein. — If $\dim X = 2$, we have the following sufficient condition for $X/\mathbb{Z}$ to be a Stein surface.

Proposition 2.5. — If the $\mathbb{C}$–action on X^* is proper and if the Riemann surface $X^*/\mathbb{C}$ is not compact, then $X/\mathbb{Z}$ is Stein.

Proof. — Under the above hypothesis we have the $\mathbb{C}$–principal bundle $X^* \to X^*/\mathbb{C}$. If the base $X^*/\mathbb{C}$ is not compact, then this bundle is holomorphically trivial, i. e. X^* is biholomorphic to $\mathbb{C} \times R$ where R is a non-compact Riemann surface. Since R is Stein, the same is true for X^* and for $X^*/\mathbb{Z} \cong \mathbb{C}^* \times R$. Since $X/\mathbb{Z}$ is locally Stein, see [Mie08], in the Stein manifold $X^*/\mathbb{Z}$, the claim follows from [DG60]. $\square$

Therefore, the crucial step in the proof of our main result consists in showing that $\mathbb{C}$ acts properly on X^* under the assumption $\dim X = 2$.

3. Local properness

Let X be a hyperbolic Stein $\mathbb{R}$–manifold. Suppose that X is taut or that it admits the Bergman metric and $H^1(X, \mathbb{R}) = \{0\}$. We show that then $\mathbb{C}$ acts locally properly on X^*.

3.1. Locally proper actions. — Recall that the action of a Lie group G on a manifold M is called locally proper if every point in M admits a G–invariant open neighborhood on which G acts properly.

Lemma 3.1. — Let $G \times M \to M$ be locally proper.

(1) *For every $x \in M$ the isotropy group G_x is compact.*
(2) *Every G–orbit admits a geometric slice.*
(3) *The orbit space M/G is a smooth manifold which is in general not Hausdorff.*
(4) *All G–orbits are closed in M.*
(5) *The G–action on M is proper if and only if M/G is Hausdorff.*

Proof. — The first claim is elementary to check. The second claim is proven in [DK00]. The third one is a consequence of (2) since the slices yield charts on M/G which are smoothly compatible because the transitions are given by the smooth action of G on M. Assertion (4) follows from (3) because in locally Euclidian topological spaces points are closed. The last claim is proven in [Pal61]. $\square$

Remark. — Since $\mathbb{R}$ acts properly on X and $\mathbb{C} \cdot X = X^*$, the $\mathbb{R}$-action on X^* is locally proper.

3.2. Local properness of the $\mathbb{C}$–action on X^*. — Recall that we assume that

$$X \text{ is taut} \tag{3.1}$$

or that

$$X \text{ admits the Bergman metric and } H^1(X, \mathbb{R}) = \{0\}. \tag{3.2}$$

We first show that assumption (3.1) implies that $\mathbb{C}$ acts locally properly on X^*.

Since X^* is the universal globalization of the induced local $\mathbb{C}$–action on X, we know that X is orbit-connected in X^*. This means that for every $x \in X^*$ the set $\Sigma(x) = \{t \in \mathbb{C};\ t \cdot x \in X\}$ is a strip in $\mathbb{C}$. In the following we will exploit the properties of the thickness of this strip.

Since $\Sigma(x)$ is $\mathbb{R}$–invariant, there are "numbers" $u(x) \in \mathbb{R} \cup \{-\infty\}$ and $o(x) \in \mathbb{R} \cup \{\infty\}$ for every $x \in X^*$ such that

$$\Sigma(x) = \big\{t \in \mathbb{C};\ u(x) < \mathrm{Im}(t) < o(x)\big\}.$$

The functions $u\colon X^* \to \mathbb{R}\cup\{-\infty\}$ and $o\colon X^* \to \mathbb{R}\cup\{\infty\}$ so obtained are upper and lower semicontinuous, respectively. Moreover, u und o are $\mathbb{R}$–invariant and $i\mathbb{R}$–equivariant:

$$u(it \cdot x) = u(x) - t \quad \text{and} \quad o(it \cdot x) = o(x) - t.$$

Proposition 3.2. — The functions $u, -o\colon X^* \to \mathbb{R} \cup \{-\infty\}$ are plurisubharmonic. Moreover, u and o are continuous on $X^*\setminus\{u = -\infty\}$ and $X^*\setminus\{o = \infty\}$, respectively.

Proof. — It is proven in [For96] that u and $-o$ are plurisubharmonic on X. By equivariance, we obtain this result for X^*.

Now we prove that the function $u\colon X\setminus\{u = -\infty\} \to \mathbb{R}$ is continuous which was remarked without complete proof in [Ian03]. For this let (x_n) be a sequence in X which converges to $x_0 \in X\setminus\{u = -\infty\}$. Since u is upper semi-continuous, we have $\limsup_{n\to\infty} u(x_n) \leq u(x_0)$. Suppose that u is not continuous in x_0. Then, after replacing (x_n) by a subsequence, we find $\varepsilon > 0$ such that $u(x_n) \leq u(x_0) - \varepsilon < u(x_0)$ holds for all $n \in \mathbb{N}$. Consequently, we have $\Sigma(x_0) = \big\{t \in \mathbb{C};\ u(x_0) < \mathrm{Im}(t) < o(x_0)\big\} \subset \Sigma := \big\{t \in \mathbb{C};\ u(x_0) - \varepsilon < \mathrm{Im}(t) < o(x_0)\big\} \subset \Sigma(x_n)$ for all n and hence obtain the sequence of holomorphic functions $f_n\colon \Sigma \to X$, $f_n(t) := t{\cdot}x_n$. Since X is taut and $f_n(0) = x_n \to x_0$, the sequence (f_n) has a subsequence which compactly converges to a holomorphic function $f_0\colon \Sigma \to X$. Because of $f_0\big(iu(x_0)\big) = \lim_{n\to\infty} f_n\big(iu(x_0)\big) = \lim_{n\to\infty} iu(x_0)\cdot x_n = iu(x_0)\cdot x_0 \notin X$ we arrive at a contradiction. Thus the function $u\colon X\setminus\{u=-\infty\}\to\mathbb{R}$ is continuous. By $(i\mathbb{R})$–equivariance, u is also continuous on $X^*\setminus\{u=-\infty\}$. A similar argument shows continuity of $-o\colon X^*\setminus\{o=\infty\}\to\mathbb{R}$. □

Let us consider the sets

$$\mathcal{N}(o) := \big\{x\in X^*;\ o(x)=0\big\} \quad\text{and}\quad \mathcal{P}(o) := \big\{x\in X^*;\ o(x)=\infty\big\}.$$

The sets $\mathcal{N}(u)$ and $\mathcal{P}(u)$ are similarly defined. Since $X=\big\{x\in X^*;\ u(x)<0<o(x)\big\}$, we can recover X from X^* with the help of u and o.

Lemma 3.3. — The action of $\mathbb{R}$ on X^* is proper.

Proof. — Let ∂^*X denote the boundary of X in X^*. Since the functions u and $-o$ are continuous on $X^*\setminus\mathcal{P}(u)$ and $X^*\setminus\mathcal{P}(o)$ one verifies directly that $\partial^*X=\mathcal{N}(u)\cup\mathcal{N}(o)$ holds. As a consequence, we note that if $x\in\partial^*X$, then for every $\varepsilon>0$ the element $(i\,\varepsilon)\cdot x$ is not contained in ∂^*X.
Let (t_n) and (x_n) be sequences in $\mathbb{R}$ and X^* such that $(t_n\cdot x_n, x_n)$ converges to (y_0,x_0) in $X^*\times X^*$. We may assume without loss of generality that x_0 and hence x_n are contained in X for all n. Consequently, we have $y_0\in X\cup\partial^*X$. If $y_0\in\partial^*X$ holds, we may choose an $\varepsilon>0$ such that $(i\,\varepsilon)\cdot y_0$ and $(i\,\varepsilon)\cdot x_0$ lie in X. Since the $\mathbb{R}$–action on X is proper, we find a convergent subsequence of (t_n) which was to be shown. □

Lemma 3.4. — We have:

1. *$\mathcal{N}(u)$ and $\mathcal{N}(o)$ are $\mathbb{R}$–invariant.*
2. *We have $\mathcal{N}(u)\cap\mathcal{N}(o)=\emptyset$.*
3. *The sets $\mathcal{P}(u)$ and $\mathcal{P}(o)$ are closed, $\mathbb{C}$–invariant and pluripolar in X^*.*
4. *$\mathcal{P}(u)\cap\mathcal{P}(o)=\emptyset$.*

Proof. — The first claim follows from the $\mathbb{R}$–invariance of u and o.
The second claim follows from $u(x)<o(x)$.
The third one is a consequence of the $\mathbb{R}$–invariance and $i\mathbb{R}$–equivariance of u and o.
If there was a point $x\in\mathcal{P}(u)\cap\mathcal{P}(o)$, then $\mathbb{C}\cdot x$ would be a subset of X which is impossible since X is hyperbolic. □

Lemma 3.5. — If o is not identically ∞, then the map

$$\varphi\colon i\mathbb{R}\times\mathcal{N}(o)\to X^*\setminus\mathcal{P}(o),\quad \varphi(it,z)=it\cdot z,$$

is an $i\mathbb{R}$–equivariant homeomorphism. Since $\mathbb{R}$ acts properly on $\mathcal{N}(o)$, it follows that $\mathbb{C}$ acts properly on $X^*\setminus\mathcal{P}(o)$. The same holds when o is replaced by u.

Proof. — The inverse map φ^{-1} is given by $x\mapsto\big(-io(x), io(x)\cdot x\big)$. □

Corollary 3.6. — The $\mathbb{C}$–action on X^* is locally proper. If $\mathcal{P}(o)=\emptyset$ or $\mathcal{P}(u)=\emptyset$ hold, then $\mathbb{C}$ acts properly on X^*.

From now on we suppose that X fulfills the assumption (3.2). Recall that the Bergman form ω is a Kähler form on X invariant under the action of $\mathrm{Aut}(X)$. Let ξ denote the complete holomorphic vector field on X which corresponds to the $\mathbb{R}$–action, i.e. we have $\xi(x) = \frac{\partial}{\partial t}\big|_0 \varphi_t(x)$. Hence, $\iota_\xi\omega = \omega(\cdot, \xi)$ is a 1–form on X and since $H^1(X, \mathbb{R}) = \{0\}$ there exists a function $\mu^\xi \in \mathcal{C}^\infty(X)$ with $d\mu^\xi = \iota_\xi\omega$.

Remark. — This means that μ^ξ is a momentum map for the $\mathbb{R}$–action on X.

Lemma 3.7. — The map $\mu^\xi\colon X \to \mathbb{R}$ is an $\mathbb{R}$–invariant submersion.

Proof. — The claim follows from $d\mu^\xi(x)J\xi_x = \omega_x(J\xi_x, \xi_x) > 0$. □

Proposition 3.8. — The $\mathbb{C}$–action on X^* is locally proper.

Proof. — Since μ^ξ is a submersion, the fibers $(\mu^\xi)^{-1}(c)$, $c \in \mathbb{R}$, are real hypersurfaces in X. Then

$$\frac{d}{dt}\bigg|_0 \mu^\xi(it \cdot x) = \omega_x(J\xi_x, \xi_x) > 0$$

implies that every $i\mathbb{R}$–orbit intersects $(\mu^\xi)^{-1}(c)$ transversally. Since X is orbit-connected in X^*, the map $i\mathbb{R} \times (\mu^\xi)^{-1}(c) \to X^*$ is injective and therefore a diffeomorphism onto its open image. Together with the fact that $(\mu^\xi)^{-1}(c)$ is $\mathbb{R}$–invariant this yields the existence of differentiable local slices for the $\mathbb{C}$–action. □

3.3. A necessary condition for $X/\mathbb{Z}$ to be Stein. — We have the following necessary condition for $X/\mathbb{Z}$ to be a Stein manifold.

Proposition 3.9. — If the quotient manifold $X/\mathbb{Z}$ is Stein, then X^* is Stein and the $\mathbb{C}$–action on X^* is proper.

Proof. — Suppose that $X/\mathbb{Z}$ is a Stein manifold. By [CTIT00] this implies that X^* is Stein as well.
Next we will show that the $\mathbb{C}^*$–action on $X^*/\mathbb{Z}$ is proper. For this we will use as above a moment map for the S^1–action on $X^*/\mathbb{Z}$.
By compactness of S^1 we may apply the complexification theorem from [Hei91] which shows that $X^*/\mathbb{Z}$ is also a Stein manifold and in particular Hausdorff. Hence, there exists a smooth strictly plurisubharmonic exhaustion function $\rho\colon X^*/\mathbb{Z} \to \mathbb{R}^{>0}$ invariant under S^1. Consequently, $\omega := \frac{i}{2}\partial\overline{\partial}\rho \in \mathcal{A}^{1,1}(X^*)$ is an S^1–invariant Kähler form. Associated to ω we have the S^1–invariant moment map

$$\mu\colon X^*/\mathbb{Z} \to \mathbb{R}, \quad \mu^\xi(x) := \frac{d}{dt}\bigg|_0 \rho\big(\exp(it\xi) \cdot x\big),$$

where ξ is the complete holomorphic vector field on $X^*/\mathbb{Z}$ which corresponds to the S^1–action. Now we can apply the same argument as above in order to deduce that $\mathbb{C}^*$ acts locally properly on $X^*/\mathbb{Z}$.

We still must show that $(X^*/\mathbb{Z})/\mathbb{C}^*$ is Hausdorff. To see this, let $\mathbb{C}^* \cdot x_j$, $j = 0, 1$, be two different orbits in $X^*/\mathbb{Z}$. Since $\mathbb{C}^*$ acts locally properly, these are closed and therefore there exists a function $f \in \mathcal{O}(X^*/\mathbb{Z})$ with $f|_{\mathbb{C}^* \cdot x_j} = j$ for $j = 0, 1$. Again we may assume that f is S^1– and consequently $\mathbb{C}^*$–invariant. Hence, there is a continuous function on $(X^*/\mathbb{Z})/\mathbb{C}^*$ which separates the two orbits, which implies that $(X^*/\mathbb{Z})/\mathbb{C}^*$ is Hausdorff. This proves that $\mathbb{C}^*$ acts properly on $X^*/\mathbb{Z}$.

Since we know already that the $\mathbb{C}$–action on X^* is locally proper, it is enough to show that $X^*/\mathbb{C}$ is Hausdorff. But this follows from the properness of the $\mathbb{C}^*$–action on $X^*/\mathbb{Z}$ since $X^*/\mathbb{C} \cong (X^*/\mathbb{Z})/\mathbb{C}^*$ is Hausdorff. □

4. Properness of the $\mathbb{C}$–action

Let X be a hyperbolic Stein $\mathbb{R}$–manifold. Suppose that X fulfills (3.1) or (3.2). We have seen that $\mathbb{C}$ acts locally properly on X^*. In this section we prove that under the additional assumption $\dim X = 2$ the orbit space $X^*/\mathbb{C}$ is Hausdorff. This implies that $\mathbb{C}$ acts properly on X^* if $\dim X = 2$.

4.1. Stein surfaces with $\mathbb{C}$–actions.

— For every function $f \in \mathcal{O}(\Delta)$ which vanishes only at the origin, we define

$$X_f := \big\{(x, y, z) \in \Delta \times \mathbb{C}^2;\ f(x)y - z^2 = 1\big\}.$$

Since the differential of the defining equation of X_f is given by $\big(f'(x)y\ f(x)\ -2z\big)$, we see that 1 is a regular value of $(x, y, z) \mapsto f(x)y - z^2$. Hence, X_f is a smooth Stein surface in $\Delta \times \mathbb{C}^2$.

There is a holomorphic $\mathbb{C}$–action on X_f defined by

$$t \cdot (x, y, z) := \big(x, y + 2tz + t^2 f(x), z + tf(x)\big).$$

Lemma 4.1. — The $\mathbb{C}$–action on X_f is free, and all orbits are closed.

Proof. — Let $t \in \mathbb{C}$ such that $\big(x, y + 2tz + t^2 f(x), z + tf(x)\big) = (x, y, z)$ for some $(x, y, z) \in X_f$. If $f(x) \neq 0$, then $z + tf(x) = z$ implies $t = 0$. If $f(x) = 0$, then $z \neq 0$ and $y + 2tz = y$ gives $t = 0$.

The map $\pi\colon X_f \to \Delta$, $(x, y, z) \mapsto x$, is $\mathbb{C}$–invariant. If $a \in \Delta^*$, then $f(a) \neq 0$ and we have

$$\frac{z}{f(a)} \cdot \big(a, f(a)^{-1}, 0\big) = (a, y, z) \in X_f,$$

which implies $\pi^{-1}(a) = \mathbb{C} \cdot \big(a, f(a)^{-1}, 0\big)$. A similar calculation gives $\pi^{-1}(0) = \mathbb{C} \cdot p_1 \cup \mathbb{C} \cdot p_2$ with $p_1 = (0, 0, i)$ and $p_2 = (0, 0, -i)$. Consequently, every $\mathbb{C}$–orbit is closed. □

Remark. — The orbit space $X_f/\mathbb{C}$ is the unit disc with a doubled origin and in particular not Hausdorff.

We calculate slices at the point p_j, $j = 1, 2$, as follows. Let $\varphi_j\colon \Delta \times \mathbb{C} \to X_f$ be given by $\varphi_1(z,t) := t \cdot (z, 0, i)$ and $\varphi_2(w,s) = s \cdot (w, 0, -i)$. Solving the equation $s \cdot (w, 0, -i) = t \cdot (z, 0, i)$ for (w, s) yields the transition function $\varphi_{12} = \varphi_2^{-1} \circ \varphi_1 \colon \Delta^* \times \mathbb{C} \to \Delta^* \times \mathbb{C}$,

$$(z,t) \mapsto \left(z, t + \frac{2i}{f(z)}\right).$$

The function $\frac{1}{f}$ is a meromorphic function on Δ without zeros and with the unique pole 0.

Lemma 4.2. — Let $\mathbb{R}$ act on X_f via $\mathbb{R} \hookrightarrow \mathbb{C}$, $t \mapsto ta$, for some $a \in \mathbb{C}^*$. Then there is no $\mathbb{R}$–invariant domain $D \subset X_f$ with $D \cap \mathbb{C} \cdot p_j \neq \emptyset$ for $j = 1, 2$ on which $\mathbb{R}$ acts properly.

Proof. — Suppose that $D \subset X_f$ is an $\mathbb{R}$–invariant domain with $D \cap \mathbb{C} \cdot p_j \neq \emptyset$ for $j = 1, 2$. Without loss of generality we may assume that $p_1 \in D$ and $\zeta \cdot p_2 = (0, -2\zeta i, -i) \in D$ for some $\zeta \in \mathbb{C}$. We will show that the orbits $\mathbb{R} \cdot p_1$ and $\mathbb{R} \cdot (\zeta \cdot p_2)$ cannot be separated by $\mathbb{R}$–invariant open neighborhoods.
Let $U_1 \subset D$ be an $\mathbb{R}$–invariant open neighborhood of p_1. Then there are $r, r' > 0$ such that $\Delta_r^* \times \Delta_{r'} \times \{i\} \subset U_1$ holds. Here, $\Delta_r = \{z \in \mathbb{C};\ |z| < r\}$. For $(\varepsilon_1, \varepsilon_2) \in \Delta_r^* \times \Delta_{r'}$ and $t \in \mathbb{R}$ we have

$$t \cdot (\varepsilon_1, \varepsilon_2, i) = \big(\varepsilon_1, \varepsilon_2 + 2(ta)i + (ta)^2 f(\varepsilon_1), i + (ta) f(\varepsilon_1)\big) \in U_1.$$

We have to show that for all $r_2, r_3 > 0$ there exist $(\widetilde{\varepsilon_2}, \widetilde{\varepsilon_3}) \in \Delta_{r_2} \times \Delta_{r_3}$, $(\varepsilon_1, \varepsilon_2) \in \Delta_r^* \times \Delta_{r'}$ and $t \in \mathbb{R}$ such that

$$\big(\varepsilon_1, \varepsilon_2 + 2(ta)i + (ta)^2 f(\varepsilon_1), i + (ta) f(\varepsilon_1)\big) = (\varepsilon_1, -2\zeta i + \widetilde{\varepsilon_2}, -i + \widetilde{\varepsilon_3}) \tag{4.1}$$

holds.
Let $r_2, r_3 > 0$ be given. From (4.1) we obtain $\widetilde{\varepsilon_3} = taf(\varepsilon_1) + 2i$ or, equivalently, $ta = \frac{\widetilde{\varepsilon_3} - 2i}{f(\varepsilon_1)}$. Setting $\widetilde{\varepsilon_2} = \varepsilon_2$ we obtain from $2(ta)i + (ta)^2 f(\varepsilon_1) = -2\zeta i$ the equivalent expression

$$f(\varepsilon_1) = -2i \frac{\zeta + ta}{(ta)^2}. \tag{4.2}$$

for $t \neq 0$. Choosing a real number $t \gg 1$, we find an $\varepsilon_1 \in \Delta_r^*$ such that (4.2) is fulfilled. After possibly enlarging t we have $\widetilde{\varepsilon_3} := taf(\varepsilon_1) + 2i = -2i\frac{\zeta}{ta} \in \Delta_{r_3}$. Together with $\varepsilon_2 = \widetilde{\varepsilon_2}$ equation (4.1) is fulfilled and the proof is finished. □

Thus, the Stein surface X_f cannot be obtained as globalization of the local $\mathbb{C}$–action on any $\mathbb{R}$–invariant domain $D \subset X_f$ on which $\mathbb{R}$ acts properly.

4.2. The quotient $X^*/\mathbb{C}$ is Hausdorff. — Suppose that $X^*/\mathbb{C}$ is not Hausdorff and let $x_1, x_2 \in X$ be such that the corresponding $\mathbb{C}$–orbits cannot be separated in $X^*/\mathbb{C}$. Since we already know that $\mathbb{C}$ acts locally proper on X^* we find local holomorphic slices $\varphi_j \colon \Delta \times \mathbb{C} \to U_j \subset X$, $\varphi_j(z,t) = t \cdot s_j(z)$ at each $\mathbb{C} \cdot x_j$ where $s_j \colon \Delta \to X$ is holomorphic with $s_j(0) = x_j$. Consequently, we obtain the transition function $\varphi_{12} \colon (\Delta \setminus A) \times \mathbb{C} \to (\Delta \setminus A) \times \mathbb{C}$ for some closed subset $A \subset \Delta$ which must be of the form $(z,t) \mapsto \big(z, t + f(z)\big)$ for some $f \in \mathcal{O}(\Delta \setminus A)$. The following lemma applies to show that A is discrete and that f is meromorphic on Δ. Hence, we are in one of the model cases discussed in the previous subsection.

Lemma 4.3. — Let Δ_1 and Δ_2 denote two copies of the unit disk $\{z \in \mathbb{C};\ |z| < 1\}$. Let $U \subset \Delta_j$, $j = 1, 2$, be a connected open subset and $f\colon U \subset \Delta_1 \to \mathbb{C}$ a non-constant holomorphic function on U. Define the complex manifold

$$M := (\Delta_1 \times \mathbb{C}) \cup (\Delta_2 \times \mathbb{C})/{\sim},$$

where $\sim$ is the relation $(z_1, t_1) \sim (z_2, t_2) :\Leftrightarrow z_1 = z_2 =: z \in U$ and $t_2 = t_1 + f(z)$.
Suppose that M is Hausdorff. Then the complement A of U is discrete and f extends to a meromorphic function on Δ_1.

Proof. — We first prove that for every sequence (x_n), $x_n \in U$, with $\lim_{n\to\infty} x_n = p \in \partial U$, one has $\lim_{n\to\infty} |f(x_n)| = \infty \in \mathbb{P}_1(\mathbb{C})$. Assume the contrary, i.e. there is a sequence (x_n), $x_n \in U$, with $\lim_{n\to\infty} x_n = p \in \partial U$ such that $\lim_{n\to\infty} f(x_n) = a \in \mathbb{C}$. Choose now $t_1 \in \mathbb{C}$, consider the two points $(p, t_1) \in \Delta_1 \times \mathbb{C}$ and $(p, t_1 + a) \in \Delta_2 \times \mathbb{C}$ and note their corresponding points in M as q_1 and q_2. Then $q_1 \neq q_2$. The sequences $(x_n, t_1) \in \Delta_1 \times \mathbb{C}$ and $(x_n, t_1 + f(x_n)) \in \Delta_2 \times \mathbb{C}$ define the same sequence in M having q_1 and q_2 as accumulation points. So M is not Hausdorff, a contradiction.
In particular we have proved that the zeros of f do not accumulate to ∂U in Δ_1. So there is an open neighborhood V of ∂U in Δ_1 such that the restriction of f to $W := U \cap V$ does not vanish. Let $g := 1/f$ on W. Then g extends to a continuous function on V taking the value zero outside of U. The theorem of Rado implies that this function is holomorphic on V. It follows that the boundary ∂U is discrete in Δ_1 and that f has a pole in each of the points of this set, so f is a meromorphic function on Δ_1. □

Theorem 4.4. — The orbit space $X^*/\mathbb{C}$ is Hausdorff. Consequently, $\mathbb{C}$ acts properly on X^*.

Proof. — By virtue of the above lemma, in a neighborhood of two non-separable $\mathbb{C}$–orbits X is isomorphic to a domain in one of the model Stein surfaces discussed in the previous subsection. Since we have seen there that these surfaces are never globalizations, we arrive at a contradiction. Hence, all $\mathbb{C}$–orbits are separable. □

5. Examples

In this section we discuss several examples which illustrate our results.

5.1. Hyperbolic Stein surfaces with proper $\mathbb{R}$–actions. — Let R be a compact Riemann surface of genus $g \geq 2$. It follows that the universal covering of R is given by the unit disc $\Delta \subset \mathbb{C}$ and hence that R is hyperbolic. The fundamental group $\pi_1(R)$ of R contains a normal subgroup N such that $\pi_1(R)/N \cong \mathbb{Z}$. Let $\widetilde{R} \to R$ denote the corresponding normal covering. Then $\widetilde{R}$ is a hyperbolic Riemann surface with a holomorphic $\mathbb{Z}$–action such that $\widetilde{R}/\mathbb{Z} = R$. Note that $\mathbb{Z}$ is not contained in a one parameter group of automorphisms of $\widetilde{R}$.
We have two mappings

$$\begin{tikzcd} X := \mathbb{H} \times_{\mathbb{Z}} \widetilde{R} \arrow[r, "q"] \arrow[d, "p"'] & \widetilde{R}/\mathbb{Z} = R \\ \mathbb{H}/\mathbb{Z} \cong \Delta \setminus \{0\}. & \end{tikzcd}$$

The map $p\colon X \to \Delta \setminus \{0\}$ is a holomorphic fiber bundle with fiber $\widetilde{R}$. Since the Serre problem has a positive answer if the fiber is a non-compact Riemann surface ([Mok82]), the suspension $X = \mathbb{H} \times_{\mathbb{Z}} \widetilde{R}$ is a hyperbolic Stein surface. The group $\mathbb{R}$ acts on $\mathbb{H} \times \widetilde{R}$ by $t \cdot (z, x) = (z + t, x)$ and this action commutes with the diagonal action of $\mathbb{Z}$. Consequently, we obtain an action of $\mathbb{R}$ on X.

Lemma 5.1. — The universal globalization of the local $\mathbb{C}$–action on X is given by $X^* = \mathbb{C} \times_{\mathbb{Z}} \widetilde{R}$. Moreover, $\mathbb{C}$ acts properly on X^*.

Proof. — One checks directly that $t \cdot [z, x] := [z + t, x]$ defines a holomorphic $\mathbb{C}$–action on $X^* = \mathbb{C} \times_{\mathbb{Z}} \widetilde{R}$ which extends the $\mathbb{R}$–action on X. We will show that X is orbit-connected in X^*: Since $[z + t, x]$ lies in X if and only if there exist elements $(z', x') \in \mathbb{H} \times \widetilde{R}$ and $m \in \mathbb{Z}$ such that $(z+t, x) = \big(z' + m, m \cdot x'\big)$, we conclude $\mathbb{C}[z, x] = \big\{t \in \mathbb{C};\ \mathrm{Im}(t) > -\mathrm{Im}(z)\big\}$ which is connected.
In order to show that $\mathbb{C}$ acts properly on X^* it is sufficient to show that $\mathbb{C} \times \mathbb{Z}$ acts properly on $\mathbb{C} \times \widetilde{R}$. Hence, we choose sequences $\{t_n\}$ in $\mathbb{C}$, $\{m_n\}$ in $\mathbb{Z}$ and $\big\{(z_n, x_n)\big\}$ in $\mathbb{C} \times \widetilde{R}$ such that

$$\begin{aligned} &\big((t_n, m_n) \cdot (z_n, x_n), (z_n, x_n)\big) = \\ &\qquad = \big((z_n + t_n + m_n, m_n \cdot x_n), (z_n, x_n)\big) \to \big((z_1, x_1), (z_0, x_0)\big) \end{aligned}$$

holds. Since $\mathbb{Z}$ acts properly on $\widetilde{R}$, it follows that $\{m_n\}$ has a convergent subsequence, which in turn implies that $\{t_n\}$ has a convergent subsequence. Hence, the lemma is proven. $\square$

Proposition 5.2. — The quotient $X/\mathbb{Z} \cong \Delta^* \times R$ is not holomorphically separable and in particular not Stein. The quotient $X^*/\mathbb{C}$ is biholomorphically equivalent to $\widetilde{R}/\mathbb{Z} = R$.

Proof. — It is sufficient to note that the map $\Phi\colon X = \mathbb{H} \times_{\mathbb{Z}} \widetilde{R} \to \Delta^* \times R$, $\Phi[z, x] := \big(e^{2\pi i z}, [x]\big)$, induces a biholomorphic map $X/\mathbb{Z} \to \Delta^* \times R$. $\square$

Thus we have found an example for a hyperbolic Stein surface X endowed with a proper $\mathbb{R}$–action such that the associated $\mathbb{Z}$–quotient is not holomorphically separable. Moreover, the $\mathbb{R}$–action on X extends to a proper $\mathbb{C}$–action on a Stein manifold X^* containing X as an orbit-connected domain such that $X^*/\mathbb{C}$ is any given compact Riemann surface of genus $g \geq 2$.

5.2. Counterexamples with domains in $\mathbb{C}^n$. — There is a bounded Reinhardt domain D in $\mathbb{C}^2$ endowed with a holomorphic action of $\mathbb{Z}$ such that $D/\mathbb{Z}$ is not Stein. However, this $\mathbb{Z}$–action does not extend to an $\mathbb{R}$–action. We give quickly the construction.
Let $\lambda := \frac{1}{2}(3+\sqrt{5})$ and

$$D := \{(x,y) \in \mathbb{C}^2 \mid |x| > |y|^\lambda, |y| > |x|^\lambda\}.$$

It is obvious that D is a bounded Reinhardt domain in $\mathbb{C}^2$ avoiding the coordinate hyperplanes. The holomorphic automorphism group of D is a semidirect product $\Gamma \ltimes (S^1)^2$, where the group $\Gamma \simeq \mathbb{Z}$ is generated by the automorphism $(x,y) \mapsto (x^3y^{-1}, x)$ and $(S^1)^2$ is the rotation group. Therefore the group Γ is not contained in a one-parameter group. Furthermore the quotient D/Γ is the (non-Stein) complement of the singular point in a 2-dimensional normal complex Stein space, a so-called "cusp singularity". These singularities are intensively studied in connection with Hilbert modular surfaces and Inoue-Hirzebruch surfaces, see e.g. [vdG88] and [Zaf01].
In the rest of this subsection we give an example of a hyperbolic domain of holomorphy in a 3–dimensional Stein solvmanifold endowed with a proper $\mathbb{R}$–action such that the $\mathbb{Z}$–quotient is not Stein. While this domain is not simply-connected, its fundamental group is much simpler than the fundamental groups of our two-dimensional examples.
Let $G := \left\{ \left(\begin{smallmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{smallmatrix}\right);\ a,b,c \in \mathbb{C} \right\}$ be the complex Heisenberg group and let us consider its discrete subgroup

$$\Gamma := \left\{ \begin{pmatrix} 1 & m & \frac{m^2}{2} + 2\pi ik \\ 0 & 1 & m + 2\pi il \\ 0 & 0 & 1 \end{pmatrix};\ m,k,l \in \mathbb{Z} \right\}.$$

Note that Γ is isomorphic to $\mathbb{Z}_m \ltimes \mathbb{Z}^2_{(k,l)}$. We let Γ act on $\mathbb{C}^2$ by

$$(z,w) \mapsto \left(z + mw - \frac{m^2}{2} - 2\pi ik, w - m - 2\pi il \right).$$

Proposition 5.3. — The group Γ acts properly and freely on $\mathbb{C}^2$, and the quotient manifold $\mathbb{C}^2/\Gamma$ is holomorphically separable but not Stein.

Proof. — Since $\Gamma' \cong \mathbb{Z}^2$ is a normal subgroup of Γ, we obtain $\mathbb{C}^2/\Gamma \cong (\mathbb{C}^2/\Gamma')/(\Gamma/\Gamma')$. The map $\mathbb{C}^2 \to \mathbb{C}^* \times \mathbb{C}^*$, $(z,w) \mapsto \big(\exp(z), \exp(w)\big)$, identifies $\mathbb{C}^2/\Gamma'$ with $\mathbb{C}^* \times \mathbb{C}^*$. The induced action of $\Gamma/\Gamma' \cong \mathbb{Z}$ on $\mathbb{C}^* \times \mathbb{C}^*$ is given by

$$(z,w) \mapsto \left(e^{-m^2/2} z w^m, e^{-m} w \right)$$

which shows that Γ acts properly and freely on $\mathbb{C}^2$. Moreover, we obtain the commutative diagram

$$\begin{tikzcd} \mathbb{C}^* \times \mathbb{C}^* \arrow[r] \arrow[d, "(z,w)\mapsto w"'] & Y := (\mathbb{C}^* \times \mathbb{C}^*)/\mathbb{Z} \arrow[d] \\ \mathbb{C}^* \arrow[r] & T := \mathbb{C}^*/\mathbb{Z}. \end{tikzcd}$$

The group $\mathbb{C}^*$ acts by multiplication in the first factor on $\mathbb{C}^* \times \mathbb{C}^*$ and this action commutes with the $\mathbb{Z}$–action. One checks directly that the joint $(\mathbb{C}^* \times \mathbb{Z})$–action on $\mathbb{C}^* \times \mathbb{C}^*$ is proper which implies that the map $Y \to T$ is a $\mathbb{C}^*$–principal bundle. Consequently, Y is not Stein.
In order to show that Y is holomorphically separable, note that by [Oel92] this $\mathbb{C}^*$–principal bundle $Y \to T$ extends to a line bundle $p\colon L \to T$ with first Chern class $c_1(L) = -1$. Therefore the zero section of $p\colon L \to T$ can be blown down and we obtain a singular normal Stein space $\overline{Y} = Y \cup \{y_0\}$ where $y_0 = \mathrm{Sing}(\overline{Y})$ is the blown down zero section. Thus Y is holomorphically separable. $\square$

Let us now choose a neighborhood of the singularity $y_0 \in \overline{Y}$ biholomorphic to the unit ball and let U be its inverse image in $\mathbb{C}^2$. It follows that U is a hyperbolic domain with smooth strictly Levi-convex boundary in $\mathbb{C}^2$ and in particular Stein. In order to obtain a proper action of $\mathbb{R}$ we form the suspension $D = \mathbb{H} \times_\Gamma U$ where Γ acts on $\mathbb{H} \times U$ by $(t, z, w) \mapsto (t + m, z + mw - \frac{m^2}{2} - 2\pi ik, w - m - 2\pi il)$.

Proposition 5.4. — *The suspension $D = \mathbb{H} \times_\Gamma U$ is isomorphic to a Stein domain in the Stein manifold G/Γ.*

Proof. — We identify $\mathbb{H} \times U$ with the $\mathbb{R} \times \Gamma$–invariant domain

$$\Omega := \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} ;\ \mathrm{Im}(a) > 0, (c, b) \in U \right\}$$

in G.
Since $\mathbb{H} \times U$ is Stein, it follows that $\mathbb{H} \times_\Gamma U$ is locally Stein in G/Γ. Hence, by virtue of [DG60] we only have to show that G/Γ is Stein.
For this we note first that G is a closed subgroup of $\mathrm{SL}(2, \mathbb{C}) \ltimes \mathbb{C}^2$ which implies that G/Γ is a closed complex submanifold of $X := \big(\mathrm{SL}(2, \mathbb{C}) \ltimes \mathbb{C}^2\big)/\Gamma$. By [Oel92] the manifold X is holomorphically separable, hence G/Γ is holomorphically separable. Since G is solvable, a result of Huckleberry and Oeljeklaus ([HO86]) yields the Steinness of G/Γ.
One checks directly that the action of $\mathbb{R} \times \Gamma$ on $\mathbb{H} \times U$ is proper which implies that $\mathbb{R}$ acts properly on $\mathbb{H} \times_\Gamma U$. $\square$

Because of $(\mathbb{H} \times_\Gamma U)/\mathbb{Z} \cong \Delta^* \times (U/\Gamma)$ this quotient manifold is not Stein but holomorphically separable.

6. Bounded domains with proper $\mathbb{R}$–actions

In this section we give the proof of our main result.

6.1. Proper $\mathbb{R}$–actions on D. — Let $D \subset \mathbb{C}^n$ be a bounded domain and let $\mathrm{Aut}(D)^0$ be the connected component of the identity in $\mathrm{Aut}(D)$.

Lemma 6.1. — A proper $\mathbb{R}$–action by holomorphic transformations on D exists if and only if the group $\mathrm{Aut}(D)^0$ is non-compact.

Proof. — We note first that effective $\mathbb{R}$–actions by holomorphic transformations on D correspond bijectively to one parameter subgroups $\mathbb{R} \hookrightarrow \mathrm{Aut}(D)^0$, $t \mapsto \varphi_t$, where the correspondence is given by $t \cdot z = \varphi_t(z)$ for $t \in \mathbb{R}$ and $z \in D$. Since the group $\mathrm{Aut}(D)^0$ acts properly on D, proper $\mathbb{R}$–actions correspond to closed embeddings $\mathbb{R} \hookrightarrow \mathrm{Aut}(D)^0$. If $\mathrm{Aut}(D)^0$ admits such an embedding, it cannot be compact.
Conversely, suppose that $\mathrm{Aut}(D)^0$ is not compact. By Theorem 3.1 in [Ho65] there are a maximal compact subgroup K of $\mathrm{Aut}(D)^0$ and a linear subspace V of the Lie algebra of $\mathrm{Aut}(D)^0$ such that the map $K \times V \to \mathrm{Aut}(D)^0$, $(k, \xi) \mapsto k \exp(\xi)$, is a diffeomorphism. Since $\mathrm{Aut}(D)^0$ is not compact, the vector space V has positive dimension and the map $t \mapsto \varphi_t := \exp(t\xi)$, for some $0 \neq \xi \in V$, defines a closed embedding of $\mathbb{R}$ into $\mathrm{Aut}(D)^0$ and hence a proper $\mathbb{R}$–action by holomorphic transformations on D. □

6.2. Steinness of $D/\mathbb{Z}$. — Now we give the proof of our main result.

Theorem 6.2. — Let D be a simply-connected bounded domain of holomorphy in $\mathbb{C}^2$. Suppose that the group $\mathbb{R}$ acts properly by holomorphic transformations on D. Then the complex manifold $D/\mathbb{Z}$ is biholomorphically equivalent to a domain of holomorphy in $\mathbb{C}^2$.

Proof. — Let $D \subset \mathbb{C}^2$ be a simply-connected bounded domain of holomorphy. Since the Serre problem is solvable if the fiber is D, see [Siu76], the universal globalization D^* is a simply-connected Stein surface, [CTIT00]. Moreover, we have shown in Theorem 4.4, that $\mathbb{C}$ acts properly on D^*. Since the Riemann surface $D^*/\mathbb{C}$ is also simply-connected, it must be Δ, $\mathbb{C}$ or $\mathbb{P}_1(\mathbb{C})$. In all three cases the bundle $D^* \to D^*/\mathbb{C}$ is holomorphically trivial. So we can exclude the case that $D^*/\mathbb{C}$ is compact and it follows that $D/\mathbb{Z} \cong \mathbb{C}^* \times (D^*/\mathbb{C})$ is a Stein domain in $\mathbb{C}^2$. □

6.3. A normal form for domains with non-compact $\mathrm{Aut}(D)^0$. — Let $D \subset \mathbb{C}^2$ be a simply-connected bounded domain of holomorphy such that the identity component of its automorphism group is non-compact. As we have seen, this yields a proper $\mathbb{R}$–action on D by holomorphic transformations and the universal globalization of the induced local $\mathbb{C}$–action on D is isomorphic to $\mathbb{C} \times S$ where S is either Δ or $\mathbb{C}$ and where $\mathbb{C}$ acts by translation in the first factor.

Moreover, there are plurisubharmonic functions $u, -o\colon \mathbb{C} \times S \to \mathbb{R} \cup \{-\infty\}$ which fulfill

$$u\big(t \cdot (z_1, z_2)\big) = u(z_1, z_2) - \mathrm{Im}(t) \quad \text{and} \quad o\big(t \cdot (z_1, z_2)\big) = o(z_1, z_2) - \mathrm{Im}(t)$$

such that $D = \big\{(z_1, z_2) \in \mathbb{C} \times S;\ u(z_1, z_2) < 0 < o(z_1, z_2)\big\}$. From this we conclude $u(z_1, z_2) = u(0, z_2) - \mathrm{Im}(z_1)$, $o(z_1, z_2) = o(0, z_2) - \mathrm{Im}(z_1)$ and define $u'(z_2) := u(0, z_2)$, $o'(z_2) := o(0, z_2)$.
We summarize our remarks in the following

Theorem 6.3. — Let D be a simply-connected bounded domain of holomorphy in $\mathbb{C}^2$ admitting a non-compact connected identity component of its automorphism group. Then D is biholomorphic to a domain of the form

$$\widetilde{D} = \big\{(z_1, z_2) \in \mathbb{C} \times S;\ u'(z_2) < \mathrm{Im}(z_1) < o'(z_2)\big\},$$

where the functions $u', -o'$ are subharmonic in S.

Remark. — As a consequence of this normal form we see that the domain D admits a continuous fibration over the contractible domain S such that every fiber is a strip in $\mathbb{C}$. Hence, it follows a posteriori that the simply-connected domain of holomorphy D is contractible.

References

[CTIT00] E. Casadio Tarabusi, A. Iannuzzi & S. Trapani – "Globalizations, fiber bundles, and envelopes of holomorphy", *Math. Z.* 233 (2000), no. 3, p. 535–551.

[DG60] F. Docquier & H. Grauert – "Levisches Problem und Rungescher Satz für Teilgebiete Steinscher Mannigfaltigkeiten", *Math. Ann.* 140 (1960), p. 94–123.

[DK00] J. J. Duistermaat & J. A. C. Kolk – *Lie groups*, Universitext, Springer-Verlag, Berlin, 2000.

[FI01] C. de Fabritiis & A. Iannuzzi – "Quotients of the unit ball of $\mathbb{C}^n$ for a free action of $\mathbb{Z}$", *J. Anal. Math.* 85 (2001), p. 213–224.

[For96] F. Forstneric – "Actions of $(\mathbf{R}, +)$ and $(\mathbf{C}, +)$ on complex manifolds", *Math. Z.* 223 (1996), no. 1, p. 123–153.

[Hei91] P. Heinzner – "Geometric invariant theory on Stein spaces", *Math. Ann.* 289 (1991), no. 4, p. 631–662.

[HI97] P. Heinzner & A. Iannuzzi – "Integration of local actions on holomorphic fiber spaces", *Nagoya Math. J.* 146 (1997), p. 31–53.

[Ho65] G. Hochschild – *The structure of Lie groups*, Holden-Day, Inc., San Francisco-London-Amsterdam, 1965.

[HO86] A. T. Huckleberry & E. Oeljeklaus – "On holomorphically separable complex solv-manifolds ", *Ann. Inst. Fourier (Grenoble)* 36 (1986), no. 3, p. 57–65.

[Ian03] A. Iannuzzi – "Induced local actions on taut and Stein manifolds", *Proc. Amer. Math. Soc.* 131 (2003), no. 12, p. 3839–3843 (electronic).

[IST04] A. Iannuzzi, A. Spiro & S. Trapani – "Complexifications of holomorphic actions and the Bergman metric", *Internat. J. Math.* 15 (2004), no. 8, p. 735–747.

[Kob98] S. Kobayashi – *Hyperbolic complex spaces*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 318, Springer-Verlag, Berlin, 1998.

[Mie08] C. Miebach – "Quotients of bounded homogeneous domains by cyclic groups", 2008, arxiv:math.CV/0803.4476v2.

[Mok82] N. Mok – "The Serre problem on Riemann surfaces", *Math. Ann.* 258 (1981/82), no. 2, p. 145–168.

[Oel92] K. Oeljeklaus – "On the holomorphic separability of discrete quotients of complex Lie groups", *Math. Z.* 211 (1992), no. 4, p. 627–633.

[Pal57] R. S. Palais – "A global formulation of the Lie theory of transformation groups", *Mem. Amer. Math. Soc. No.* 22 (1957), p. iii+123.

[Pal61] ______, "On the existence of slices for actions of non-compact Lie groups", *Ann. of Math. (2)* 73 (1961), p. 295–323.

[Siu76] Y. T. Siu – "Holomorphic fiber bundles whose fibers are bounded Stein domains with zero first Betti number", *Math. Ann.* 219 (1976), no. 2, p. 171–192.

[vdG88] G. van der Geer – *Hilbert modular surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 16, Springer-Verlag, Berlin, 1988.

[Zaf01] D. Zaffran – "Serre problem and Inoue-Hirzebruch surfaces", *Math. Ann.* 319 (2001), no. 2, p. 395–420.

Christian Miebach
LATP-UMR(CNRS) 6632
CMI-Université d'Aix-Marseille I
39 rue Joliot-Curie
F-13453 Marseille Cedex 13
France
miebach@cmi.univ-mrs.fr

Karl Oeljeklaus
LATP-UMR(CNRS) 6632
CMI-Université d'Aix-Marseille I
39 rue Joliot-Curie
F-13453 Marseille Cedex 13
France
karloelj@cmi.univ-mrs.fr

Affine Simplices in Oka Manifolds

Finnur Lárusson

Received: June 8, 2009
Revised: July 27, 2009

Communicated by Thomas Peternell

Abstract. We show that the homotopy type of a complex manifold X satisfying the Oka property is captured by holomorphic maps from the affine spaces $\mathbb{C}^n$, $n \geq 0$, into X. Among such X are all complex Lie groups and their homogeneous spaces. We present generalisations of this result, one of which states that the homotopy type of the space of continuous maps from any smooth manifold to X is given by a simplicial set whose simplices are holomorphic maps into X.

2000 Mathematics Subject Classification: Primary 32Q55. Secondary 18G30, 32C18, 55U10.
Keywords and Phrases: Complex manifold, Stein manifold, Oka manifold, Oka property, simplicial set, singular set, affine simplex, homotopy type, weak equivalence.

1. Introduction

Motivated by Gromov's comments in his seminal paper [12], Sec. 3.5.G and 3.5.G', we prove in Sec. 2 that the homotopy type of an Oka manifold X (as a topological space) is captured by holomorphic maps from the affine spaces $\mathbb{C}^n$, $n \geq 0$, into X. In Sec. 3 we present generalisations of this result. We start with a very brief review of some background material.

The concept of an Oka manifold has evolved from Gromov's paper and subsequent work, mainly due to Forstnerič, see in particular [4] and [5]. By a *Stein inclusion* we mean the inclusion into a reduced Stein space S (or a Stein manifold: the choice is immaterial) of a closed analytic subvariety T. A complex manifold X has the *basic Oka property with interpolation* (BOPI) with respect to $T \hookrightarrow S$ if every continuous map $h : S \to X$ with $h|T$ holomorphic can be deformed to a holomorphic map $S \to X$ with $h|T$ fixed. Also, X has the *interpolation property* with respect to $T \hookrightarrow S$ if every holomorphic map $h : T \to X$ extends to a holomorphic map $S \to X$. The following are equivalent (see [15]) and define what it means for X to be Oka:

(1) X has BOPI with respect to every Stein inclusion.
(2) X has the interpolation property, or equivalently BOPI, with respect to every Stein inclusion $T \hookrightarrow \mathbb{C}^n$, $n \geq 1$, where T is contractible (holomorphically or topologically: the choice is immaterial).

The Oka property has several other equivalent formulations. Each of these has a parametric version, where instead of a single map h as above we have a family of maps depending continuously on a parameter. The parametric Oka properties are all equivalent [4], and are equivalent to the Oka property [7].

A holomorphic map $f : X \to Y$ has the *parametric Oka property with interpolation* (POPI) if for every Stein inclusion $T \hookrightarrow S$, every finite polyhedron P with a subpolyhedron Q, and every continuous map $g : S \times P \to X$ such that the restriction $g|S \times Q$ is holomorphic along S (meaning that $g(\cdot, q) : S \to X$ is holomorphic for each $q \in Q$), the restriction $g|T \times P$ is holomorphic along T, and the composition $f \circ g$ is holomorphic along S, there is a continuous map $G : S \times P \times I \to X$, where $I = [0, 1]$, such that:

(1) $G(\cdot, \cdot, 0) = g$,
(2) $G(\cdot, \cdot, 1) : S \times P \to X$ is holomorphic along S,
(3) $G(\cdot, \cdot, t) = g$ on $S \times Q$ and on $T \times P$ for all $t \in I$,
(4) $f \circ G(\cdot, \cdot, t) = f \circ g$ on $S \times P$ for all $t \in I$.

Equivalently, $Q \hookrightarrow P$ may be taken to be any cofibration between cofibrant topological spaces, such as the inclusion of a subcomplex in a CW-complex, and the existence of G can be replaced by the stronger statement that the inclusion into the space, with the compact-open topology, of continuous maps $h : S \times P \to X$ with $h = g$ on $S \times Q$ and on $T \times P$ and $f \circ h = f \circ g$ on $S \times P$ of the subspace of maps that are holomorphic along S is acyclic, that is, a weak homotopy equivalence (see [14], §16). Taking P to be a point and Q empty defines BOPI for f. A complex manifold X is Oka if and only if the constant map from X to a point satisfies BOPI or, equivalently, POPI. For maps in general, it is not known whether BOPI implies POPI.

The notion of a holomorphic submersion being subelliptic was defined by Forstnerič [2], generalising the concept of ellipticity due to Gromov [12]. Subellipticity is the weakest currently-known sufficient geometric condition for a holomorphic map to satisfy POPI (see Forstnerič's recently-proved parametric Oka principle for liftings [6]) and for a complex manifold to be Oka.

By the influential work of Grauert in [9] and [10], the primary examples of Oka manifolds, to which our results apply, are complex Lie groups and their homogeneous spaces, that is, complex manifolds on which a complex Lie group acts holomorphically and transitively. Among other known examples are $\mathbb{C}^n \setminus A$, where A is an algebraic or a tame analytic subvariety of codimension at least 2, $\mathbb{P}^n \setminus A$, where A is a subvariety of codimension at least 2, Hopf manifolds, Hirzebruch surfaces, and the complement of a finite set in a complex torus of dimension at least 2 (see [3] and [5]).

2. Oka manifolds are homotopically elliptic

Our results are naturally formulated in the language of simplicial sets. Simplicial sets are combinatorial objects that have a homotopy theory equivalent to that of topological spaces, but tend to be more useful or at least more convenient than topological spaces for various homotopy-theoretic purposes. For an introduction to simplicial sets, we refer the reader to [8] or [16].

We denote by $\mathbf{\Delta}$ the category of finite ordinals and order-preserving maps. The objects of $\mathbf{\Delta}$ are the sets $\mathbf{n} = \{0, 1, 2, \dots, n\}$, $n \in \mathbb{N}$, with the usual order, and a morphism $\theta : \mathbf{n} \to \mathbf{m}$ is a map such that $\theta(i) \leq \theta(j)$ whenever $0 \leq i \leq j \leq n$. A cosimplicial object in a category $\mathcal{C}$ is a functor $\mathbf{\Delta} \to \mathcal{C}$. A simplicial object in $\mathcal{C}$ is a functor from the opposite category $\mathbf{\Delta}^{\mathrm{op}}$ to $\mathcal{C}$. In particular, a simplicial set is a functor from $\mathbf{\Delta}^{\mathrm{op}}$ to the category $\mathbf{Set}$ of sets. The category of simplicial objects in $\mathcal{C}$ is denoted $s\mathcal{C}$. A cosimplicial object $A_\bullet$ in $\mathcal{C}$ induces a functor $h_{A_\bullet} : \mathcal{C} \to s\mathbf{Set}$, $X \mapsto \hom_{\mathcal{C}}(A_\bullet, X)$. We call the simplicial set $\hom_{\mathcal{C}}(A_\bullet, X)$ the homotopy type of X with respect to $A_\bullet$.

The standard n-simplex T_n, $n \geq 0$, is the subset

$$T_n = \{(t_0, \dots, t_n) \in \mathbb{R}^{n+1} : t_0 + \cdots + t_n = 1, t_0, \dots, t_n \geq 0\}$$

of $\mathbb{R}^{n+1}$ with the subspace topology. An order-preserving map $\theta : \mathbf{n} \to \mathbf{m}$ induces a continuous map $\theta_* : T_n \to T_m$ defined by the formula $\theta_*(t_0, \dots, t_n) = (s_0, \dots, s_m)$, where

$$s_i = \sum_{j \in \theta^{-1}(i)} t_j$$

(the sum is interpreted as zero if $\theta^{-1}(i)$ is empty). It is easy to check that this defines a cosimplicial object $T_\bullet$ in the category of topological spaces. The homotopy type $sX = \mathscr{C}(T_\bullet, X)$ of a topological space X with respect to $T_\bullet$ is the usual homotopy type of X. Here, for each $n \geq 0$, $\mathscr{C}(T_n, X)$ denotes the set of continuous maps $T_n \to X$. The simplicial set sX is called the singular set of X. It is a fibrant simplicial set, that is, a Kan complex.

The *affine n-simplex* A_n, $n \geq 0$, is the affine subspace

$$A_n = \{(t_0, \dots, t_n) \in \mathbb{C}^{n+1} : t_0 + \cdots + t_n = 1\}$$

of $\mathbb{C}^{n+1}$, viewed as a complex manifold biholomorphic to $\mathbb{C}^n$. An order-preserving map $\theta : \mathbf{n} \to \mathbf{m}$ induces a holomorphic map $\theta_* : A_n \to A_m$ defined by the same formula as above, and we have a cosimplicial object $A_\bullet$ in the category of complex manifolds. We call the homotopy type $eX = \mathscr{O}(A_\bullet, X)$ of a complex manifold X with respect to $A_\bullet$ the *affine homotopy type* of X. Here, for each $n \geq 0$, $\mathscr{O}(A_n, X)$ denotes the set of holomorphic maps $A_n \to X$. We also call the simplicial set eX the *affine singular set* of X.

A holomorphic map $A_n \to X$ is determined by its restriction to $T_n \subset A_n$, so we have a monomorphism, that is, a cofibration $eX \hookrightarrow sX$. The following lemma comes from basic homotopy theory.

Lemma. *For a complex manifold X, the following are equivalent.*

(a) *The affine singular set eX is fibrant and the cofibration $eX \hookrightarrow sX$ is a weak equivalence of simplicial sets.*
(b) *The cofibration $eX \hookrightarrow sX$ is the inclusion of a strong deformation retract.*

Proof. (a) $\Rightarrow$ (b) by [13], Prop. 7.6.11.
(b) $\Rightarrow$ (a) by [13], Prop. 7.8.3, and since a retract of a fibrant object is fibrant. □

We say that X is *homotopically elliptic* if conditions (a) and (b) are satisfied. Then the usual homotopy type of X as a topological space is represented by the affine singular set eX of X.
If X is connected and homotopically elliptic, then X is $\mathbb{C}$-connected, meaning that any two points in X can be joined by an entire curve. In fact, any finite subset of X lies in a holomorphic image of $\mathbb{C}$. On the other hand, if X is Brody hyperbolic, then eX is discrete.

Theorem 1. *An Oka manifold is homotopically elliptic.*

Proof. Let $Z_n = \{(z_1, \dots, z_n) \in \mathbb{C}^n : z_j = 0 \text{ for some } j\}$ be the union of the coordinate hyperplanes in $\mathbb{C}^n$, $n \geq 2$. If X is an Oka manifold, every holomorphic map $Z_n \to X$ extends to a holomorphic map $\mathbb{C}^n \to X$, but this is precisely what it means for eX to be fibrant.
The homotopy groups $\pi_m(K, *)$, $m \geq 1$, of a Kan complex K with respect to a base point $* \in K_0$ may be simply described as follows:

$$\pi_m(K, *) = \{a \in K_m : d_j a = * \text{ for } j = 0, \dots, m\}/\sim,$$

where $d_j : K_m \to K_{m-1}$ is the face map that in the case of sX and eX acts by precomposition by the map

$$\delta_j : (t_0, \dots, t_{m-1}) \mapsto (t_0, \dots, t_{j-1}, 0, t_j, \dots, t_{m-1}),$$

and $\sim$ is the equivalence relation with $a \sim b$ for $a, b \in K_m$ with all faces $*$ if there is $c \in K_{m+1}$ such that $d_j c = a$ for some j, $d_j c = b$ for another j, and $d_j c = *$ for the remaining values of j. Identifying vertices $a, b \in K_0$ if there is $c \in K_1$ with $d_0 c = a$ and $d_1 c = b$ (this is an equivalence relation) gives the set $\pi_0(K)$ of path components of K. (See e.g. [1], Th. 2.4, or [17], Sec. 8.2—homotopy groups of non-fibrant simplicial sets are not so easily dealt with.)
Since X is Oka, two points in the same path component of X can be joined by a holomorphic image of $\mathbb{C}$. Thus the inclusion $eX \hookrightarrow sX$ induces a bijection $\pi_0(eX) \to \pi_0(sX)$.
By induction over m we obtain continuous retractions $\rho_m : A_m \to T_m$, $m \geq 0$, such that $\rho_{m+1} \circ \delta_j = \delta_j \circ \rho_m$ for $j = 0, \dots, m$, so ρ_m retracts each face of A_m onto the corresponding face of T_m. The continuous surjection $\sigma_m : T_m \times I \to T_{m+1}$,

$$(t_0, \dots, t_m, s) \mapsto (t_0(1-s), t_1, \dots, t_m, t_0 s),$$

$m \geq 1$, collapses each segment $\{x\} \times I$, where x belongs to the face of T_m with $t_0 = 0$, and makes no other identifications.

Let $m \geq 1$ and choose a base point $* \in X$. To prove surjectivity of the induced map $\pi_m(eX, *) \to \pi_m(sX, *)$, we need to show that if $a \in s_m X$ has all faces $*$, then there is $b \in e_m X$ with all faces $*$ that is equivalent to a by some $c \in s_{m+1}X$. Now $a_0 = a \circ \rho_m : A_m \to X$ is continuous with all faces $*$, so since X is Oka, there is a continuous deformation a_t, $t \in I$, of a_0, such that a_1 is holomorphic and a_t has all faces $*$ for all $t \in I$. The restriction to $T_m \times I$ of the deformation factors through σ_m by a map $T_{m+1} \to X$, which is continuous since σ_m is a quotient map, and which is the desired c.

To prove injectivity of the induced map $\pi_m(eX, *) \to \pi_m(sX, *)$, we need to show that if $a, b \in e_m X$ with all faces $*$ are equivalent by $c \in s_{m+1}X$, say $dc = (a, b, *, \ldots, *)$, then a and b are also equivalent by some $c' \in e_{m+1}X$. Continuously extend c to $T_{m+1} \cup W_{m+1}$, where $W_{m+1} = \{(t_0, \ldots, t_{m+1}) \in A_{m+1} : t_j = 0 \text{ for some } j\}$, such that dc is still $(a, b, *, \ldots, *)$. Use the acyclic cofibration $T_{m+1} \cup W_{m+1} \hookrightarrow A_{m+1}$ to further extend c to a continuous map $c : A_{m+1} \to X$. Since X is Oka, c may be deformed to $c' \in e_{m+1}X$ with $dc' = dc$. □

The author has tried to directly construct a strong deformation retraction from sX onto eX, but without success.

The proof shows that a complex manifold is homotopically elliptic if and only if it satisfies the interpolation property with respect to the Stein inclusions $Z_n \hookrightarrow \mathbb{C}^n$, $n \geq 2$, and a weak version of BOPI with respect to the Stein inclusions $W_n \hookrightarrow A_n \cong \mathbb{C}^n$, $n \geq 1$.

3. Generalisations

Theorem 1 is a special case of a more general result. Let $f : X \to Y$ be a holomorphic map between complex manifolds and $T \hookrightarrow S$ be a Stein inclusion. Let

$$\begin{array}{ccc} T & \longrightarrow & X \\ \downarrow & & \downarrow \\ S & \longrightarrow & Y \end{array}$$

be a commuting square of holomorphic maps. Let $L_{\mathscr{O}}$ be the space, with the compact-open topology, of holomorphic liftings in the square, and $L_{\mathscr{C}}$ be the space of continuous liftings. Let $eL_{\mathscr{O}}$ be the simplicial set whose n-simplices, $n \geq 0$, are the holomorphic maps $\lambda : S \times A_n \to X$ such that $\lambda(\cdot, t)$ is a lifting in the square for every $t \in A_n$, and whose maps taking m-simplices to n-simplices are given by precomposing in the second variable by the holomorphic maps $\theta_* : A_n \to A_m$ described above. There are inclusions

$$eL_{\mathscr{O}} \overset{i'}{\hookrightarrow} sL_{\mathscr{O}} \overset{i''}{\hookrightarrow} sL_{\mathscr{C}}.$$

If f satisfies POPI, then i'' is a weak equivalence (see [14], §16). Also, the proof of the Theorem is easily generalised to show that if f satisfies BOPI, then $eL_{\mathscr{O}}$ is fibrant and $i'' \circ i'$ is a weak equivalence. Thus, if f satisfies POPI, i' is a weak equivalence of Kan complexes.

Theorem 1 is the case when T is empty and S and Y are points. A less special case is when T is empty and Y is a point. Then liftings in the square are simply maps $S \to X$, so we write $e\mathscr{O}(S, X)$ for $eL_{\mathscr{O}}$ and conclude that if X is Oka, then the inclusions $e\mathscr{O}(S, X) \hookrightarrow s\mathscr{O}(S, X) \hookrightarrow s\mathscr{C}(S, X)$ are weak equivalences of Kan complexes.

Generalising this in a different direction, we can represent the the homotopy type of the space $\mathscr{C}(M, X)$ of continuous maps from any smooth manifold M to an Oka manifold X by a simplicial set whose simplices are holomorphic maps into X. Namely, assuming as we may that M is real-analytic, by a well-known result of Grauert [11], M can be real-analytically embedded into a Stein manifold S such that M is a strong deformation retract of S. Then, if X is Oka, the homotopy type of $\mathscr{C}(M, X)$ is given by the Kan complex $e\mathscr{O}(S, X)$.

For ease of reference, we summarise the above as a theorem.

Theorem 2. *Let X be an Oka manifold.*

(1) *For every Stein manifold S, the inclusions*

$$e\mathscr{O}(S, X) \hookrightarrow s\mathscr{O}(S, X) \hookrightarrow s\mathscr{C}(S, X)$$

are weak equivalences of Kan complexes.

(2) *For every smooth manifold M, there is a Stein manifold S such that the homotopy type of $\mathscr{C}(M, X)$ is given by the Kan complex $e\mathscr{O}(S, X)$.*

References

[1] E. B. Curtis. *Simplicial homotopy theory.* Adv. Math. 6 (1971) 107–209.

[2] F. Forstnerič. *The Oka principle for sections of subelliptic submersions.* Math. Z. 241 (2002) 527–551.

[3] F. Forstnerič. *The homotopy principle in complex analysis: a survey.* Explorations in complex and Riemannian geometry, 73–99, Contemp. Math., 332, Amer. Math. Soc., 2003.

[4] F. Forstnerič. *Extending holomorphic mappings from subvarieties in Stein manifolds.* Ann. Inst. Fourier (Grenoble) 55 (2005) 733–751.

[5] F. Forstnerič. *Runge approximation on convex sets implies the Oka property.* Ann. of Math. (2) 163 (2006) 689–707.

[6] F. Forstnerič. *Invariance of the parametric Oka property.* In Complex analysis: Several complex variables and connections with PDE theory and geometry, volume in honour of Linda P. Rothschild, Trends in Mathematics, Birkhäuser, to appear, `arXiv:0901.4373`.

[7] F. Forstnerič. *Oka manifolds.* C. R. Math. Acad. Sci. Paris 347 (2009) 1017–1020.

[8] P. G. Goerss and J. F. Jardine. *Simplicial homotopy theory.* Progress in Mathematics, 174. Birkhäuser Verlag, 1999.

[9] H. Grauert. *Holomorphe Funktionen mit Werten in komplexen Lieschen Gruppen.* Math. Ann. 133 (1957) 450–472.

[10] H. Grauert. *Analytische Faserungen über holomorph-vollständigen Räumen.* Math. Ann. 135 (1958) 263–273.

[11] H. Grauert. *On Levi's problem and the imbedding of real-analytic manifolds.* Ann. of Math. (2) 68 (1958) 460–472.

[12] M. Gromov. *Oka's principle for holomorphic sections of elliptic bundles.* J. Amer. Math. Soc. 2 (1989) 851–897.

[13] P. S. Hirschhorn. *Model categories and their localizations.* Mathematical Surveys and Monographs, 99. Amer. Math. Soc., 2003.

[14] F. Lárusson. *Model structures and the Oka Principle.* J. Pure Appl. Algebra 192 (2004) 203–223.

[15] F. Lárusson. *Applications of a parametric Oka principle for liftings.* In Complex analysis: Several complex variables and connections with PDE theory and geometry, volume in honour of Linda P. Rothschild, Trends in Mathematics, Birkhäuser, to appear, `arXiv:0901.4388`.

[16] J. P. May. *Simplicial objects in algebraic topology.* Van Nostrand Mathematical Studies, 11. D. Van Nostrand Co., 1967. Chicago Lectures in Mathematics. University of Chicago Press, 1992.

[17] P. Selick. *Introduction to homotopy theory.* Fields Institute Monographs, 9. Amer. Math. Soc., 1997.

Finnur Lárusson
School of Mathematical Sciences
University of Adelaide
Adelaide SA 5005
Australia.
finnur.larusson@adelaide.edu.au

Thom Spectra that Are Symmetric Spectra

Christian Schlichtkrull

Received: November 7, 2007
Revised: December 18, 2009

Communicated by Stefan Schwede

Abstract. We analyze the functorial and multiplicative properties of the Thom spectrum functor in the setting of symmetric spectra and we establish the relevant homotopy invariance.

2000 Mathematics Subject Classification: 55P43
Keywords and Phrases: Thom spectra, symmetric spectra.

1. Introduction

The purpose of this paper is to develop the theory of Thom spectra in the setting of symmetric spectra. In particular, we establish the relevant homotopy invariance and we investigate the multiplicative properties. Classically, given a sequence of spaces $X_0 \to X_1 \to X_2 \to \dots$, equipped with a compatible sequence of maps $f_n\colon X_n \to BO(n)$, the Thom spectrum $T(f)$ is defined by pulling the universal bundles $V(n)$ over $BO(n)$ back via the f_n's and letting

$$T(f)_n = \overline{f^*V(n)}/X_n.$$

Here the bar denotes fibre-wise one-point compactification. More generally, one may consider compatible families of maps $X_n \to BF(n)$, where $F(n)$ is the topological monoid of base point preserving self-homotopy equivalences of S^n, and similarly define a Thom spectrum by pulling back the canonical S^n-(quasi)fibration over $BF(n)$. Composing with the canonical maps $BO(n) \to BF(n)$, one sees that the latter construction generalizes the former. This generalization was suggested by Mahowald [24], [25], and has been investigated in detail by Lewis in [20].

1.1. Symmetric Thom spectra via $\mathcal{I}$-spaces. In order to translate the definition of Thom spectra into the setting of symmetric spectra, we shall modify the construction by considering certain diagrams of spaces. Let $\mathcal{I}$ be the category whose objects are the finite sets $\mathbf{n} = \{1, \ldots, n\}$, together with the empty set $\mathbf{0}$, and whose morphisms are the injective maps. The concatenation $\mathbf{m} \sqcup \mathbf{n}$ in $\mathcal{I}$ is defined by letting $\mathbf{m}$ correspond to the first m and $\mathbf{n}$ to the last n elements of $\{1, \ldots, m+n\}$. This makes $\mathcal{I}$ a symmetric monoidal category with symmetric structure given by the (m,n)-shuffles $\tau_{m,n} \colon \mathbf{m} \sqcup \mathbf{n} \to \mathbf{n} \sqcup \mathbf{m}$. We define an $\mathcal{I}$-space to be a functor from $\mathcal{I}$ to the category $\mathcal{U}$ of spaces and write $\mathcal{IU}$ for the category of such functors. The correspondence $\mathbf{n} \to BF(n)$ defines an $\mathcal{I}$-space and we show that if $X \to BF$ is a map of $\mathcal{I}$-spaces, then the Thom spectrum $T(f)$ defined as above is a symmetric spectrum. The main advantage of the category Sp^{Σ} of symmetric spectra to ordinary spectra is that it has a symmetric monoidal smash product. Similarly, the category $\mathcal{IU}/BF$ of $\mathcal{I}$-spaces over BF inherits a symmetric monoidal structure from $\mathcal{I}$ and the Thom spectrum functor is compatible with these structures in the following sense.

Theorem 1.1. *The symmetric Thom spectrum functor $T \colon \mathcal{IU}/BF \to Sp^{\Sigma}$ is strong symmetric monoidal.*

That T is strong symmetric monoidal means of course that there is a natural isomorphism of symmetric spectra $T(f) \wedge T(g) \cong T(f \boxtimes g)$, where $f \boxtimes g$ denotes the monoidal product in $\mathcal{IU}/BF$. In particular, T takes monoids in $\mathcal{IU}/BF$ to symmetric ring spectra. A similar construction can be carried out in the setting of orthogonal spectra and the idea of realizing Thom spectra as "structured ring spectra" by such a diagrammatic approach goes back to [31].

1.2. Lifting space level data to $\mathcal{I}$-spaces. Let $\mathcal{N}$ be the ordered set of non-negative integers, thought of as a subcategory of $\mathcal{I}$ via the canonical subset inclusions. Another starting point for the construction of Thom spectra is to consider maps $X \to BF_{\mathcal{N}}$, where $BF_{\mathcal{N}}$ denotes the colimit of the $\mathcal{I}$-space BF restricted to $\mathcal{N}$. Given such a map, one may choose a suitable filtration of X so as to get a map of $\mathcal{N}$-spaces $X(n) \to BF(n)$ and the definition of the Thom spectrum $T(f)$ then proceeds as above. This is the point of view taken by Lewis [20]. The space $BF_{\mathcal{N}}$ has an action of the linear isometries operad $\mathcal{L}$, and Lewis proves that if f is a map of $\mathcal{C}$-spaces where $\mathcal{C}$ is an operad that is augmented over $\mathcal{L}$, then the Thom spectrum $T(f)$ inherits an action of $\mathcal{C}$.
In the setting of symmetric spectra the problem is how to lift space level data to objects in $\mathcal{IU}/BF$. We think of $\mathcal{I}$ as some kind of algebraic structure acting on BF, and in order to pull such an action back via a space level map we should ideally map into the quotient space $BF_{\mathcal{I}}$, that is, into the colimit over $\mathcal{I}$. The problem with this approach is that the homotopy type of $BF_{\mathcal{I}}$ differs from that of $BF_{\mathcal{N}}$. For this reason we shall instead work with the homotopy colimit $BF_{h\mathcal{I}}$ which does have the correct homotopy type. We prove in Section 4 that the homotopy colimit functor has a right adjoint $U \colon \mathcal{U}/BF_{h\mathcal{I}} \to \mathcal{IU}/BF$ such

that this pair of adjoint functors defines a Quillen equivalence

$$\operatorname*{hocolim}_{\mathcal{I}}\colon\ \mathcal{IU}/BF \rightleftarrows \mathcal{U}/BF_{h\mathcal{I}}\ :U.$$

Here the model structure on $\mathcal{IU}$ is the one established by Sagave-Schlichtkrull [33]. The weak equivalences in this model structure are called *$\mathcal{I}$-equivalences* and are the maps that induce weak homotopy equivalences on the associated homotopy colimits; see Section 4.1 for details. It follows from the theorem that the homotopy theory associated to $\mathcal{IU}/BF$ is equivalent to that of $\mathcal{U}/BF_{h\mathcal{I}}$. As is often the case for functors that are right adjoints, U is only homotopically well-behaved when applied to fibrant objects. We shall usually remedy this by composing with a suitable fibrant replacement functor on $\mathcal{U}/BF_{h\mathcal{I}}$ and we write U' for the composite functor so defined.

Composing the right adjoint U with the symmetric Thom spectrum functor from Theorem 1.1 we get a Thom spectrum functor on $\mathcal{U}/BF_{h\mathcal{I}}$. However, even when restricted to fibrant objects this functor does not have all the properties one may expect from a Thom spectrum functor. Notably, one of the important properties of the Lewis-May Thom spectrum functor on $\mathcal{U}/BF_{\mathcal{N}}$ is that it preserves colimits whereas the symmetric Thom spectrum obtained by composing with U does not have this property. For this reason we shall introduce another procedure for lifting space level data to $\mathcal{I}$-spaces in the form of a functor

$$R\colon\ \mathcal{U}/BF_{h\mathcal{I}} \to \mathcal{IU}/BF$$

and we shall use this functor to associate Thom spectra to objects in $\mathcal{U}/BF_{h\mathcal{I}}$. The first statement in the following theorem ensures that the functor so defined produces Thom spectra with the correct homotopy type.

Theorem 1.2. *There is a natural level equivalence $R \xrightarrow{\sim} U'$ over BF and the symmetric Thom spectrum functor defined by the composition*

$$T\colon\ \mathcal{U}/BF_{h\mathcal{I}} \xrightarrow{R} \mathcal{IU}/BF \xrightarrow{T} Sp^{\Sigma}$$

preserves colimits.

As indicated in the theorem we shall use the notation T both for the symmetric Thom spectrum on $\mathcal{IU}/BF$ and for its composition with R; the context will always make the meaning clear. In Section 4.4 we show that in a precise sense our Thom spectrum functor becomes equivalent to that of Lewis-May when composing with the forgetful functor from symmetric spectra to spectra. We also have the following analogue of Lewis' result imposing $\mathcal{L}$-actions on Thom spectra. In our setting the relevant operad is the Barrat-Eccles operad $\mathcal{E}$, see [2] and [28], Remarks 6.5. We recall that $\mathcal{E}$ is an E_∞ operad and that a space with an $\mathcal{E}$-action is automatically an associative monoid.

Theorem 1.3. *The operad $\mathcal{E}$ acts on $BF_{h\mathcal{I}}$ and if $f\colon X \to BF_{h\mathcal{I}}$ is a map of $\mathcal{C}$-spaces where $\mathcal{C}$ is an operad that is augmented over $\mathcal{E}$, then $T(f)$ inherits an action of $\mathcal{C}$.*

We often find that the enriched functoriality obtained by working with homotopy colimits over $\mathcal{I}$ instead of colimits over $\mathcal{N}$ is very useful. For example, one may represent complexification followed by realification as maps of $\mathcal{E}$-spaces

$$BO_{h\mathcal{I}} \to BU_{h\mathcal{I}} \to BO_{h\mathcal{I}},$$

such that the composite E_∞ map represents multiplication by 2. The procedure for lifting space level data described above works quite generally for diagram categories. Implemented in the framework of orthogonal spectra, it gives an answer to the problem left open in [32], Chapter 23, on how to construct orthogonal Thom spectra from space level data; we spell out the details of this in Section 8.5. We also remark that one can define an $\mathcal{I}$-space $BGL_1(A)$ for any symmetric ring spectrum A, and that an analogous lifting procedure allows one to associate A-module Thom spectra to maps $X \to BGL_1(A)_{h\mathcal{I}}$. We hope to return to this in a future paper.

1.3. Homotopy invariance. Ideally, one would like the symmetric Thom spectrum functor to take $\mathcal{I}$-equivalences of $\mathcal{I}$-spaces over BF to stable equivalences of symmetric spectra. However, due to the fact that quasifibrations are not in general preserved under pullbacks this is not true without further assumptions on the objects in $\mathcal{IU}/BF$. We say that an object (X, f) (that is, a map $f\colon X \to BF$) is *T-good* if $T(f)$ has the same homotopy type as the Thom spectrum associated to a fibrant replacement of f; see Definition 5.1 for details.

Theorem 1.4. *If $(X, f) \to (Y, g)$ is an $\mathcal{I}$-equivalence of T-good $\mathcal{I}$-spaces over BF, then the induced map $T(f) \to T(g)$ is a stable equivalence of symmetric spectra.*

Here stable equivalence refers to the stable model structure on Sp^Σ defined in [16] and [27]. It is a subtle property of this model structure that a stable equivalence needs not induce an isomorphism of stable homotopy groups. However, if X and Y are convergent (see Section 2), then the associated Thom spectra are also convergent, and in this case a stable equivalence is indeed a π_*-isomorphism in the usual sense. The T-goodness requirement in the theorem is not a real restriction since in general any object in $\mathcal{U}/BF$ can (and should) be replaced by one that is T-good. The functor R takes values in the subcategory of convergent T-good objects and takes weak homotopy equivalences to $\mathcal{I}$-equivalences (in fact to level-wise equivalences). It follows that the Thom spectrum functor in Theorem 1.2 is a homotopy functor; see Corollary 4.13.

Example 1.5. Theorem 1.4 also has interesting consequences for Thom spectra that are not convergent. As an example, consider the Thom spectrum $MO(1)^{\wedge\infty}$ that represents the bordism theory of manifolds whose stable normal bundle splits as a sum of line bundles, see [1], [9]. This is the symmetric Thom spectrum associated to the map of $\mathcal{I}$-spaces $X(n) \to BF(n)$, where $X(n) = BO(1)^n$. It is proved in [34] that $X_{h\mathcal{I}}$ is homotopy equivalent to $Q(\mathbb{R}P^\infty)$, hence it follows that $MO(1)^{\wedge\infty}$ is stably equivalent as a symmetric ring spectrum to the Thom spectrum associated to the map of infinite loop spaces $Q(\mathbb{R}P^\infty) \to BF_{h\mathcal{I}}$.

In general any $\mathcal{I}$-space X is $\mathcal{I}$-equivalent to the constant $\mathcal{I}$-space $X_{h\mathcal{I}}$ and consequently any symmetric Thom spectrum is stably equivalent to one arising from a space-level map. However, the added flexibility obtained by working in $\mathcal{IU}$ is often very convenient. Notably, it is proved in [33] that any E_∞ monoid in $\mathcal{IU}$ is equivalent to a strictly commutative monoid; something which is well-known not to be the case in $\mathcal{U}$.

1.4. Applications to the Thom isomorphism. As an application of the techniques developed in this paper we present a strictly multiplicative version of the Thom isomorphism. A map $f\colon X \to BF_{h\mathcal{I}}$ gives rise to a morphism in $\mathcal{U}/BF_{h\mathcal{I}}$,

$$\Delta\colon (X, f) \to (X \times X, f \circ \pi_2),$$

where Δ is the diagonal inclusion and π_2 denotes the projection onto the second factor of $X \times X$. The Thom spectrum $T(f \circ \pi_2)$ is isomorphic to $X_+ \wedge T(f)$, and the Thom diagonal

$$\Delta\colon T(f) \to X_+ \wedge T(f)$$

is the map of Thom spectra induced by Δ. In Section 7.1 we define a canonical orientation $T(f) \to H$, where H denotes (a convenient model of) the Eilenberg-Mac Lane spectrum $H\mathbb{Z}/2$. Using this we get a map of symmetric spectra

$$T(f) \wedge H \xrightarrow{\Delta \wedge H} X_+ \wedge T(f) \wedge H \to X_+ \wedge H \wedge H \to X_+ \wedge H, \tag{1.6}$$

where the last map is induced by the multiplication in H. The spectrum level version of the $\mathbb{Z}/2$-Thom isomorphism theorem is the statement that this is a stable equivalence, see [26]. If f is oriented in the sense that it lifts to a map $f\colon X \to BSF_{h\mathcal{I}}$, then we define a canonical integral orientation $T(f) \to H\mathbb{Z}$ and the spectrum level version of the integral Thom isomorphism theorem is the statement that the induced map

$$T(f) \wedge H\mathbb{Z} \to X_+ \wedge H\mathbb{Z} \tag{1.7}$$

is a stable equivalence. In our framework these results lift to "structured ring spectra" in the sense of the following theorem. Here $\mathcal{C}$ again denotes an operad that is augmented over $\mathcal{E}$.

Theorem 1.8. *If $f\colon X \to BF_{h\mathcal{I}}$ (respectively $f\colon X \to BSF_{h\mathcal{I}}$) is a map of $\mathcal{C}$-spaces, then the spectrum level Thom equivalence (1.6) (respectively (1.7)) is a $\mathcal{C}$-map.*

For example, one may represent the complex cobordism spectrum MU as the Thom spectrum associated to the $\mathcal{E}$-map $BU_{h\mathcal{I}} \to BSF_{h\mathcal{I}}$ and the Thom equivalence (1.7) is then an equivalence of E_∞ symmetric ring spectra. This should be compared with the H_∞ version in [20].

1.5. Diagram Thom spectra and symmetrization. The definition of the symmetric Thom spectrum functor shows that the category $\mathcal{I}$ is closely related to the category of symmetric spectra. However, many of the Thom spectra that occur in the applications do not naturally arise from a map of $\mathcal{I}$-spaces but rather from a map of $\mathcal{D}$-spaces for some monoidal category $\mathcal{D}$ equipped with a monoidal functor $\mathcal{D} \to \mathcal{I}$. We formalize this in Section 8 where we introduce the notion of a $\mathcal{D}$-spectrum associated to such a monoidal functor. For example, the complex cobordism spectrum MU associated to the unitary groups $U(n)$ and the Thom spectrum $M\mathfrak{B}$ associated to the braid groups $\mathfrak{B}(n)$ can be realized as diagram ring spectra in this way. It is often convenient to replace the $\mathcal{D}$-Thom spectrum associated to a map of $\mathcal{D}$-spaces $f\colon X \to BF$ by a symmetric spectrum, and our preferred way of doing this is to first transform f to a map of $\mathcal{I}$-spaces and then evaluate the symmetric Thom spectrum functor on this transformed map. In this way we end up with a symmetric spectrum to which we can exploit the structural relationship to the category of $\mathcal{I}$-spaces. We shall discuss various ways of carrying out this "symmetrization" process and in particular we shall see how to realize the Thom spectra MU and $M\mathfrak{B}$ as (in the case of MU commutative) symmetric ring spectra.

1.6. Organization of the paper. We begin by recalling the basic facts about Thom spaces and Thom spectra in Section 2, and in Section 3 we introduce the symmetric Thom spectrum functor and show that it is strong symmetric monoidal. The $\mathcal{I}$-space lifting functor R is introduced in Section 4, where we prove Theorem 1.2 in a more precise form; this is the content of Proposition 4.10 and Corollary 4.13. Here we also compare the Lewis-May Thom spectrum functor to our construction. We prove the homotopy invariance result Theorem 1.4 in Section 5, and in Section 6 we analyze to what extent the constructions introduced in the previous sections are preserved under operad actions. In particular, we prove Theorem 1.3 in a more precise form; this is the content of Corollary 6.9. The Thom isomorphism theorem is proved in Section 7 and in Section 8 we discuss how to symmetrize other types of diagram Thom spectra and how the analogue of the lifting functor R works in the context of orthogonal spectra. Finally, we have included some background material on homotopy colimits in Appendix A.

1.7. Notation and conventions. We shall work in the categories $\mathcal{U}$ and $\mathcal{T}$ of unbased and based compactly generated weak Hausdorff spaces. By a cofibration we understand a map having the homotopy extension property, see [39]. A based space is well-based if the inclusion of the base point is a cofibration. In this paper S^n always denotes the one-point compactification of $\mathbb{R}^n$. By a spectrum E we understand a sequence $\{E_n\colon n \geq 0\}$ of based spaces together with a sequence of based structure maps $S^1 \wedge E_n \to E_{n+1}$. A map of spectra $f\colon E \to F$ is a sequence of based maps $f_n\colon E_n \to F_n$ that commute with the structure maps and we write *Sp* for the category of spectra so defined. A spectrum is *connective* if $\pi_n(E) = 0$ for $n < 0$ and *convergent* if there is an unbounded, non-decreasing sequence of integers $\{\lambda_n : n \geq 0\}$ such that the adjoint structure maps $E_n \to \Omega E_{n+1}$ are $(\lambda_n + n)$-connected for all n.

2. Preliminaries on Thom spaces and Thom spectra

In this section we recall the basic facts about Thom spaces and Thom spectra that we shall need. The main reference for this material is Lewis' account in [20], Section IX. Here we emphasize the details relevant for the construction of symmetric Thom spectra in Section 3. We begin by recalling the two-sided simplicial bar construction and some of its properties, referring to [30] for more details. Given a topological monoid G, a right G-space Y, and a left G-space X, this is the simplicial space $B_\bullet(Y, G, X)$ with k-simplices $Y \times G^k \times X$ and simplicial operators

$$d_i(y, g_1, \dots, g_k, x) = \begin{cases} (yg_1, g_2 \dots, g_k, x), & \text{for } i = 0 \\ (y, g_1, \dots, g_i g_{i+1}, \dots, g_k, x), & \text{for } 0 < i < k \\ (y, g_1, \dots, g_k x), & \text{for } i = k, \end{cases}$$

and

$$s_i(y, g_1, \dots, g_k, x) = (y, \dots, g_{i-1}, 1, g_i, \dots, x), \quad \text{for } 0 \leq i \leq k.$$

We write $B(Y, G, X)$ for the topological realization. In the case where X and Y equal the one-point space $*$, this is the usual simplicial construction of the classifying space BG. The projection of X onto $*$ induces a map

$$p\colon\ B(Y, G, X) \to B(Y, G, *)$$

whose fibres are homeomorphic to X. Furthermore, if X has a G-invariant basepoint, then the inclusion of the base point defines a section

$$s : B(Y, G, *) \to B(Y, G, X).$$

Recall that a topological monoid is *grouplike* if the set of components with the induced monoid structure is a group.

Proposition 2.1 ([19],[30]). *If G is a well-based grouplike monoid, then the projection p is a quasifibration, and if X has a G-invariant base point such that X is (non-equivariantly) well-based, then the section s is a cofibration.* $\square$

In general we say that a sectioned quasifibration is *well-based* if the section is a cofibration. Let $F(n)$ be the topological monoid of base point preserving homotopy equivalences of S^n, where we recall that the latter denotes the one-point compactification of $\mathbb{R}^n$. It follows from [19], Theorem 2.1, that this is a well-based monoid and we let $V(n) = B(*, F(n), S^n)$. Then $BF(n)$ is a classifying space for sectioned fibrations with fibre equivalent to S^n and the projection $p_n\colon V(n) \to BF(n)$ is a well-based quasifibration. Given a map $f\colon X \to BF(n)$, let $p_X\colon f^*V(n) \to X$ be the pull-back of $V(n)$ along f, and notice that the section s gives rise to a section $s_X\colon X \to f^*V(n)$. The associated Thom space is the quotient space

$$T(f) = f^*V(n)/s_X(X).$$

This construction is clearly functorial on the category $\mathcal{U}/BF(n)$ of spaces over $BF(n)$. We often use the notation (X,f) for an object $f\colon X \to BF(n)$ in this category. In order for the Thom space functor to be homotopically well-behaved we would like p_X to be a quasifibration and s_X to be a cofibration, but unfortunately this is not true in general. This is the main technical difference compared to working with sectioned fibrations. For our purpose it will not do to replace the quasifibration p_n by an equivalent fibration since we then loose the strict multiplicative properties of the bar construction required for the definition of strict multiplicative structures on Thom spectra. We say that f *classifies a well-based quasifibration* if p_X is a quasi-fibration and s_X is a cofibration. The following well-known results are included here for completeness.

Lemma 2.2 ([20]). *Given a well-based sectioned quasifibration $p\colon V \to B$ and a Hurewicz fibration $f\colon X \to B$, the pullback $p_X\colon f^*V \to X$ is again a well-based quasifibration.*

Proof. Since f is a fibration the pullback diagram defining f^*V is homotopy cartesian, hence $f^*V \to X$ is a quasifibration. In order to see that the section s_X is a cofibration, notice that it is the pullback of the section of p along the Hurewicz fibration $f^*V \to V$. The result then follows from Theorem 12 of [40] which states that the pullback of a cofibration along a Hurewicz fibration is again a cofibration. □

Let $Top(n)$ be the topological group of base point preserving homeomorphisms of S^n. The next result is the main reason why the objects in $\mathcal{U}/BF(n)$ that factor through $B\,Top(n)$ are easier to handle than general objects.

Lemma 2.3. *If $f\colon X \to BF(n)$ factors through $B\,Top(n)$, then f classifies a well-based Hurewicz fibration, hence a well-based quasifibration.* □

Proof. Let $W(n) = B(*, Top(n), S^n)$. The projection $W(n) \to B\,Top(n)$ is a fibre bundle by [30], Corollary 8.4, and in particular a Hurewicz fibration. Suppose that f factors through a map $g\colon X \to B\,Top(n)$. Then $f^*V(n)$ is homeomorphic to $g^*W(n)$ and thus p_X is a Hurewicz fibration. We must prove that the section is a cofibration. Let us use the Strøm model structure [41] on $\mathcal{U}$ to get a factorization $g = g_2 g_1$ where g_1 is a cofibration and g_2 is a Hurewicz fibration. From this we get a factorization of the pullback diagram defining $g^*W(n)$,

$$\begin{tikzcd} g^*W(n) \arrow[r] \arrow[d, "p_X"] & g_2^*W(n) \arrow[r] \arrow[d, "p_Y"] & W(n) \arrow[d] \\ X \arrow[r, "g_1"] & Y \arrow[r, "g_2"] & B\,Top(n), \end{tikzcd}$$

and it follows from Lemma 2.2 that the section s_Y of p_Y is a cofibration. Since p_Y is a Hurewicz fibration it follows by the same argument that the induced map $g^*W(n) \to g_2^*W(n)$ is also a cofibration. It is clear that the composition $X \to g^*W(n) \to g_2^*W(n)$ is a cofibration and the conclusion thus follows from

Lemma 5 of [41], which states that if $h = i \circ j$ is a composition of maps in which h and i are both cofibrations, then j is a cofibration as well. □

This lemma applies in particular if f factors through $BO(n)$. In order to get around the difficulty that the Thom space functor is not a homotopy functor on the whole category $\mathcal{U}/BF(n)$ we follow Lewis [20], Section IX, and define a functor

$$\Gamma\colon \mathcal{U}/BF(n) \to \mathcal{U}/BF(n), \quad (X, f) \mapsto (\Gamma_f(X), \Gamma(f)) \tag{2.4}$$

by replacing a map by a (Hurewicz) fibration in the usual way,

$$\Gamma_f(X) = \{(x, \omega) \in X \times BF(n)^I \colon\ f(x) = \omega(0)\}, \quad \Gamma(f)(x, \omega) = \omega(1).$$

We sometimes write $\Gamma(X)$ instead of $\Gamma_f(X)$ when the map f is clear from the context. The natural inclusion $X \to \Gamma_f(X)$, whose second coordinate is the constant path at $f(x)$, defines a natural equivalence from the identity functor on $\mathcal{U}/BF(n)$ to Γ. It follows from Lemma 2.2 that the composition of the Thom space functor T with Γ is a homotopy functor. We think of $T(\Gamma(f))$ as representing the correct homotopy type of the Thom space and say that f is *T-good* if the natural map $T(f) \to T(\Gamma(f))$ is a weak homotopy equivalence. In particular, f is T-good if it classifies a well-based quasifibration. It follows from the above discussion that the restriction of T to the subcategory of T-good objects is a homotopy functor.

Remark 2.5. The fibrant replacement functors used in [20] and [30] are defined using Moore paths instead of paths defined on the unit interval I. The use of Moore paths is less convenient for our purposes since we shall use Γ in combination with more general homotopy pullback constructions.

The following basic lemma is needed in order to establish the connectivity and convergence properties of the Thom spectrum functor. It may for example be deduced from the dual Blakers-Massey Theorem in [14].

Lemma 2.6. *Let*

$$\begin{tikzcd} V_1 \arrow[r] \arrow[d, "p_1"] & V_2 \arrow[d, "p_2"] \\ B_1 \arrow[r, "\beta"] & B_2 \end{tikzcd}$$

be a pullback diagram of well-based quasifibrations p_1 and p_2. Suppose that p_1 and p_2 are n-connected with $n > 1$ and that β is k-connected. Then the quotient spaces V_1/B_1 and V_2/B_2 are well-based and $(n-1)$-connected, and the induced map $V_1/B_1 \to V_2/B_2$ is $(k+n)$-connected. □

We now turn to Thom spectra. Let $\mathcal{N}$ be as in Section 1.2, and write $\mathcal{N}\mathcal{U}$ for the category of $\mathcal{N}$-spaces, that is, functors $X\colon \mathcal{N} \to \mathcal{U}$. Consider the $\mathcal{N}$-space BF defined by the sequence of cofibrations

$$BF(0) \xrightarrow{i_0} BF(1) \xrightarrow{i_1} BF(2) \xrightarrow{i_2} \dots$$

obtained by applying B to the monoid homomorphisms $F(n) \to F(n+1)$ that take an element u to $1_{S^1} \wedge u$, the smash product with the identity on S^1. Notice, that there are pullback diagrams

$$\begin{array}{ccc} S^1 \bar{\wedge} V(n) & \longrightarrow & V(n+1) \\ \Big\downarrow & & \Big\downarrow \\ BF(n) & \xrightarrow{i_n} & BF(n+1), \end{array} \tag{2.7}$$

where $S^1 \bar{\wedge} -$ denotes fibre-wise smash product with S^1. Indeed, there clearly is such a pullback diagram of the underlying simplicial spaces, and topological realization preserves pullback diagrams. We let $\mathcal{NU}/BF$ be the category of $\mathcal{N}$-spaces over BF. Thus, an object is a sequence of maps

$$f_n \colon X(n) \to BF(n)$$

that are compatible with the structure maps. Again we may specify the domain by writing the objects in the form (X, f).

Definition 2.8. The Thom spectrum functor $T \colon \mathcal{NU}/BF \to Sp$ is defined by applying the Thom space construction level-wise, $T(f)_n = T(f_n)$, with structure maps given by

$$S^1 \wedge T(f_n) \cong T(i_n \circ f_n) \to T(f_{n+1}).$$

A morphism in $\mathcal{NU}/BF$ induces a map of Thom spectra in the obvious way. As for the Thom space functor, the Thom spectrum functor is not homotopically well-behaved on the whole category $\mathcal{NU}/BF$. We define a functor $\Gamma \colon \mathcal{NU}/BF \to \mathcal{NU}/BF$ by applying the functor Γ level-wise, and we say that an object (X, f) is *T-good* if the induced map $T(f) \to T(\Gamma(f))$ is a stable equivalence. We say that f is *level-wise T-good* if the induced map is a level-wise equivalence. The following proposition is an immediate consequence of Lemma 2.2 and Lemma 2.6.

Proposition 2.9. *If $f \colon X \to BF$ is T-good, then $T(f)$ is connective.* □

An $\mathcal{N}$-space X is said to be *convergent* if there exists an unbounded, non-decreasing sequence of integers $\{\lambda_n \colon n \geq 0\}$ such that $X(n) \to X(n+1)$ is λ_n-connected for each n.

Proposition 2.10. *If $f \colon X \to BF$ is level-wise T-good and X is convergent, then $T(f)$ is also convergent.*

Proof. We may assume that f is a level-wise fibration, hence classifies a well-based quasifibration at each level. If $X(n) \to X(n+1)$ is λ_n-connected, it follows from Lemma 2.6 that the structure map $S^1 \wedge T(f_n) \to T(f_{n+1})$ is $(\lambda_n + n)$-connected. The convergence of X thus implies that of $T(f)$. □

Given an $\mathcal{N}$-space X, write $X_{h\mathcal{N}}$ for its homotopy colimit. This is homotopy equivalent to the usual telescope construction on X. We say that a morphism $(X, f) \to (Y, g)$ in $\mathcal{NU}/BF$ is an *$\mathcal{N}$-equivalence* if the induced map of homotopy colimits $X_{h\mathcal{N}} \to Y_{h\mathcal{N}}$ is a weak homotopy equivalence. The following theorem can be deduced from [20], Proposition 4.9. We shall indicate a more direct proof in Section 5.1.

Theorem 2.11. *If $(X,f) \to (Y,g)$ is an $\mathcal{N}$-equivalence of T-good $\mathcal{N}$-spaces over BF, then the induced map $T(f) \to T(g)$ is a stable equivalence.*

In particular, it follows that $T \circ \Gamma$ takes $\mathcal{N}$-equivalences to stable equivalences.

3. Symmetric Thom spectra

We begin by recalling the definition of a (topological) symmetric spectrum. The basic references are the papers [16] and [27] that deal respectively with the simplicial and the topological version of the theory. See also [35].

3.1. Symmetric spectra. By definition a symmetric spectrum X is a spectrum in which each of the spaces $X(n)$ is equipped with a base point preserving left Σ_n-action, such that the iterated structure maps

$$\sigma^n\colon\ S^m \wedge X(n) \to X(m+n)$$

are $\Sigma_m \times \Sigma_n$-equivariant. A map of symmetric spectra $f\colon\ X \to Y$ is a sequence of Σ_n-equivariant based maps $X(n) \to Y(n)$ that strictly commute with the structure maps. We write Sp^{Σ} for the category of symmetric spectra. Following [27] we shall view symmetric spectra as diagram spectra, and for this reason we introduce some notation which will be convenient for our purposes. Let the category $\mathcal{I}$ be as in Section 1.1. Given a morphism $\alpha\colon\ \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$, let $\mathbf{n}-\alpha$ denote the complement of $\alpha(\mathbf{m})$ in $\mathbf{n}$ and let $S^{n-\alpha}$ be the one-point compactification of $\mathbb{R}^{\mathbf{n}-\alpha}$. Consider then the topological category $\mathcal{I}_S$ that has the same objects as $\mathcal{I}$, but whose morphism spaces are defined by

$$\mathcal{I}_S(\mathbf{m},\mathbf{n}) = \bigvee_{\alpha\in\mathcal{I}(\mathbf{m},\mathbf{n})} S^{n-\alpha}.$$

We view $\mathcal{I}_S$ as a category enriched in the category of based spaces $\mathcal{T}$. Writing the morphisms in the form $(\mathbf{x},\alpha)$ for $\mathbf{x} \in S^{n-\alpha}$, the composition is defined by

$$\mathcal{I}_S(\mathbf{m},\mathbf{n}) \wedge \mathcal{I}_S(\mathbf{l},\mathbf{m}) \to \mathcal{I}_S(\mathbf{l},\mathbf{n}), \quad (\mathbf{x},\alpha)\wedge(\mathbf{y},\beta) \mapsto (\mathbf{x}\wedge\alpha_*\mathbf{y},\alpha\beta),$$

where $\mathbf{x}\wedge\alpha_*\mathbf{y}$ is defined by the canonical homeomorphism

$$S^{n-\alpha} \wedge S^{m-\beta} \cong S^{n-\alpha\beta}, \quad \mathbf{x}\wedge\mathbf{y} \mapsto \mathbf{x}\wedge\alpha_*\mathbf{y},$$

obtained by reindexing the coordinates of $S^{m-\beta}$ via α. This choice of notation has the advantage of making some of our constructions self-explanatory. By a functor between categories enriched in $\mathcal{T}$ we understand a functor such that the maps of morphism spaces are based and continuous. Thus, if $X\colon\ \mathcal{I}_S \to \mathcal{T}$ is a functor in this sense, then we have for each morphism $\alpha\colon\ \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$ a based continuous map

$$\alpha_*\colon\ S^{n-\alpha} \wedge X(m) \to X(n).$$

One easily checks that the maps $S^1 \wedge X(n) \to X(n+1)$ induced by the morphisms $\mathbf{n} \to 1 \sqcup \mathbf{n}$ give X the structure of a symmetric spectrum and that the category of (based continuous) functors $\mathcal{I}_S \to \mathcal{T}$ may be identified with Sp^Σ in this way. The symmetric monoidal structure of $\mathcal{I}$ gives rise to a symmetric monoidal structure on $\mathcal{I}_S$. On morphism spaces this is given by the continuous maps

$$\sqcup\colon \mathcal{I}_S(\mathbf{m}_1,\mathbf{n}_1)\times\mathcal{I}_S(\mathbf{m}_2,\mathbf{n}_2)\to\mathcal{I}_S(\mathbf{m}_1\sqcup\mathbf{m}_2,\mathbf{n}_1\sqcup\mathbf{n}_2),$$

that map a pair of morphisms $\big((\mathbf{x},\alpha),(\mathbf{y},\beta)\big)$ to $(\mathbf{x}\wedge\mathbf{y},\alpha\sqcup\beta)$. As noted in [27], this implies that the category of symmetric spectra inherits a symmetric monoidal smash product. Given symmetric spectra X and Y, this is defined by the left Kan extension,

$$X\wedge Y(n) = \operatorname*{colim}_{\mathbf{n}_1\sqcup\mathbf{n}_2\to\mathbf{n}} X(n_1)\wedge Y(n_2), \tag{3.1}$$

where the colimit is over the topological category $(\sqcup\downarrow\mathbf{n})$ of objects and morphisms in $\mathcal{I}_S\times\mathcal{I}_S$ over $\mathbf{n}$. More explicitly, we may rewrite this as

$$X\wedge Y(n) = \operatorname*{colim}_{\alpha\colon\,\mathbf{n}_1\sqcup\mathbf{n}_2\to\mathbf{n}} S^{n-\alpha}\wedge X(n_1)\wedge Y(n_2),$$

where the colimit is now over the discrete category $(\sqcup\downarrow\mathbf{n})$ of objects and morphisms in $\mathcal{I}\times\mathcal{I}$ over $\mathbf{n}$. Given a morphism in this category of the form

$$(\beta_1,\beta_2)\colon\,(\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_1\sqcup\mathbf{n}_2\xrightarrow{\alpha}\mathbf{n})\to(\mathbf{n}_1',\mathbf{n}_2',\mathbf{n}_1'\sqcup\mathbf{n}_2'\xrightarrow{\alpha'}\mathbf{n}),$$

the morphism α' specifies a homeomorphism

$$S^{n-\alpha}\cong S^{n-\alpha'}\wedge S^{n_1'-\beta_1}\wedge S^{n_2'-\beta_2},$$

and the induced map in the diagram is defined by

$$\begin{aligned} S^{n-\alpha}\wedge X(n_1)\wedge Y(n_2) &\to S^{n-\alpha'}\wedge S^{n_1'-\beta_1}\wedge X(n_1)\wedge S^{n_2'-\beta_2}\wedge Y(n_2)\\ &\to S^{n-\alpha'}\wedge X(n_1')\wedge Y(n_2'). \end{aligned}$$

The unit for the smash product is the sphere spectrum S with $S(n)=S^n$. By definition, a *symmetric ring spectrum* is a monoid in this monoidal category. Spelling this out using the above notation, a symmetric ring spectrum is a symmetric spectrum X equipped with a based map $1_X\colon S^0\to X(0)$ and a collection of based maps

$$\mu_{m,n}\colon\, X(m)\wedge X(n)\to X(m+n),$$

such that the usual unitality and associativity conditions hold, and such that the diagrams

$$\begin{array}{ccc} S^{n-\alpha}\wedge X(m)\wedge S^{n'-\alpha'}\wedge X(m') & \xrightarrow{\mu_{m,m'}\circ tw} & S^{n+n'-\alpha\sqcup\alpha'}\wedge X(m+m') \\ \big\downarrow{\scriptstyle\alpha\wedge\alpha'} & & \big\downarrow{\scriptstyle\alpha\sqcup\alpha'} \\ X(n)\wedge X(n') & \xrightarrow{\mu_{n,n'}} & X(n+n') \end{array}$$

commute for each pair of morphisms $\alpha\colon\mathbf{m}\to\mathbf{n}$ and $\alpha'\colon\mathbf{m}'\to\mathbf{n}'$ in $\mathcal{I}$. Here *tw* flips the factors $X(m)$ and $S^{n'-\alpha'}$. A ring spectrum is *commutative* if it defines a commutative monoid in Sp^{Σ}. Explicitly, this means that there are commutative diagrams

$$\begin{array}{ccc} X(m) \wedge X(n) & \xrightarrow{\mu_{m,n}} & X(m+n) \\ \Big\downarrow tw & & \Big\downarrow \tau_{m,n} \\ X(n) \wedge X(m) & \xrightarrow{\mu_{n,m}} & X(n+m), \end{array}$$

where the right hand vertical map is given by the left action of the (m,n)-shuffle $\tau_{m,n}$.

3.2. Symmetric Thom spectra via $\mathcal{I}$-spaces. As in Section 1.1 we write $\mathcal{IU}$ for the category of $\mathcal{I}$-spaces. This category inherits the structure of a symmetric monoidal category from that of $\mathcal{I}$ in the usual way: given $\mathcal{I}$-spaces X and Y, their product $X \boxtimes Y$ is defined by the Kan extension

$$(X \boxtimes Y)(n) = \operatorname*{colim}_{\mathbf{n_1} \sqcup \mathbf{n_2} \to \mathbf{n}} X(n_1) \times Y(n_2),$$

where the colimit is again over the category $(\sqcup \downarrow \mathbf{n})$. The unit for the monoidal structure is the constant $\mathcal{I}$-space $\mathcal{I}(\mathbf{0}, -)$. We use term *$\mathcal{I}$-space monoid* for a monoid in this category. This amounts to an $\mathcal{I}$-space X equipped with a unit $1_X \in X(0)$ and a natural transformation of $\mathcal{I} \times \mathcal{I}$-diagrams

$$\mu_{m,n} : X(m) \times X(n) \to X(m+n)$$

that satisfies the obvious associativity and unitality conditions. An $\mathcal{I}$-space monoid X is commutative if it defines a commutative monoid in $\mathcal{IU}$, that is, the diagrams

$$\begin{array}{ccc} X(m) \times X(n) & \xrightarrow{\mu_{m,n}} & X(m+n) \\ \Big\downarrow tw & & \Big\downarrow \tau_{m,n} \\ X(n) \times X(m) & \xrightarrow{\mu_{n,m}} & X(n+m) \end{array}$$

are commutative.

The family of topological monoids $F(n)$ define a functor from $\mathcal{I}$ to the category of topological monoids: a morphism $\alpha\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$ induces a monoid homomorphism $\alpha\colon F(m) \to F(n)$ by associating to an element f in $F(m)$ the composite map

$$S^n \cong S^{n-\alpha} \wedge S^m \xrightarrow{S^{n-\alpha} \wedge f} S^{n-\alpha} \wedge S^m \cong S^n. \tag{3.2}$$

As usual the homeomorphism $S^{n-\alpha} \wedge S^m \cong S^n$ is induced by the bijection $(\mathbf{n} - \alpha) \sqcup \mathbf{m} \to \mathbf{n}$ specified by α and the inclusion of $\mathbf{n} - \alpha$ in $\mathbf{n}$. We also have the natural monoid homomorphisms

$$F(m) \times F(n) \to F(m+n), \quad (f,g) \mapsto f \wedge g$$

defined by the usual smash product of based spaces. Applying the classifying space functor degree-wise and using that it commutes with products, we get from this the commutative $\mathcal{I}$-space monoid $BF\colon \mathbf{n} \mapsto BF(n)$. We write $\mathcal{IU}/BF$ for the category of $\mathcal{I}$-spaces over BF with objects (X, f) given by a map $f\colon X \to BF$ of $\mathcal{I}$-spaces. This category inherits a symmetric monoidal structure from that of $\mathcal{IU}$: given objects (X,f) and (Y,g), the product is defined by the composition

$$f \boxtimes g\colon X \boxtimes Y \xrightarrow{f\boxtimes g} BF \boxtimes BF \to BF,$$

where the last map is the multiplication in BF. The meaning of the symbol $f \boxtimes g$ will always be clear from the context. By definition, a monoid in this monoidal structure is a pair (X,f) given by an $\mathcal{I}$-space monoid X together with a monoid morphism $f\colon X \to BF$.

Definition 3.3. The symmetric Thom spectrum functor $T\colon \mathcal{IU}/BF \to Sp^{\Sigma}$ is defined by the level-wise Thom space construction $T(f)(n) = T(f_n)$. A morphism $\alpha\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$ gives rise to a pullback diagram

$$\begin{array}{ccc} S^{n-\alpha}\bar{\wedge} V(m) & \longrightarrow & V(n) \\ \downarrow & & \downarrow \\ BF(m) & \xrightarrow{\alpha} & BF(n) \end{array}$$

in which $\bar{\wedge}$ denotes the fibre-wise smash product. On fibres this restricts to the homeomorphism $S^{n-\alpha} \wedge S^m \to S^n$ specified by α. Pulling this diagram back via f and applying the Thom space construction, we get the required structure maps

$$S^{n-\alpha} \wedge T(f_m) \cong T(\alpha \circ f_m) \to T(f_n).$$

Notice, that this Thom spectrum functor is related to that in Section 2 by a commutative diagram of functors

$$\begin{array}{ccc} \mathcal{IU}/BF & \xrightarrow{T} & Sp^{\Sigma} \\ \downarrow & & \downarrow \\ \mathcal{NU}/BF & \xrightarrow{T} & Sp, \end{array} \tag{3.4}$$

where the vertical arrows represent the obvious forgetful functors. Recall the notion of a strong symmetric monoidal functor from [22], Section XI.2. We now prove Theorem 1.1 stating that the symmetric Thom spectrum is strong symmetric monoidal.

Proof of Theorem 1.1. It is clear that we have a canonical isomorphism $S \to T(*)$. We must show that given objects (X,f) and (Y,g) in $\mathcal{IU}/BF$ there is a natural isomorphism

$$T(f) \wedge T(g) \cong T(f \boxtimes g).$$

By definition, $X \boxtimes Y(n)$ is the colimit of the $(\sqcup \downarrow \mathbf{n})$-diagram

$$(\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_1 \sqcup \mathbf{n}_2 \xrightarrow{\alpha} \mathbf{n}) \mapsto X(n_1) \times Y(n_2).$$

Given $\alpha\colon \mathbf{n}_1 \sqcup \mathbf{n}_2 \to \mathbf{n}$, let $\alpha(f,g)$ be the composite map

$$X(n_1) \times Y(n_2) \xrightarrow{f_{n_1} \times g_{n_2}} BF(n_1) \times BF(n_2) \to BF(n_1+n_2) \xrightarrow{\alpha} BF(n).$$

Using these structure maps we view the above $(\sqcup \downarrow \mathbf{n})$-diagram as a diagram over $BF(n)$, and since the Thom space functor preserves colimits by [20], Propositions 1.1 and 1.4, we get the homeomorphism

$$T(f \boxtimes g)(n) \cong \operatorname*{colim}_{(\sqcup\downarrow\mathbf{n})} T(\alpha(f,g)).$$

Furthermore, since pullback commutes with topological realization and fibre-wise smash products, we have an isomorphism

$$\alpha(f,g)^* V(n) \cong S^{n-\alpha} \bar{\wedge} f_{n_1}^* V(n_1) \bar{\wedge} g_{n_2}^* V(n_2)$$

of sectioned spaces over $BF(n)$, hence a homeomorphism of the associated Thom spaces

$$T(\alpha(f,g)) \cong S^{n-\alpha} \wedge T(f_{n_1}) \wedge T(g_{n_2}).$$

Combining the above, we get the homeomorphism

$$T(f) \wedge T(g)(n) \cong \operatorname*{colim}_{\alpha} S^{n-\alpha} \wedge T(f_{n_1}) \wedge T(g_{n_2}) \cong T(f \boxtimes g)(n),$$

specifying the required isomorphism of symmetric spectra. □

Corollary 3.5. *If X is an $\mathcal{I}$-space monoid and $f\colon X \to BF$ a monoid morphism, then $T(f)$ is a symmetric ring spectrum which is commutative if X is.* □

Recall that the *tensor* of an unbased space K with a symmetric spectrum X is defined by the level-wise smash product $X \wedge K_+$. Similarly, the tensor of K with an $\mathcal{I}$-space X is defined by the level-wise product $X \times K$. For an object (X, f) in $\mathcal{IU}/BF$, the tensor is given by $(X \times K, f \circ \pi_X)$, where π_X denotes the projection onto X. We refer to [6], Chapter 6, for a general discussion of tensors in enriched categories.

Proposition 3.6. *The symmetric Thom spectrum functor preserves colimits and tensors with unbased spaces.*

Proof. The first statement follows [20], Proposition 1.1 and Corollary 1.4, which combine to show that the Thom space functor preserves colimits. The second claim is that $T(f \circ \pi_X)$ is isomorphic to $T(f) \wedge K_+$ which follows directly from the definition. □

4. Lifting space level data to $\mathcal{I}$-spaces

The homotopy colimit construction induces a functor

$$\operatorname*{hocolim}_{\mathcal{I}}\colon \mathcal{IU}/BF \to \mathcal{U}/BF_{h\mathcal{I}}, \quad (X \to BF) \mapsto (X_{h\mathcal{I}} \to BF_{h\mathcal{I}}), \tag{4.1}$$

where, given an $\mathcal{I}$-space X, we write $X_{h\mathcal{I}}$ for its homotopy colimit. Our first task in this section is to verify that this is the left adjoint in a Quillen equivalence.

4.1. The right adjoint of $\operatorname{hocolim}_{\mathcal{I}}$. Recall first that the homotopy colimit functor $\mathcal{IU} \to \mathcal{U}$ has a right adjoint that to a space Y associates the $\mathcal{I}$-space $\mathbf{n} \mapsto \operatorname{Map}(B(\mathbf{n} \downarrow \mathcal{I}), Y)$. Here $(\mathbf{n} \downarrow \mathcal{I})$ denotes the category of objects in $\mathcal{I}$ under $\mathbf{n}$. We refer to [7], Section XII.2.2, and [17] for the details of this adjunction. The right adjoint in turn induces a functor

$$U\colon \mathcal{U}/BF_{h\mathcal{I}} \to \mathcal{IU}/BF, \quad (X, f) \mapsto (U_f(X), U(f))$$

by associating to a map $f\colon X \to BF_{h\mathcal{I}}$ the map of $\mathcal{I}$-spaces defined by the upper row in the pullback diagram

$$\begin{tikzcd} U_f(X) \arrow[r, "U(f)"] \arrow[d] & BF \arrow[d] \\ \operatorname{Map}(B(- \downarrow \mathcal{I}), X) \arrow[r] & \operatorname{Map}(B(- \downarrow \mathcal{I}), BF_{h\mathcal{I}}) \end{tikzcd}$$

The map on the right is the unit of the adjunction. It is immediate that U is right adjoint to the homotopy colimit functor in (4.1) and we shall prove in Proposition 4.5 below that the adjunction

$$\operatorname*{hocolim}_{\mathcal{I}}\colon \mathcal{IU}/BF \rightleftarrows \mathcal{U}/BF_{h\mathcal{I}} : U \qquad (4.2)$$

is a Quillen equivalence when we give $\mathcal{U}/BF_{h\mathcal{I}}$ the model structure induced by the Quillen model structure on $\mathcal{U}$ and $\mathcal{IU}/BF$ the model structure induced by the $\mathcal{I}$-model structure on $\mathcal{IU}$ established by Sagave-Schlichtkrull [33]. Before describing the $\mathcal{I}$-model structure we recall that $\mathcal{IU}$ has a level model structure in which the weak equivalences and fibrations are defined level-wise. Given $d \geq 0$, let $F_d\colon \mathcal{U} \to \mathcal{IU}$ be the functor that to a space K associates the $\mathcal{I}$-space $F_d(K) = \mathcal{I}(\mathbf{d}, -) \times K$. Thus, F_d is left adjoint to the evaluation functor that takes an $\mathcal{I}$-space X to $X(d)$. The level structure on $\mathcal{IU}$ is cofibrantly generated with set of generating cofibrations

$$FI = \{F_d(S^{n-1}) \to F_d(D^n) : d \geq 0, n \geq 0\}$$

obtained by applying the functors F_d to the set I of generating cofibrations for the Quillen model structure on $\mathcal{U}$. By a relative cell complex in $\mathcal{IU}$ we understand a map $X \to Y$ that can be written as the transfinite composition of a sequence of maps

$$X = Y_0 \to Y_1 \to Y_2 \to \cdots \to \operatorname*{colim}_{n \geq 0} Y_n = Y$$

where each map $Y_n \to Y_{n+1}$ is the pushout of a coproduct of generating cofibrations. It follows from the general theory for cofibrantly generated model categories that a cofibration in $\mathcal{IU}$ is a retract of a cell complex. We refer the reader to [15], Section 11.6, for a general discussion of level model structures on diagram categories.

As explained in Section 1.2, the weak equivalences in the $\mathcal{I}$-model structure on $\mathcal{IU}$, that is, the $\mathcal{I}$-equivalences, are the maps that induce weak homotopy equivalences of homotopy colimits. The cofibrations in the $\mathcal{I}$-model structure are the same as for the level structure and the fibrations can be characterized as the maps having the right lifting property with respect to acyclic cofibrations. Again, the $\mathcal{I}$-model structure is cofibrantly generated with FI the set of generating cofibrations. There also is an explicit description of the generating acyclic cofibrations and the fibrations but we shall not need this here.

Lemma 4.3. *The adjunction in (4.2) is a Quillen adjunction*

Proof. We claim that $\mathrm{hocolim}_{\mathcal{I}}$ preserves cofibrations and acyclic cofibrations. By definition, $\mathrm{hocolim}_{\mathcal{I}}$ preserves weak equivalences in general and the first claim therefore implies the second. For the first claim it suffices to show that $\mathrm{hocolim}_{\mathcal{I}}$ takes the generating cofibrations for $\mathcal{IU}$ to cofibrations in $\mathcal{U}$. The homotopy colimit of a map $F_d(S^{n-1}) \to F_d(D^n)$ may be identified with the map

$$B(\mathbf{d} \downarrow \mathcal{I}) \times S^{n-1} \to B(\mathbf{d} \downarrow \mathcal{I}) \times D^n$$

and the claim follows since $B(\mathbf{d} \downarrow \mathcal{I})$ is a cell complex. □

In preparation for the proof that the above Quillen adjunction is in fact a Quillen equivalence we make some general comments on homotopy colimits of $\mathcal{I}$-spaces. In general, given an $\mathcal{I}$-space X, the homotopy type of $X_{h\mathcal{I}}$ may be very different from that of $X_{h\mathcal{N}}$. However, if the underlying $\mathcal{N}$-space is convergent, then the natural map $X_{h\mathcal{N}} \to X_{h\mathcal{I}}$ is a weak homotopy equivalence by the following lemma due to Bökstedt; see [23], Lemma 2.3.7, for a published version.

Lemma 4.4 ([8]). *Let X be an $\mathcal{I}$-space and suppose that there exists an unbounded non-decreasing sequence of integers λ_m such that any morphism $\mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$ induces a λ_m-connected map $X(m) \to X(n)$. Then the inclusion $\{\mathbf{n}\} \to \mathcal{I}$ induces a $(\lambda_n - 1)$-connected map $X(n) \to X_{h\mathcal{I}}$ for all n.* □

The structure maps $F(m) \to F(n)$ are $(m-1)$-connected by the Freudentahl suspension theorem and consequently the induced maps $BF(m) \to BF(n)$ are m-connected. Thus, the proposition applies to the $\mathcal{I}$-space BF and we see that the canonical map $BF(n) \to BF_{h\mathcal{I}}$ is $(n-1)$-connected. This map can be written as the composition

$$BF(n) \to \mathrm{Map}(B(\mathbf{n} \downarrow \mathcal{I}), BF_{h\mathcal{I}}) \to BF_{h\mathcal{I}}$$

where the first map is the unit of the adjunction and the second map is defined by evaluating at the vertex represented by the initial object. Since the second map is clearly a homotopy equivalence it follows that the first map is also $(n-1)$-connected.

Proposition 4.5. *The adjunction (4.2) is a Quillen equivalence.*

Proof. Given a cofibrant object $f\colon X \to BF$ in $\mathcal{IU}/BF$ and a fibrant object $g\colon Y \to BF_{h\mathcal{I}}$ in $\mathcal{U}/BF_{h\mathcal{I}}$ we must show that a morphism $\phi\colon X_{h\mathcal{I}} \to Y$ of spaces over $BF_{h\mathcal{I}}$ is a weak homotopy equivalence if and only if the adjoint $\psi\colon X \to U_g(Y)$ is an $\mathcal{I}$-equivalence of $\mathcal{I}$-spaces over BF. The maps ϕ and ψ

are related by the commutative diagram

$$\begin{tikzcd} X_{h\mathcal{I}} \arrow[rr, "\psi_{h\mathcal{I}}"] \arrow[dr, "\phi"'] & & U_g(Y)_{h\mathcal{I}} \arrow[dl, "\varepsilon"] \\ & Y & \end{tikzcd}$$

where ε denotes the counit for the adjunction. It therefore suffices to show that ε is a weak homotopy equivalence. The assumption that (Y,g) be a fibrant object means that g is a fibration and the pullback diagram

$$\begin{array}{ccc} U_g(Y)(n) & \longrightarrow & BF(n) \\ \downarrow & & \downarrow \\ \mathrm{Map}(B(\mathbf{n}\downarrow\mathcal{I}),Y) & \longrightarrow & \mathrm{Map}(B(\mathbf{n}\downarrow\mathcal{I}),BF_{h\mathcal{I}}) \end{array}$$

used to define $U_g(Y)$ is therefore homotopy cartesian. By the remarks following Lemma 4.4 it follows that the vertical maps are $(n-1)$-connected. The counit ε admits a factorization

$$U_g(Y)_{h\mathcal{I}} \to \mathrm{Map}(B(-\downarrow\mathcal{I}),Y)_{h\mathcal{I}} \to Y$$

where the first map is a weak homotopy equivalence by the above discussion and the second map is a weak homotopy equivalence since $B(-\downarrow\mathcal{I})$ is level-wise contractible. This completes the proof. □

The functor U is only homotopically well-behaved when applied to fibrant objects. We define a (Hurewicz) fibrant replacement functor Γ on $\mathcal{U}/BF_{h\mathcal{I}}$ as in (2.4) (replacing $BF(n)$ by $BF_{h\mathcal{I}}$) and we write U' for the composite functor $U\circ\Gamma$. This is up to natural homeomorphism the same as the functor obtained by evaluating the homotopy pullback instead of the pullback in the diagram defining $U_f(X)$.

4.2. The $\mathcal{I}$-space lifting functor R. As discussed in Section 1.2, the functor U does not have all the properties one may wish when constructing Thom spectra from maps to $BF_{h\mathcal{I}}$. In this section we introduce the $\mathcal{I}$-space lifting functor R and we establish some of its properties. Given a space X and a map $f\colon X\to BF_{h\mathcal{I}}$, we shall view this as a map of constant $\mathcal{I}$-spaces. In order to lift it to a map with target BF, consider the $\mathcal{I}$-space $\overline{BF}$ defined by

$$\overline{BF}(n) = \operatorname*{hocolim}_{(\mathcal{I}\downarrow\mathbf{n})} BF\circ\pi_n,$$

where $\pi_n\colon(\mathcal{I}\downarrow\mathbf{n})\to\mathcal{I}$ is the forgetful functor that maps an object $\mathbf{m}\to\mathbf{n}$ to $\mathbf{m}$. By definition, $\overline{BF}$ is the homotopy left Kan extension of BF along the identity functor on $\mathcal{I}$, see Appendix A.1. Since the identity on $\mathbf{n}$ is a terminal object in $(\mathcal{I}\downarrow\mathbf{n})$ there results a canonical homotopy equivalence $t_n\colon\overline{BF}(n)\to BF(n)$ for each n.

Lemma 4.6. *The map $\pi_n\colon\overline{BF}(n)\to BF_{h\mathcal{I}}$ induced by the functor π_n is $(n-1)$-connected.*

Proof. The homotopy equivalence t_n has a section induced by the inclusion of the terminal object in $(\mathcal{I} \downarrow \mathbf{n})$, such that the canonical map $BF(n) \to BF_{h\mathcal{I}}$ factors through $\overline{BF}(n)$. The result therefore follows from Lemma 4.4 and the above discussion. □

Consider now the diagram of $\mathcal{I}$-spaces

$$BF_{h\mathcal{I}} \xleftarrow{\pi} \overline{BF} \xrightarrow{t} BF,$$

where the right hand map is the level-wise equivalence specified above and the left hand map is induced by the functors π_n. Here we again view $BF_{h\mathcal{I}}$ as a constant $\mathcal{I}$-space. We define the $\mathcal{I}$-space $R_f(X)$ to be the level-wise homotopy pullback of the diagram of $\mathcal{I}$-spaces

$$X \xrightarrow{f} BF_{h\mathcal{I}} \xleftarrow{\pi} \overline{BF}, \tag{4.7}$$

that is, $R_f(X)(n)$ is the space

$$\{(x, \omega, b) \in X \times BF^I_{h\mathcal{I}} \times \overline{BF}(n) \colon\ \omega(0) = f(x),\ \omega(1) = \pi(b)\}.$$

Notice, that the two projections $R_f(X) \to X$ and $R_f(X) \to \overline{BF}$ are level-wise Hurewicz fibrations of $\mathcal{I}$-spaces. The functor R is defined by
(4.8)

$$R\colon \mathcal{U}/BF_{h\mathcal{I}} \to \mathcal{I}\mathcal{U}/BF, \quad (f\colon X \to BF_{h\mathcal{I}}) \mapsto (R(f)\colon R_f(X) \to \overline{BF} \xrightarrow{t} BF).$$

When there is no risk of confusion we write $R(X)$ instead of $R_f(X)$.

Proposition 4.9. *The $\mathcal{I}$-space $R_f(X)$ is convergent and $R(f)$ is level-wise T-good.*

Proof. Since $R_f(X)$ is defined as a homotopy pullback, we see from Lemma 4.6 that the map $R_f(X)(n) \to X$ is $(n-1)$-connected for each n, hence $R_f(X)$ is convergent. We claim that $R(f)$ classifies a well-based quasifibration at each level. In order to see this we first observe that $t^*V(n)$ is a well-based quasifibration over $\overline{BF}(n)$ by Lemma A.4. Thus, $R(f)^*V(n)$ is a pullback of a well-based quasifibration along the Hurewicz fibration $R_f(X)(n) \to \overline{BF}(n)$, hence is itself a well-based quasifibration by Lemma 2.2. □

Proposition 4.10. *There is a natural level-wise equivalence $R_f(X) \xrightarrow{\sim} U'_f(X)$ over BF.*

Proof. Given a map $f\colon X \to BF_{h\mathcal{I}}$, consider the diagram of $\mathcal{I}$-spaces

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d] & BF_{h\mathcal{I}} \arrow[d] & \overline{BF} \arrow[l, "\pi"'] \arrow[d, "t"] \\
\mathrm{Map}(B(- \downarrow \mathcal{I}), X) \arrow[r, "f"] & \mathrm{Map}(B(- \downarrow \mathcal{I}), BF_{h\mathcal{I}}) & BF \arrow[l]
\end{tikzcd}$$

where we view X and $BF_{h\mathcal{I}}$ as constant $\mathcal{I}$-spaces and the corresponding vertical maps are induced by the projection $B(- \downarrow \mathcal{I}) \to *$. The left hand square is strictly commutative and we claim that the right hand square is homotopy commutative. Indeed, with notation as in Appendix A.1, $\overline{BF}$ is the homotopy

Kan extension $id^h_* BF$ along the identity functor on $\mathcal{I}$ and the adjoints of the two compositions in the diagram are the two maps $\mathrm{hocolim}_{\mathcal{I}} \overline{BF} \to BF_{h\mathcal{I}}$ shown to be homotopic in Lemma A.3. Using the canonical homotopy from that lemma we therefore get a canonical homotopy relating the two composites in the right hand square. The latter homotopy in turn gives rise to a natural maps of the associated homotopy pullbacks, that is, to a natural map $R_f(X) \to U'_f(X)$. Since the vertical maps in the above diagram are level-wise equivalences the same holds for the map of homotopy pullbacks. □

Corollary 4.11. *The functors R and* $\mathrm{hocolim}_{\mathcal{I}}$ *are homotopy inverses in the sense that there is a chain of natural weak homotopy equivalences $R_f(X)_{h\mathcal{I}} \simeq X$ of spaces over $BF_{h\mathcal{I}}$ and a chain of natural $\mathcal{I}$-equivalences $R_{f_{h\mathcal{I}}}(X_{h\mathcal{I}}) \simeq X$ of $\mathcal{I}$-spaces over BF.*

Proof. It follows easily from Proposition 4.5 and its proof that the functor U' has this property and the same therefore holds for R by Proposition 4.11. □

The functor R has good properties both formally and homotopically.

Proposition 4.12. *The functor R in (4.8) takes weak homotopy equivalences over $BF_{h\mathcal{I}}$ to level-wise equivalences over BF and preserves colimits and tensors with unbased spaces.*

Proof. The first statement follows from the homotopy invariance of homotopy pullbacks. In order to verify that R preserves colimits, we first observe that $BF_{h\mathcal{I}}$ is locally equiconnected (the diagonal $BF_{h\mathcal{I}} \to BF_{h\mathcal{I}} \times BF_{h\mathcal{I}}$ is a cofibration) by [19], Corollary 2.4. We then view $R_f(X)$ as the pullback of X along the level-wise Hurewicz fibrant replacement $\Gamma_\pi(\overline{BF}) \to BF_{h\mathcal{I}}$ and the result follows from [20], Propositions 1.1 and 1.2, which together state that the pullback functor along a Hurewicz fibration preserves colimits provided the base space is locally equiconnected. The last statement about preservation of tensors is the claim that if K is an unbased space and (X, f) an object of $\mathcal{U}/BF_{h\mathcal{I}}$, then R takes $(X \times K, f \circ \pi_X)$ to $R_f(X) \times K$; this follows immediately from the definition. □

Combining this result with Proposition 2.10, Proposition 3.6 and Proposition 4.9, we get the following corollary in which we define the Thom spectrum functor on $\mathcal{U}/BF_{h\mathcal{I}}$ using R.

Corollary 4.13. *The Thom spectrum functor*

$$T\colon \mathcal{U}/BF_{h\mathcal{I}} \xrightarrow{R} \mathcal{IU}/BF \xrightarrow{T} Sp^\Sigma \tag{4.14}$$

takes values in the subcategory of well-based, connective and convergent symmetric spectra. It takes weak homotopy equivalences over $BF_{h\mathcal{I}}$ to level-wise equivalences and preserves colimits and tensors with unbased spaces. □

The functor R also behaves well with respect to cofibrations as we explain next. We follow [27] in using the term *h-cofibration* for a morphism having the homotopy extension property. Thus, a map $i\colon A \to X$ in $\mathcal{U}$ is an h-cofibration if and only if the induced map from the mapping cylinder

$$X \cup_i (A \times I) \to X \times I \tag{4.15}$$

admits a retraction. By our conventions, this is precisely what we mean by a cofibration of spaces in this paper. Given a base space B, a morphism i in $\mathcal{U}/B$ is an h-cofibration if the analogous morphism (4.15) admits a retraction in $\mathcal{U}/B$; we emphasize this by saying that i is a *fibre-wise* h-cofibration. These conventions also apply to define h-cofibrations in $\mathcal{IU}$ and, given an $\mathcal{I}$-space B, fibre-wise h-cofibrations in $\mathcal{IU}/B$ with the corresponding mapping cylinders defined level-wise. A morphism $i\colon A \to X$ in Sp^{Σ} is an h-cofibration if the mapping cylinder $X \cup_i (A \wedge I_+)$ is a retract of $X \wedge I_+$.

Proposition 4.16. *The functor R takes maps over $BF_{h\mathcal{I}}$ that are cofibrations in $\mathcal{U}$ to fibre-wise h-cofibrations in $\mathcal{IU}/BF$ and the Thom spectrum functor (4.14) takes such maps to h-cofibrations of symmetric spectra.*

Proof. Notice first that we may view $R_f(X)$ as the pullback of $\overline{BF}$ along the fibrant replacement $\Gamma_f(X) \to BF_{h\mathcal{I}}$. Given a morphism $(A, f) \to (X, g)$ in $\mathcal{U}/BF_{h\mathcal{I}}$ such that $A \to X$ is a cofibration, the induced map $\Gamma_f(A) \to \Gamma_g(X)$ is a fibre-wise h-cofibration by [20], IX, Proposition 1.11. Since fibre-wise h-cofibrations are preserved under pullback, this in turn implies that $R_f(A) \to R_g(X)$ is a fibre-wise h-cofibration over $\overline{BF}$, hence over BF. It follows from Proposition 3.6 that the Thom spectrum functor on $\mathcal{IU}/BF$ takes fibre-wise h-cofibrations to h-cofibrations. Combining this with the above gives the result. $\square$

4.3. Preservation of monoidal structures. Recall from [22], Section XI.2, that given monoidal categories $(\mathcal{A}, \Box, 1_{\mathcal{A}})$ and $(\mathcal{B}, \triangle, 1_{\mathcal{B}})$, a *monoidal functor* $\Phi\colon \mathcal{A} \to \mathcal{B}$ is a functor Φ together with a morphism $1_{\mathcal{B}} \to \Phi(1_{\mathcal{A}})$ and a natural transformation

$$\Phi(X) \triangle \Phi(Y) \to \Phi(X \Box Y),$$

satisfying the usual associativity and unitality conditions. It follows from the definition that if A is a monoid in $\mathcal{A}$, then $\Phi(A)$ inherits the structure of a monoid in $\mathcal{B}$. Since (unbased) homotopy colimits commute with products, we may view $\operatorname{hocolim}_{\mathcal{I}}$ as a monoidal functor $\mathcal{IU} \to \mathcal{U}$ with structure maps

$$\operatorname*{hocolim}_{\mathcal{I}} X \times \operatorname*{hocolim}_{\mathcal{I}} Y \cong \operatorname*{hocolim}_{\mathcal{I}\times\mathcal{I}} X \times Y \to \operatorname*{hocolim}_{\mathcal{I}} X \boxtimes Y$$

induced by the universal natural transformation $X(\mathbf{m}) \times Y(\mathbf{n}) \to X \boxtimes Y(\mathbf{m} \sqcup \mathbf{n})$ of $\mathcal{I} \times \mathcal{I}$-diagrams. The unit morphism is induced by the inclusion of the initial object $\mathbf{0}$, thought of as a vertex in $B\mathcal{I}$. Since BF is a monoid in $\mathcal{IU}$, $BF_{h\mathcal{I}}$ inherits the structure of a topological monoid. It follows that we may also view $\mathcal{U}/BF_{h\mathcal{I}}$ as a monoidal category and the following result is then clear from the definition.

Proposition 4.17. *The functor* $\operatorname{hocolim}_{\mathcal{I}}$ *in (4.1) is monoidal.* $\square$

However, the functor $\operatorname{hocolim}_{\mathcal{I}}$ is not symmetric monoidal, hence does not take commutative monoids in $\mathcal{IU}$ to commutative topological monoids. In particular, $BF_{h\mathcal{I}}$ is not a commutative monoid which is already clear from the fact that it is not equivalent to a product of Eilenberg-Mac Lane spaces. We prove in Section 6.1 that $BF_{h\mathcal{I}}$ has a canonical E_∞ structure and that more generally $\operatorname{hocolim}_{\mathcal{I}}$ takes E_∞ objects in $\mathcal{IU}$ to E_∞ spaces.

PROPOSITION 4.18. *The functor R in (4.8) is monoidal.*

Proof. By definition, $R(*)(0)$ is the loop space of $BF_{h\mathcal{I}}$ and we let $* \to R(*)$ be the map of $\mathcal{I}$-spaces that is the inclusion of the constant loop in degree 0. We must define an associative and unital natural transformation of $\mathcal{I}$-spaces $R(X) \boxtimes R(Y) \to R(X \times Y)$ over BF. By the universal property of the $\boxtimes$-product, this amounts to an associative and unital natural transformation of $\mathcal{I}^2$-diagrams

$$R(X)(\mathbf{m}) \times R(Y)(\mathbf{n}) \to R(X \times Y)(\mathbf{m} \sqcup \mathbf{n}).$$

The domain is the homotopy pullback of the diagram

$$X \times Y \to BF_{h\mathcal{I}} \times BF_{h\mathcal{I}} \leftarrow \overline{BF}(m) \times \overline{BF}(n),$$

and the target is the homotopy pullback of the diagram

$$X \times Y \to BF_{h\mathcal{I}} \leftarrow \overline{BF}(m+n).$$

The $\mathcal{I}$-space $\overline{BF}$ inherits a monoid structure from that of BF such that $\pi\colon \overline{BF} \to BF_{h\mathcal{I}}$ is a map of monoids. Using these structure maps, we define a map from the first diagram to the second, giving the required multiplication. $\square$

Since the monoids in the monoidal category $\mathcal{U}/BF_{h\mathcal{I}}$ are precisely the topological monoids over $BF_{h\mathcal{I}}$, this has the following corollary.

COROLLARY 4.19. *If X is a topological monoid and $f\colon X \to BF_{h\mathcal{I}}$ a monoid morphism, then $T(f)$ is a symmetric ring spectrum.* $\square$

This may be reformulated as saying that the Thom spectrum functor preserves the action of the associativity operad whose kth space is the symmetric group Σ_k, see [28], Section 3. More generally, we show in Section 6 that T preserves all operad actions of operads that are augmented over the Barratt-Eccles operad.

4.4. COMPARISON WITH THE LEWIS-MAY THOM SPECTRUM FUNCTOR. Let as before $BF_{\mathcal{N}}$ denote the colimit of BF over $\mathcal{N}$. In this section we recall the Thom spectrum functor on $\mathcal{U}/BF_{\mathcal{N}}$ considered in [20], Section IX, and we relate this to our symmetric Thom spectrum functor on $\mathcal{U}/BF_{h\mathcal{I}}$. We shall use the same notation for the $\mathcal{I}$-space BF and its restriction to an $\mathcal{N}$-space. The colimit functor $\mathcal{NU}/BF \to \mathcal{U}/BF_{\mathcal{N}}$ has a right adjoint, again denoted U, that to an object $f\colon X \to BF_{\mathcal{N}}$ associates the map of $\mathcal{N}$-spaces defined by the upper row in the pullback diagram

$$\begin{array}{ccc} U_f(X) & \xrightarrow{U(f)} & BF \\ \Big\downarrow & & \Big\downarrow \\ X & \xrightarrow{f} & BF_{\mathcal{N}} \end{array}$$

where the vertical map on the right is the unit of the adjunction relating the colimit and the constant functors. Here we view X and $BF_{\mathcal{N}}$ as constant $\mathcal{N}$-spaces. We again write U' for the functor obtained by composing with the Hurewicz fibrant replacement functor Γ on $\mathcal{U}/BF_{\mathcal{N}}$. The Thom spectrum functor considered in [20] is the composition

$$\mathcal{U}/BF_{\mathcal{N}} \xrightarrow{U} \mathcal{N}\mathcal{U}/BF \xrightarrow{T} Sp,$$

where T is the functor from Section 2. (In the language of [20] this is the Thom *prespectrum* associated to f. The authors go on to define a spectrum $M(f)$ with the property that the adjoint structure maps are homeomorphisms, but this will not be relevant for the discussion here). The first step in the comparison to our symmetric Thom spectrum functor on $\mathcal{U}/BF_{h\mathcal{I}}$ is to relate the spaces $BF_{\mathcal{N}}$ and $BF_{h\mathcal{I}}$. Consider the diagram of weak homotopy equivalences

$$BF_{h\mathcal{I}} \xleftarrow{i} BF_{h\mathcal{N}} \xrightarrow{t} BF_{\mathcal{N}},$$

where i is induced from the inclusion $i\colon \mathcal{N} \to \mathcal{I}$ and t is the canonical projection from the homotopy colimit to the colimit. The former is a weak homotopy equivalence by Lemma 4.4 and the latter is a weak homotopy equivalence since the structure maps are cofibrations. Let us choose a homotopy inverse $j\colon BF_{h\mathcal{I}} \to BF_{h\mathcal{N}}$ of i and a homotopy relating $i \circ j$ to the identity on $BF_{h\mathcal{I}}$. Here we of course use that these spaces have the homotopy type of a CW-complex. The precise formulation of the comparison will depend on these choices. Let ζ be the composite homotopy equivalence

$$\zeta\colon BF_{h\mathcal{I}} \xrightarrow{j} BF_{h\mathcal{N}} \xrightarrow{t} BF_{\mathcal{N}}.$$

In general, given a map $\phi\colon B_1 \to B_2$ in $\mathcal{U}$, we write $\phi_*\colon \mathcal{U}/B_1 \to \mathcal{U}/B_2$ for the functor defined by post-composing with ϕ.

Lemma 4.20. *Suppose that ϕ and ψ are maps from B_1 to B_2 that are homotopic by a homotopy $h\colon B_1 \times I \to B_2$. Then the functors ϕ_* and ψ_* from $\mathcal{U}/B_1$ to $\mathcal{U}/B_2$ are related by a chain of natural weak homotopy equivalences depending on h.*

Proof. Let $h_*\colon \mathcal{U}/B_1 \to \mathcal{U}/B_2$ be the functor that takes $f\colon X \to B_1$ to

$$X \times I \xrightarrow{f \times I} B_1 \times I \xrightarrow{h} B_2.$$

The two endpoint inclusions of X in $X \times I$ then give rise to the natural weak homotopy equivalences $\phi_* \to h_* \leftarrow \psi_*$. $\square$

Applied to the homotopy relating $i \circ j$ to the identity on $BF_{h\mathcal{I}}$ this result gives a chain of natural weak homotopy equivalences relating the composite functor

$$\mathcal{U}/BF_{h\mathcal{I}} \xrightarrow{j_*} \mathcal{U}/BF_{h\mathcal{N}} \xrightarrow{i_*} \mathcal{U}/BF_{h\mathcal{I}}$$

to the identity on $\mathcal{U}/BF_{h\mathcal{I}}$.

Lemma 4.21. *The two compositions in the diagram*

$$\begin{tikzcd} \mathcal{U}/BF_{h\mathcal{I}} \arrow[r, "U'"] \arrow[d, "\zeta_*"] & \mathcal{I}\mathcal{U}/BF \arrow[d, "i^*"] \\ \mathcal{U}/BF_{\mathcal{N}} \arrow[r, "U'"] & \mathcal{N}\mathcal{U}/BF \end{tikzcd}$$

are related by a chain of natural level-wise equivalences.

Proof. We shall interpolate between these functors by relating both to the $\mathcal{N}$-space analogue of the functor U' on $\mathcal{U}/BF_{h\mathcal{I}}$. Thus, given $f\colon X \to BF_{h\mathcal{N}}$, the diagram of $\mathcal{N}$-spaces

$$\mathrm{Map}(B(-\downarrow\mathcal{N}), X) \xrightarrow{f} \mathrm{Map}(B(-\downarrow\mathcal{N}), BF_{h\mathcal{N}}) \leftarrow BF$$

is related by evident chains of term-wise level equivalences to the diagrams

$$\mathrm{Map}(i^*B(-\downarrow\mathcal{I}), X) \xrightarrow{i\circ f} \mathrm{Map}(i^*B(-\downarrow\mathcal{I}), BF_{h\mathcal{I}}) \leftarrow BF$$

and

$$X \xrightarrow{t\circ f} BF_{\mathcal{N}} \leftarrow BF.$$

Evaluating the homotopy pullbacks of these diagrams we get a chain of natural level-wise equivalences relating the two compositions in the diagram

$$\begin{tikzcd} \mathcal{U}/BF_{\mathcal{N}} \arrow[d, "U'"] & \mathcal{U}/BF_{h\mathcal{N}} \arrow[l, "t_*"'] \arrow[r, "i_*"] & \mathcal{U}/BF_{h\mathcal{I}} \arrow[d, "U'"] \\ \mathcal{N}\mathcal{U}/BF & & \mathcal{I}\mathcal{U}/BF. \arrow[ll, "i^*"'] \end{tikzcd}$$

By the remarks following Lemma 4.20 we therefore get a chain of natural transformations

$$i^* \circ U' \sim i^* \circ U' \circ i_* \circ j_* \sim U' \circ t_* \circ j_* \sim U' \circ \zeta_*, \tag{4.22}$$

each of which is a level-wise weak homotopy equivalence. □

We can now compare our symmetric Thom spectrum functor to the Lewis-May Thom spectrum functor on $\mathcal{U}/BF_{\mathcal{N}}$. Since the functors TR and TU' on $\mathcal{U}/BF_{h\mathcal{I}}$ are level-wise equivalent by Proposition 4.10, it suffices to consider TU'.

Proposition 4.23. *The two compositions in the diagram*

$$\begin{array}{ccc} \mathcal{U}/BF_{h\mathcal{I}} & \xrightarrow{TU'} & Sp^{\Sigma} \\ \downarrow{\scriptstyle \zeta_*} & & \downarrow \\ \mathcal{U}/BF_{\mathcal{N}} & \xrightarrow{TU'} & Sp \end{array}$$

are related by a chain of level-wise equivalences.

Proof. The diagram in question is obtained by composing the diagram in Proposition 4.21 with the commutative diagram (3.4). Since the chain of weak homotopy equivalences in (4.22) is contained in the full subcategory of level-wise T-good objects in $\mathcal{NU}/BF$, applying T gives a chain of level-wise equivalences. □

5. Homotopy invariance of symmetric Thom spectra

In this section we prove the homotopy invariance result stated in Theorem 1.4 and we show how the proof can be modified to give the $\mathcal{N}$-space analogue in Theorem 2.11. As for the Thom space functor, the symmetric Thom spectrum functor is not homotopically well-behaved on the whole domain category $\mathcal{IU}/BF$. We define a level-wise Hurewicz fibrant replacement functor on $\mathcal{IU}/BF$ by applying the functor Γ in (2.4) at each level.

Definition 5.1. An object (X, f) in $\mathcal{IU}/BF$ is *T-good* if the canonical map $T(f) \to T(\Gamma(f))$ is a stable equivalence (a weak equivalence in the stable model structure) of symmetric spectra.

As before we say that (X, f) is level-wise T-good if $T(f) \to T(\Gamma(f))$ is a level-wise equivalence. The first step in the proof of Theorem 1.4 is to generalize the definition of $\overline{BF}$ to any $\mathcal{I}$-space X by associating to X the $\mathcal{I}$-space $\overline{X}$ defined by

$$\overline{X}(n) = \operatorname*{hocolim}_{(\mathcal{I}\downarrow \mathbf{n})} X \circ \pi_n.$$

We then have a diagram of $\mathcal{I}$-spaces

$$X_{h\mathcal{I}} \xleftarrow{\pi} \overline{X} \xrightarrow{t} X,$$

where we view $X_{h\mathcal{I}}$ as a constant $\mathcal{I}$-space. The map t is a level-wise equivalence and π is an $\mathcal{I}$-equivalence by Lemma A.2. If $f\colon X \to BF$ is a map of $\mathcal{I}$-spaces, then we have a commutative diagram

$$\begin{array}{ccc} \overline{X} & \xrightarrow{\bar{f}} & \overline{BF} \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ X_{h\mathcal{I}} & \xrightarrow{f_{h\mathcal{I}}} & BF_{h\mathcal{I}}, \end{array}$$

hence there is an induced morphism

$$(\overline{X}, t \circ \bar{f}) \to (R_{f_{h\mathcal{I}}}(X_{h\mathcal{I}}), R(f_{h\mathcal{I}})) \tag{5.2}$$

of $\mathcal{I}$-spaces over BF.

Proposition 5.3. *Applying $T \circ \Gamma$ to the morphism (5.2) gives a stable equivalence of symmetric spectra.*

In order to prove this proposition we shall make use of the level model structure on $\mathcal{IU}$ recalled in Section 4.1. Let $F_d\colon \mathcal{U} \to \mathcal{IU}$ be the functor defined in that section and let us write $F_d(u)\colon\ F_d(K) \to BF$ for the map of $\mathcal{I}$-spaces associated to a map of spaces $u\colon\ K \to BF(d)$.

Lemma 5.4. *If u is a Hurewicz fibration, then $F_d(u)$ is level-wise T-good.*

Proof. The pullback of $V(n)$ along $F_d(u)$ is isomorphic to the coproduct of the pullbacks along u of the fibre-wise suspensions $S^{n-\alpha}\bar{\wedge}V(d)$ over $BF(d)$, where α runs through the injective maps $\mathbf{d} \to \mathbf{n}$. These are well-based quasifibrations by Proposition 2.1 and since u is a fibration, the same holds for the pullbacks by Lemma 2.2 and the claim follows. □

The idea is to first prove Proposition 5.3 for objects of the form $F_d(u)$.

Lemma 5.5. *Applying $T \circ \Gamma$ to the map of $\mathcal{I}$-spaces $\overline{F_d(K)} \to R(F_d(K)_{h\mathcal{I}})$ over BF gives a stable equivalence of symmetric spectra.*

The proof of this requires some preparation. We view K as a space over $BF_{h\mathcal{I}}$ via the map

$$\tilde{u}\colon\ K \to BF(d) \to BF_{h\mathcal{I}}, \tag{5.6}$$

where the second map is induced by the inclusion of $\{\mathbf{d}\}$ in $\mathcal{I}$.

Lemma 5.7. *There is a weak homotopy equivalence $K \to F_d(K)_{h\mathcal{I}}$ of spaces over $BF_{h\mathcal{I}}$.*

Proof. By definition of the homotopy colimit we may identify $F_d(K)_{h\mathcal{I}}$ with $B(\mathbf{d} \downarrow \mathcal{I}) \times K$, where $(\mathbf{d} \downarrow \mathcal{I})$ is the category of objects in $\mathcal{I}$ under $\mathbf{d}$. Since this category has an initial object its classifying space is contractible and the result follows. □

In the case of the $\mathcal{I}$-space $F_d(K)$, the level-wise equivalence $t\colon\ \overline{F_d(K)} \to F_d(K)$ has a section induced by the canonical map $K \to \overline{F_d(K)}(d)$. Using this, we get a commutative diagram in $\mathcal{IU}/BF$,

$$\begin{array}{ccc} F_d(K) & \xrightarrow{\sim} & \overline{F_d(K)} \\ \downarrow & & \downarrow \\ R(K) & \xrightarrow{\sim} & R(F_d(K)_{h\mathcal{I}}). \end{array} \tag{5.8}$$

The upper horizontal map is a level-wise equivalence since t is and the lower horizontal map is a level-wise equivalence by the above lemma. Thus, in order to prove Lemma 5.5, we may equally well consider the vertical map on the left hand side of the diagram.

Given a based space T, let $F_d^S(T)$ be the symmetric spectrum $\mathcal{I}_S(\mathbf{d},-) \wedge T$. The functor F_d^S so defined is left adjoint to the functor $Sp^\Sigma \to \mathcal{T}$ that takes a symmetric spectrum to its dth space, see [27]. In particular it follows from the definition that $F_0^S(T)$ is the suspension spectrum of T. Notice also that the Thom spectrum $T(F_d(u))$ associated to $F_d(u)$ may be identified with $F_d^S(T(u))$, where as usual $T(u)$ denotes the Thom space of the map u. Let $T(\tilde{u})$ be the symmetric Thom spectrum of the map $\tilde{u}$ in (5.6) and let $\Sigma_L^d T(\tilde{u})$ be the left shift by $\mathbf{d}$, that is, the composition of $T(\tilde{u})$ with the concatenation functor $\mathcal{I}_S \to \mathcal{I}_S$, $\mathbf{n} \mapsto \mathbf{d} \sqcup \mathbf{n}$. Thus, the nth space of $\Sigma_L^d T(\tilde{u})$ is $T(\tilde{u})(\mathbf{d} \sqcup \mathbf{n})$ with Σ_n acting via the inclusion $\Sigma_n \to \Sigma_{d+n}$ induced by $\mathbf{n} \mapsto \mathbf{d} \sqcup \mathbf{n}$. The condition that u be a Hurewicz fibration in the following lemma is unnecessarily restrictive, but the present formulation is sufficient for our purposes.

Lemma 5.9. *If u is a Hurewicz fibration, then the canonical map of spaces $T(u) \to T(\tilde{u})(d)$ induces a π_*-isomorphism $F_0^S(T(u)) \to \Sigma_L^d T(\tilde{u})$.*

Proof. In spectrum degree n this is the map of Thom spaces induced by the map

$$K \to R_{\tilde{u}}(K)(\mathbf{d}) \to R_{\tilde{u}}(K)(\mathbf{d} \sqcup \mathbf{n}),$$

viewed as a map of T-good spaces over $BF(d+n)$. This is also a map of spaces over K via the projection $R_{\tilde{u}}(K) \to K$, and it therefore follows from the proof of Proposition 4.9 that its connectivity tends to infinity with n. The result then follows from Lemma 2.6. □

We shall prove Lemma 5.5 using the detection functor D from [37]. We recall that this functor associates to a symmetric spectrum T the symmetric spectrum DT whose nth space is the based homotopy colimit

$$DT(n) = \operatorname*{hocolim}_{\mathbf{m} \in \mathcal{I}} \Omega^m (T(m) \wedge S^n).$$

By [37], Theorem 3.1.2, a map of (level-wise well-based) symmetric spectra $T \to T'$ is a stable equivalence if and only if the induced map $DT \to DT'$ is a π_*-isomorphism. There is a closely related functor $T \mapsto MT$, where MT is the symmetric spectrum with nth space

$$MT(n) = \operatorname*{hocolim}_{\mathbf{m} \in \mathcal{I}} \Omega^m (T(\mathbf{m} \sqcup \mathbf{n})).$$

Thus, MT is the homotopy colimit of the $\mathcal{I}$-diagram of symmetric spectra $\mathbf{m} \mapsto \Omega^m(\Sigma_L^m T)$. There is a canonical map $DT \to MT$, which is a level-wise equivalence if T is convergent and level-wise well-based.

Proof of Lemma 5.5. We claim that applying $T \circ \Gamma$ to the vertical map on the left hand side of (5.8) gives a stable equivalence, and for this we may assume without loss of generality that u is a Hurewicz fibration. Then $F_d(u)$ is T-good by Lemma 5.4 and since $R(\tilde{u})$ is T-good by Proposition 4.9, it suffices to show that $T(F_d(u)) \to T(\tilde{u})$ is a stable equivalence. Furthermore, by [27], Theorem 8.12, it is enough to show that this map is a stable equivalence after smashing with S^d and by the above remarks this in turn follows if applying D gives a π_*-isomorphism. We identify $T(F_d(u))$ with $F_d^S(T(u))$ and claim that there is a commutative diagram

$$\begin{tikzcd} F_0^S(T(u)) \arrow[r, "\sim"] \arrow[d, "\sim"] & \Sigma_L^d T(\tilde{u}) \arrow[r, "\sim"] & \Omega^d(S^d \wedge \Sigma_L^d T(\tilde{u})) \arrow[d, "\sim"] \\ D(S^d \wedge F_d^S(T(u))) \arrow[r] & D(S^d \wedge T(\tilde{u})) \arrow[r, "\sim"] & M(S^d \wedge T(\tilde{u})), \end{tikzcd}$$

where the maps are π_*-isomorphisms as indicated. The vertical map on the left hand side is induced by the space-level map

$$T(u) \to F_d^S(T(u))(d) \to \Omega^d(S^d \wedge F_d^S(T(u))(d)) \to D(S^d \wedge F_d^S(T(u)))(0).$$

It is a fundamental property of the model structure on Sp^Σ that the induced map of symmetric spectra is a π_*-isomorphism, see the proof of [37], Lemma 3.2.5. The first map in the upper row is the stable equivalence from Lemma 5.9, and the remaining indicated arrows are π_*-isomorphisms since $T(\tilde{u})$ is connective and convergent. This proves the claim. $\square$

We now wish to prove Proposition 5.3 by an inductive argument based on the filtration

$$\emptyset = X_0 \to X_1 \to X_2 \to \cdots \to \operatorname{colim}_n X_n = X \tag{5.10}$$

of a cell complex X in $\mathcal{IU}$, cf. the discussion of the level model structure in Section 4.1. In order to carry out the induction step, we need to ensure that the induced maps of Thom spectra are h-cofibrations in the sense of Section 4.2. The following is the $\mathcal{I}$-space analogue of [20], IX, Lemma 1.9 and Proposition 1.11. The proof is essentially the same as in the space-level case.

Proposition 5.11. *The functor Γ on $\mathcal{IU}/BF$ preserves colimits and takes morphisms in $\mathcal{IU}/BF$ that are h-cofibrations in $\mathcal{IU}$ to fibre-wise h-cofibrations.* $\square$

Since the symmetric Thom spectrum functor on $\mathcal{IU}/BF$ preserves colimits and takes fibre-wise h-cofibrations to h-cofibrations by Proposition 3.6, this has the following consequence.

Proposition 5.12. *The composite functor $T \circ \Gamma$ preserves colimits and takes morphisms in $\mathcal{IU}/BF$ that are h-cofibrations in $\mathcal{IU}$ to h-cofibrations of symmetric spectra.* $\square$

Proof of Proposition 5.3. Using the level model structure we may choose a cofibrant $\mathcal{I}$-space X' and a level-wise equivalence $X' \to X$, hence it suffices to consider the case where X is a cofibrant $\mathcal{I}$-space. Then X is a retract of a cell complex which we may view as a cell complex over BF via the retraction. By functoriality we are thus reduced to the case where X is a cell complex with a filtration by h-cofibrations as in (5.10). In order to handle this case we use that both functors in (5.2) preserve colimits and tensors with unbased spaces, hence they also preserve (not necessarily fibre-wise) h-cofibrations. Applying the functor $T \circ \Gamma$, we see that both functors in the proposition preserve colimits and take h-cofibrations of $\mathcal{I}$-spaces over BF to h-cofibrations of symmetric spectra. We prove by induction that the result holds for each of the $\mathcal{I}$-spaces X_n in the filtration. By definition, X_{n+1} is a pushout of a diagram of the form $B \leftarrow A \to X_n$, where $A \to B$ is a coproduct of generating cofibrations, hence in particular an h-cofibration. We view this as a diagram of $\mathcal{I}$-spaces over BF via the inclusion of X_{n+1} in X and get a diagram of Thom spectra

$$\begin{array}{ccccc} T\Gamma(\overline{X}_n) & \longleftarrow & T\Gamma(\overline{A}) & \longrightarrow & T\Gamma(\overline{B}) \\ \downarrow & & \downarrow & & \downarrow \\ T\Gamma(R((X_n)_{h\mathcal{I}})) & \longleftarrow & T\Gamma(R(A_{h\mathcal{I}})) & \longrightarrow & T\Gamma(R(B_{h\mathcal{I}})), \end{array}$$

such that the map for X_{n+1} is the induced map of pushouts. By the above discussion it follows that the horizontal maps on the right hand side of the diagram are h-cofibrations and the vertical maps are stable equivalences by Lemma 5.5 and the induction hypothesis. Consequently the map of pushouts is also a stable equivalence, see [27], Theorem 8.12. □

Proof of Theorem 1.4. We prove that applying the functor $T \circ \Gamma$ to an $\mathcal{I}$-equivalence $X \to Y$ over BF gives a stable equivalence of symmetric spectra. Consider the commutative diagram

$$\begin{array}{ccccc} X & \longleftarrow & \overline{X} & \longrightarrow & R(X_{h\mathcal{I}}) \\ \downarrow & & \downarrow & & \downarrow \\ Y & \longleftarrow & \overline{Y} & \longrightarrow & R(Y_{h\mathcal{I}}) \end{array}$$

of $\mathcal{I}$-spaces over BF. Applying $T \circ \Gamma$ to this diagram we get a diagram of symmetric spectra where the horizontal maps are stable equivalence by Proposition 5.3 and the fact that $T \circ \Gamma$ preserves level-wise equivalences. The result now follows from Corollary 4.13 which ensures that the map $R(X_{h\mathcal{I}}) \to R(Y_{h\mathcal{I}})$ induces a stable equivalence. □

Notice, that as a consequence of the theorem, the composite functor $T \circ \Gamma$ is a homotopy functor on $\mathcal{IU}/BF$ in the sense that it takes $\mathcal{I}$-equivalences to stable equivalences.

5.1. The proof of Theorem of 2.11. The proof of Theorem 2.11 is similar to but simpler than the proof of Theorem 1.4. We first introduce a functor

$$R^{\mathcal{N}} \colon \mathcal{U}/BF_{h\mathcal{N}} \to \mathcal{NU}/BF,$$

which is the $\mathcal{N}$-space analogue of the functor R. Let us temporarily write $\overline{BF}$ for the homotopy Kan extension of the $\mathcal{N}$-space BF along the identity functor of $\mathcal{N}$, that is,

$$\overline{BF}(n) = \operatorname*{hocolim}_{(\mathcal{N}\downarrow\mathbf{n})} BF \circ \pi_n,$$

where π_n is the forgetful functor $(\mathcal{N} \downarrow \mathbf{n}) \to \mathcal{N}$. Given a map $f \colon X \to BF_{h\mathcal{N}}$, we define $R^{\mathcal{N}}_f(X)$ to be the level-wise homotopy pullback of the diagram of $\mathcal{N}$-spaces

$$X \xrightarrow{f} BF_{h\mathcal{N}} \xleftarrow{\pi} \overline{BF},$$

and we define $R^{\mathcal{N}}(f)$ to be the composite map of $\mathcal{N}$-spaces

$$R^{\mathcal{N}}(f)\colon\ R_f^{\mathcal{N}}(X) \to \overline{BF} \xrightarrow{t} BF.$$

Exactly as in the $\mathcal{I}$-space case there is a map of $\mathcal{N}$-spaces

$$(\overline{X}, t\circ \overline{f}) \to (R^{\mathcal{N}}_{f_{h\mathcal{N}}}(X_{h\mathcal{N}}), R^{\mathcal{N}}(f_{h\mathcal{N}})) \tag{5.13}$$

over BF, where we again use the (temporary) notation $\overline{X}$ for the homotopy Kan extension along the identity on $\mathcal{N}$. Theorem 2.11 then follows from the following proposition in the same way that Theorem 1.4 follows from Proposition 5.3.

Proposition 5.14. *Applying $T\circ\Gamma$ to (5.13) gives a stable equivalence of spectra.*

In order to prove this we first consider the $\mathcal{N}$-spaces $F_d(K)$ defined by $\mathcal{N}(\mathbf{d},-)\times K$, where $\mathbf{d}$ is an object in $\mathcal{N}$ and K is a space. Given a map $u\colon K \to BF(d)$, we have the following $\mathcal{N}$-space analogue of (5.8),

$$\begin{tikzcd} F_d(K) \arrow[r] \arrow[d] & \overline{F_d(K)} \arrow[d] \\ R(K) \arrow[r] & R(F_d(K)_{h\mathcal{N}}). \end{tikzcd}$$

However, in contrast to the $\mathcal{I}$-space setting, this is a diagram of convergent $\mathcal{N}$-spaces and the connectivity of the maps in degree n tends to infinity with n. Thus, the $\mathcal{N}$-space analogue of Lemma 5.5 holds with a simpler proof.

Proof of Proposition 5.14. We use that $\mathcal{N}\mathcal{U}$ has a cofibrantly generated level model structure and as in the $\mathcal{I}$-space case we reduce to the case of a cell complex. Using that the functors in (5.13) preserve colimits and h-cofibrations, the inductive argument used in the proof of Proposition 5.3 then also applies in the $\mathcal{N}$-space setting. $\square$

6. Preservation of operad actions

Let $\mathcal{C}$ be an operad as defined in [28] and notice that $\mathcal{C}$ defines a monad C on the symmetric monoidal category $\mathcal{I}\mathcal{U}$ in the usual way by letting

$$C(X) = \coprod_{k=0}^{\infty} \mathcal{C}(k)\times_{\Sigma_k} X^{\boxtimes k}.$$

We define a *$\mathcal{C}$-$\mathcal{I}$-space* to be an algebra for this monad and write $\mathcal{I}\mathcal{U}[\mathcal{C}]$ for the category of such algebras. More explicitly, a $\mathcal{C}$-$\mathcal{I}$-space is an $\mathcal{I}$-space X together with a sequence of maps of $\mathcal{I}$-spaces

$$\theta_k\colon\ \mathcal{C}(k)\times X^{\boxtimes k}\to X,$$

satisfying the associativity, unitality and equivariance relations listed in [28], Lemma 1.4. By the universal property of the $\boxtimes$-product, θ_k is determined by a natural transformation of $\mathcal{I}^k$-diagrams

$$\theta_k\colon\ \mathcal{C}(k)\times X(n_1)\times\cdots\times X(n_k)\to X(n_1+\cdots+n_k) \tag{6.1}$$

and the equivariance condition amounts to the commutativity of the diagram

$$\begin{tikzcd}
\mathcal{C}(k)\times X(n_1)\times\cdots\times X(n_k) \arrow[r, "\theta_k\circ(\sigma\times\mathrm{id})"] \arrow[d, "\mathrm{id}\times\sigma"] & X(n_1+\cdots+n_k) \arrow[d, "\sigma(n_1,\dots,n_k)_*"] \\
\mathcal{C}(k)\times X(n_{\sigma^{-1}(1)})\times\cdots\times X(n_{\sigma^{-1}(k)}) \arrow[r, "\theta_k"] & X(n_{\sigma^{-1}(1)}+\cdots+n_{\sigma^{-1}(k)})
\end{tikzcd}$$

for all elements σ in Σ_k. Here σ permutes the factors on the left hand side of the diagram and $\sigma(n_1,\dots,n_k)$ denotes the permutation of $\mathbf{n}_1\sqcup\cdots\sqcup\mathbf{n}_k$ that permutes the k summands as σ permutes the elements of $\mathbf{k}$. As defined in [28], the 0th space of $\mathcal{C}$ is a one-point space, so that θ_0 specifies a base point of X. Notice, that an action of the one-point operad $*$ on an $\mathcal{I}$-space X is the same thing as a commutative monoid structure on X. In this case the projection $\mathcal{C}\to *$ induces a $\mathcal{C}$-action on X for any operad $\mathcal{C}$. This applies in particular to the commutative $\mathcal{I}$-space monoid BF.

In similar fashion an operad $\mathcal{C}$ defines a monad C on the category Sp^{Σ} by letting

$$C(X)=\bigvee_{k=0}^{\infty}\mathcal{C}(k)_+\wedge_{\Sigma_k}X^{\wedge k}$$

and we write $Sp^{\Sigma}[\mathcal{C}]$ for the category of algebras for this monad. Thus, an object of $Sp^{\Sigma}[\mathcal{C}]$ is a symmetric spectrum X together with a sequence of maps of symmetric spectra

$$\theta_k\colon\ \mathcal{C}(k)_+\wedge X^{\wedge k}\to X,$$

satisfying the analogous associativity, unitality and equivariance relations. By the universal property of the smash product, θ_k is determined by a natural transformation of $\mathcal{I}_S^k$-diagrams,

$$\theta_k\colon\ \mathcal{C}(k)_+\wedge X(n_1)\wedge\cdots\wedge X(n_k)\to X(n_1+\cdots+n_k). \tag{6.2}$$

The naturality condition can be formulated explicitly as follows. Given a family of morphisms $\alpha_i\colon\ \mathbf{m}_i\to\mathbf{n}_i$ in $\mathcal{I}$ for $i=1,\dots,k$, let $\alpha=\alpha_1\sqcup\cdots\sqcup\alpha_k$. Writing $n=n_1+\cdots+n_k$ and making the identification

$$S^{n_1-\alpha_1}\wedge\cdots\wedge S^{n_k-\alpha_k}=S^{n-\alpha},$$

we require that the diagram

$$\begin{tikzcd}
S^{n-\alpha}\wedge\mathcal{C}(k)_+\wedge X(m_1)\wedge\cdots\wedge X(m_k) \arrow[r, "S^{n-\alpha}\wedge\theta_k"] \arrow[d] & S^{n-\alpha}\wedge X(m_1+\cdots+m_k) \arrow[d] \\
\mathcal{C}(k)_+\wedge X(n_1)\wedge\cdots\wedge X(n_k) \arrow[r, "\theta_k"] & X(n_1+\cdots+n_k)
\end{tikzcd}$$

be commutative. We now show that the symmetric Thom spectrum functor behaves well with respect to operad actions. Given an operad $\mathcal{C}$ and a map of $\mathcal{I}$-spaces $f\colon\ X\to BF$, let $C(f)$ be the composite map

$$C(f)\colon\ C(X)\to C(BF)\to BF.$$

The following is the analogue in our setting of [20], Theorem IX 7.1. It is a formal consequence of the fact that T is a strong symmetric monoidal functor that preserves colimits and tensors with unbased spaces.

PROPOSITION 6.3. *There is a canonical isomorphism of symmetric spectra*

$$T(C(f)) = C(T(f)). \qquad \square$$

COROLLARY 6.4. *The Thom spectrum functor on $\mathcal{IU}/BF$ preserves operad actions in the sense that there is an induced functor*

$$T\colon \mathcal{IU}[\mathcal{C}]/BF \to Sp^{\Sigma}[\mathcal{C}]. \qquad \square$$

6.1. OPERAD ACTIONS PRESERVED BY $\mathrm{hocolim}_{\mathcal{I}}$. As in Section 1.2 we use the notation $\mathcal{E}$ for the Barratt-Eccles operad. We recall that the kth space $\mathcal{E}(k)$ is the classifying space of the translation category $\tilde{\Sigma}_k$ that has the elements of Σ_k as its objects. A morphism $\rho\colon \sigma \to \tau$ in $\tilde{\Sigma}_k$ is an element $\rho \in \Sigma_k$ such that $\rho\sigma = \tau$; see [29], Section 4 (but notice that the order of the composition in $\tilde{\Sigma}_k$ is defined differently here). In the following proposition, $\mathcal{C}$ denotes an arbitrary operad and $\mathcal{E} \times \mathcal{C}$ denotes the product operad whose kth space is the product $\mathcal{E}(k) \times \mathcal{C}(k)$.

PROPOSITION 6.5. *The functor* $\mathrm{hocolim}_{\mathcal{I}}$ *induces a functor*

$$\operatorname*{hocolim}_{\mathcal{I}}\colon \mathcal{IU}[\mathcal{C}] \to \mathcal{U}[\mathcal{E} \times \mathcal{C}].$$

Proof. Let $\mathcal{I}(X)$ be the topological category whose space of objects is the disjoint union of the spaces $X(n)$ indexed by the objects $\mathbf{n}$ in $\mathcal{I}$, and in which a morphism $(\mathbf{m}, x) \to (\mathbf{n}, y)$ is specified by a morphism $\alpha\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$ such that $\alpha_*(x) = y$. Then it follows from the definition of the homotopy colimit that $X_{h\mathcal{I}}$ may be identified with the classifying space $B\mathcal{I}(X)$; see Appendix A for details. In the following we shall view the spaces $\mathcal{C}(k)$ as topological categories with only identity morphisms. For each k, consider the functor of topological categories

$$\psi_k\colon \tilde{\Sigma}_k \times \mathcal{C}(k) \times \mathcal{I}(X)^k \to \mathcal{I}(X),$$

that maps a tuple of objects $\sigma \in \Sigma_k$, $c \in \mathcal{C}(k)$, and $(\mathbf{n}_1, x_1), \dots, (\mathbf{n}_k, x_k)$, to

$$(\mathbf{n}_{\sigma^{-1}(1)} \sqcup \dots \sqcup \mathbf{n}_{\sigma^{-1}(k)}, \sigma(n_1, \dots, n_k)_* \theta_k(c, x_1, \dots, x_k)).$$

Here θ_k denotes the $\mathcal{C}(k)$-action on X and $\sigma(n_1, \dots, n_k)$ is defined as at the beginning of this section. If $\rho\colon \sigma \to \tau$ is a morphism in $\tilde{\Sigma}_k$, then the induced morphism in $\mathcal{I}(X)$ is specified by

$$\psi_k(\rho) = \rho(n_{\sigma^{-1}(1)}, \dots, n_{\sigma^{-1}(k)}),$$

and if $\vec{\alpha}$ denotes a k-tuple of morphisms in $\mathcal{I}(X)$ whose ith component is specified by $\alpha_i\colon \mathbf{n}_i \to \mathbf{m}_i$, then the induced morphism in $\mathcal{I}(X)$ is specified by

$$\psi_k(\vec{\alpha}) = \alpha_{\sigma^{-1}(1)} \sqcup \dots \sqcup \alpha_{\sigma^{-1}(k)}.$$

Since the classifying space functor preserves products, these functors give rise to maps

$$\psi_k\colon \mathcal{E}(k) \times \mathcal{C}(k) \times B\mathcal{I}(X)^k \to B\mathcal{I}(X),$$

and it is straightforward to check that this defines an $\mathcal{E}\times\mathcal{C}$-action on $B\mathcal{I}(X)$. The associativity, unitality, and equivariance conditions may all be checked on the categorical level. $\square$

Letting $\mathcal{C}$ be the commutativity operad $*$, it follows in particular that if X is a commutative $\mathcal{I}$-space monoid, then $X_{h\mathcal{I}}$ inherits an $\mathcal{E}$-action. In this case the action is induced by a permutative structure on the category $\mathcal{I}(X)$ introduced in the above proof, cf. [29], Section 4. This applies in particular to the $\mathcal{I}$-space BF giving an $\mathcal{E}$-action on $BF_{h\mathcal{I}}$. We say that an operad $\mathcal{C}$ is augmented over $\mathcal{E}$ if there is a specified morphism of operads $\mathcal{C}\to\mathcal{E}$. In this case we may restrict an $(\mathcal{E}\times\mathcal{C})$-action to the diagonal $\mathcal{C}$-action via the morphism $\mathcal{C}\to\mathcal{E}\times\mathcal{C}$.

Corollary 6.6. *If $\mathcal{C}$ is augmented over $\mathcal{E}$, then* $\operatorname{hocolim}_{\mathcal{I}}$ *induces a functor*

$$\operatorname*{hocolim}_{\mathcal{I}}\colon \mathcal{I}\mathcal{U}[\mathcal{C}]/BF\to\mathcal{U}[\mathcal{C}]/BF_{h\mathcal{I}}. \qquad \square$$

6.2. Operad actions preserved by R. In order to prove that the $\mathcal{I}$-space lifting functor R preserves operad actions, we need the following lemma in which we view $BF_{h\mathcal{I}}$ as a constant $\mathcal{E}$-$\mathcal{I}$-space.

Lemma 6.7. *The $\mathcal{I}$-space $\overline{BF}$ has an $\mathcal{E}$-action such that $\pi\colon\overline{BF}\to BF_{h\mathcal{I}}$ is a morphism of $\mathcal{E}$-$\mathcal{I}$-spaces.*

Proof. Consider more generally a commutative $\mathcal{I}$-space monoid X, and let $\overline{X}$ be the $\mathcal{I}$-space defined in Section 5. For each object $\mathbf{n}$ in $\mathcal{I}$, let $\mathcal{I}/\mathbf{n}(X)$ be the topological category whose classifying space is $\overline{X}(n)$. Thus, the object space is given by

$$\coprod_{\alpha\colon\ \mathbf{m}\to\mathbf{n}} X(m),$$

where the coproduct is over the objects in $(\mathcal{I}\downarrow\mathbf{n})$; see Appendix A for details. Consider for each k the functor

$$\psi_k\colon\ \tilde{\Sigma}_k\times\mathcal{I}/\mathbf{n}_1(X)\times\cdots\times\mathcal{I}/\mathbf{n}_k(X)\to\mathcal{I}/(\mathbf{n}_1\sqcup\cdots\sqcup\mathbf{n}_k)(X)$$

that maps a tuple of objects σ in $\tilde{\Sigma}_k$ and (α_i,x_i) in $\mathcal{I}/\mathbf{n}_i(X)$ for $i=1,\dots,k$, to the object

$$(\alpha,\mathbf{x}_{\sigma^{-1}(1)}\dots\mathbf{x}_{\sigma^{-1}(k)}),$$

where α is the morphism

$$\begin{aligned}\alpha\colon\ \mathbf{m}_{\sigma^{-1}(1)}\sqcup\cdots\sqcup\mathbf{m}_{\sigma^{-1}(k)} &\xrightarrow{\alpha_{\sigma^{-1}(1)}\sqcup\cdots\sqcup\alpha_{\sigma^{-1}(k)}}\mathbf{n}_{\sigma^{-1}(1)}\sqcup\cdots\sqcup\mathbf{n}_{\sigma^{-1}(k)}\\ &\xrightarrow{\sigma^{-1}(n_{\sigma^{-1}(1)},\dots,n_{\sigma^{-1}(k)})}\mathbf{n}_1\sqcup\cdots\sqcup\mathbf{n}_k\end{aligned}$$

and the second factor is the product of the elements $\mathbf{x}_{\sigma^{-1}(1)},\dots,\mathbf{x}_{\sigma^{-1}(k)}$ using the monoid structure. The induced maps of classifying spaces

$$\psi_k\colon\ \mathcal{E}(k)\times\overline{X}(n_1)\times\cdots\times\overline{X}(n_k)\to\overline{X}(n_1+\cdots+n_k)$$

then specify the required $\mathcal{E}$-action on $\overline{X}$. With this definition it is clear that the canonical morphism $\overline{X}\to X_{h\mathcal{I}}$ is a morphism of $\mathcal{E}$-$\mathcal{I}$-spaces. $\square$

Proposition 6.8. *Let $\mathcal{C}$ be an operad augmented over the Barratt-Eccles operad. Then the $\mathcal{I}$-space lifting functor R induces a functor*

$$R\colon \mathcal{U}[\mathcal{C}]/BF_{h\mathcal{I}} \to \mathcal{I}\mathcal{U}[\mathcal{C}]/BF.$$

Proof. We give $BF_{h\mathcal{I}}$ the $\mathcal{C}$-action defined by the augmentation to $\mathcal{E}$. Let $f\colon X \to BF_{h\mathcal{I}}$ be a map of $\mathcal{C}$-spaces and consider the diagram

$$X \xrightarrow{f} BF_{h\mathcal{I}} \xleftarrow{\pi} \overline{BF}$$

defining $R_f(X)$. Pulling the $\mathcal{E}$-action on $\overline{BF}$ defined in Lemma 6.7 back to a $\mathcal{C}$-action, this is a diagram of $\mathcal{C}$-$\mathcal{I}$-spaces. Thus, $R_f(X)$ is a homotopy pullback of a diagram in $\mathcal{I}\mathcal{U}[\mathcal{C}]$, hence is itself an object in this category and the projections $R_f(X) \to X$ and $R_f(X) \to \overline{BF}$ are maps of $\mathcal{C}$-$\mathcal{I}$-spaces, see [28], Section 1. Since the equivalence $t\colon \overline{BF} \to BF$ is also a map of $\mathcal{C}$-$\mathcal{I}$-spaces, the conclusion follows. □

Combining this with Corollary 6.4 we get the following.

Corollary 6.9. *If $\mathcal{C}$ is an operad that is augmented over $\mathcal{E}$, then the Thom spectrum functor on $\mathcal{U}/BF_{h\mathcal{I}}$ induces a functor $T\colon \mathcal{U}[\mathcal{C}]/BF_{h\mathcal{I}} \to Sp^{\Sigma}[\mathcal{C}]$.* □

7. The Thom isomorphism

Let MF be the symmetric Thom spectrum associated to the identity $BF \to BF$, and let MSF be the symmetric Thom spectrum associated to the inclusion $BSF \to BF$. Here $SF(n)$ denotes the submonoid of orientation preserving based homotopy equivalences (those that are homotopic to the identity) and BSF is the corresponding commutative $\mathcal{I}$-space monoid. We first construct canonical orientations of these Thom spectra, and for this we need convenient models of Eilenberg-Mac Lane spectra.

7.1. Eilenberg-Mac Lane spectra and orientations. Let A be a discrete ring, and write $A[-]$ for the functor that to a topological space X associates the free topological A-module $A[X]$ generated by X, see e.g. [42], Section 2.3. In the special case where X is the realization of a simplicial set $X_\bullet$ this may be identified with the realization of the simplicial A-module $A[X_\bullet]$. If X is based, we write $A(X)$ for the topological A-module $A[X]/A[*]$. The functor $A(-)$ defined in this way is left adjoint to the forgetful functor from topological A-modules to based spaces. It is well-known that $A(S^n)$ is a model of the Eilenberg-Mac Lane space $K(A,n)$, and that when equipped with the obvious structure maps this defines a model of the Eilenberg-Mac Lane spectrum for A as a symmetric ring spectrum. In order to define the orientations, we shall consider a variant of this construction. Let $F_A(n)$ be the topological monoid of continuous A-linear endomorphisms of $A(S^n)$ and notice that by the above remarks this is homotopy equivalent to A considered as a discrete multiplicative monoid. Writing $SF_A(n)$ for the connected component corresponding to the unit of A, this is then a contractible topological monoid. Applying the bar construction as in Section 2, we get a well-based quasifibration

$$B(*, SF_A(n), A(S^n)) \to BSF_A(n),$$

and we define the Eilenberg-Mac Lane spectrum HA to be the symmetric spectrum with nth space

$$HA(n) = B(*, SF_A(n), A(S^n))/BSF_A(n).$$

It is easy to check that this is a commutative symmetric ring spectrum which is level-wise equivalent to the usual model for the Eilenberg-Mac Lane spectrum considered above. Since HA is *flat* in the sense of [4], the functor $HA \wedge (-)$ preserves stable equivalences between well-based spectra; this follows from a slight refinement of the argument used in [4]. (Alternatively, one can check that the arguments in [35], Proposition 5.14, works equally well with Quillen cofibrations of spaces replaced by our notion of (h-)cofibrations.) Let now $A = \mathbb{Z}/2$ and observe that the functor $\mathbb{Z}/2(-)$ defines a map of sectioned quasifibrations

$$B(*, F(n), S^n) \to B(*, SF_{\mathbb{Z}/2}(n), \mathbb{Z}/2(n)).$$

The canonical orientation of MF is the induced map of commutative symmetric ring spectra $MF \to H\mathbb{Z}/2$. Similarly, the functor $\mathbb{Z}(-)$ defines a map of sectioned quasifibrations

$$B(*, SF(n), S^n) \to B(*, SF_{\mathbb{Z}}(n), \mathbb{Z}(S^n))$$

and the canonical orientation of MSF is the induced map of commutative symmetric ring spectra $MSF \to H\mathbb{Z}$.

7.2. The Thom isomorphism. We first consider the Thom isomorphism with $\mathbb{Z}/2$-coefficients. Given a map $f\colon X \to BF_{h\mathcal{I}}$, the $\mathcal{I}$-space lift $R_f(X) \to BF$ induces a map of symmetric spectra $T(f) \to MF$, and we define the $H\mathbb{Z}/2$-orientation of $T(f)$ to be the composition

$$T(f) \to MF \to H\mathbb{Z}/2.$$

As explained in Section 1.4, the orientation induces a map of symmetric spectra

$$T(f) \wedge H\mathbb{Z}/2 \to X_+ \wedge H\mathbb{Z}/2. \tag{7.1}$$

Since our construction of the Thom spectrum functor has good properties both formally and homotopically, the proof that this is a stable equivalence is almost completely formal.

Theorem 7.2. *The map of symmetric spectra (7.1) is a stable equivalence.*

Proof. Both functor in the theorem are homotopy functors on $\mathcal{U}/BF_{h\mathcal{I}}$ in the sense that they take weak homotopy equivalences to stable equivalences; this follows from Corollary 4.13 and the fact that $H\mathbb{Z}/2$ is flat. Thus, we may assume that X is a CW-complex and consider the filtration of X by skeleta

X^n such that X^{-1} is the empty set and X^n is homeomorphic to the pushout of a diagram of the form

$$X^{n-1} \leftarrow \coprod S^{n-1} \to \coprod D^n.$$

Since both functors in the theorem preserve pushouts and h-cofibrations, it suffices by [27], Theorem 8.12, to consider the case where the domain of f is of the form D^n or S^n. If f is the inclusion of the basepoint $* \to BF_{h\mathcal{I}}$, then the unit of the $\mathcal{I}$-space monoid $R_f(*)$ gives a stable equivalence $S \to T(f)$ and the composition

$$S \wedge H\mathbb{Z}/2 \xrightarrow{\sim} T(f) \wedge H\mathbb{Z}/2 \to *_+ \wedge H\mathbb{Z}/2$$

is the identity on $H\mathbb{Z}/2$. Using the homotopy invariance of the Thom spectrum functor, this easily implies the result for D^n. Identifying S^n with the pushout of the diagram $D^n \leftarrow S^{n-1} \to D^n$, the result for S^n then follows by an inductive argument. □

7.3. The integral Thom isomorphism. Using the commutative $\mathcal{I}$-space monoid BSF instead of BF, we get a a monoidal $\mathcal{I}$-space lifting functor

$$R\colon \mathcal{U}/BSF_{h\mathcal{I}} \to \mathcal{I}\mathcal{U}/BSF$$

defined in analogy with the $\mathcal{I}$-space lifting functor on $\mathcal{U}/BF_{h\mathcal{I}}$. The two lifting functors are related by a diagram

$$\begin{array}{ccc} \mathcal{U}/BSF_{h\mathcal{I}} & \xrightarrow{R} & \mathcal{I}\mathcal{U}/BSF \\ \downarrow & & \downarrow \\ \mathcal{U}/BF_{h\mathcal{I}} & \xrightarrow{R} & \mathcal{I}\mathcal{U}/BF, \end{array}$$

which is commutative up to natural $\mathcal{I}$-equivalence. Thus, the two natural ways to define a Thom spectrum functor on $\mathcal{U}/BSF_{h\mathcal{I}}$ are equivalent up to stable equivalence. For the definition of orientations it is most convenient to define the Thom spectrum functor on $\mathcal{U}/BSF_{h\mathcal{I}}$ to be the composition

$$T\colon \mathcal{U}/BSF_{h\mathcal{I}} \xrightarrow{R} \mathcal{I}\mathcal{U}/BSF \to \mathcal{I}\mathcal{U}/BF \xrightarrow{T} Sp^{\Sigma}.$$

With this definition we have a canonical integral orientation of the Thom spectrum associated to a map $X \to BSF_{h\mathcal{I}}$, defined by the composition $T(f) \to MSF \to H\mathbb{Z}$. The orientation again gives rise to a map of symmetric spectra

$$T(f) \wedge H\mathbb{Z} \to X_+ \wedge H\mathbb{Z} \tag{7.3}$$

and the proof of the integral version of the Thom isomorphism theorem is completely analogous to the $H\mathbb{Z}/2$-version.

Theorem 7.4. *The map (7.3) is a stable equivalence.* □

We can now verify the claim in Theorem 1.8 that the Thom equivalence is strictly multiplicative.

Proof of Theorem 1.8. Let H denote either one of the commutative symmetric ring spectra $H\mathbb{Z}/2$ or $H\mathbb{Z}$, and view H as an object in $Sp^{\Sigma}[\mathcal{C}]$ by projecting $\mathcal{C}$ onto the commutativity operad. We claim that (1.6) is a diagram in $Sp^{\Sigma}[\mathcal{C}]$ when we give each of the terms the diagonal $\mathcal{C}$-action. For the first two maps this follows from Proposition 6.8 and Corollary 6.9, which imply that the Thom diagonal and $T(f) \to MF$ (or $T(f) \to MSF$) are both $\mathcal{C}$-maps. For the last map the claim follows from the fact that the multiplication $H \wedge H \to H$ is a map of commutative symmetric ring spectra, hence in particular a $\mathcal{C}$-map. □

8. Symmetrization of diagram Thom spectra

In this section we first generalize the definition of the symmetric Thom spectrum functor to other types of diagram spectra. We then show how the results in the previous sections can be used to turn such diagram Thom spectra into symmetric spectra.

8.1. Diagram spaces and diagram Thom spectra. Given a small category $\mathcal{D}$, we define a *D-space* to be a functor $X\colon \mathcal{D} \to \mathcal{U}$ and we write $\mathcal{D}\mathcal{U}$ for the category of such functors. Suppose that we are given a functor $\phi\colon \mathcal{D} \to \mathcal{I}$. Then we can generalize the notion of a symmetric spectrum by introducing the topological category $\mathcal{D}_S$ that has the same objects as $\mathcal{D}$, but whose morphism spaces are defined by

$$\mathcal{D}_S(a,b) = \bigvee_{\alpha \in \mathcal{D}(a,b)} S^{b-\alpha},$$

where $S^{b-\alpha}$ is shorthand notation for $S^{\phi(b)-\phi(\alpha)}$, cf. Section 3.1. The composition is defined as for $\mathcal{I}_S$. We define a *$\mathcal{D}$-spectrum* to be a continuous based functor $\mathcal{D}_S \to \mathcal{T}$ and we write $\mathcal{D}_S\mathcal{T}$ for the category of such functors. Thus, a $\mathcal{D}$-spectrum is given by a family of based spaces $X(a)$ indexed by the objects a in $\mathcal{D}$, together with a family of based structure maps $S^{b-\alpha} \wedge X(a) \to X(b)$ indexed by the morphisms $\alpha\colon a \to b$ in $\mathcal{D}$. It is required (i) that the structure map associated to an identity morphism $1_a\colon a \to a$ is the canonical identification $S^0 \wedge X(a) \to X(a)$, and (ii) that given a pair of composable morphisms $\alpha\colon a \to b$ and $\beta\colon b \to c$, the following diagram is commutative

$$\begin{array}{ccc}
S^{c-\beta} \wedge S^{b-\alpha} \wedge X(a) & \longrightarrow & S^{c-\beta} \wedge X(b) \\
\downarrow & & \downarrow \\
S^{c-\beta\alpha} \wedge X(a) & \longrightarrow & X(c).
\end{array}$$

In particular, if ϕ denotes the identity functor on $\mathcal{I}$, then $\mathcal{I}_S\mathcal{T}$ is an alternative notation for the category of symmetric spectra. Suppose now that $\mathcal{D}$ has the structure of a strict monoidal category. As in the case of $\mathcal{I}$-spaces, $\mathcal{D}\mathcal{U}$ inherits a monoidal structure from $\mathcal{D}$ which is symmetric monoidal if $\mathcal{D}$ is. If in addition ϕ is strict monoidal, then the monoidal structure of $\mathcal{D}$ also induces a monoidal structure on $\mathcal{D}_S$ which is symmetric monoidal if $\mathcal{D}$ and ϕ are. This in turn induces a monoidal structure on the category of $\mathcal{D}$-spectra $\mathcal{D}_S\mathcal{T}$ which again is symmetric monoidal if $\mathcal{D}$ and ϕ are. The $\mathcal{I}$-space BF pulls back to a $\mathcal{D}$-space via ϕ and the definition of the symmetric Thom spectrum functor immediately generalizes to give a Thom spectrum functor

$$T\colon \mathcal{DU}/BF \to \mathcal{D}_S\mathcal{T}.$$

The proof of Theorem 1.1 generalizes to show that this is a strong monoidal functor which is symmetric monoidal if $\mathcal{D}$ and ϕ are.

8.2. Examples of diagram Thom spectra. Many examples of Thom spectra arise from compatible families of groups over the topological monoids $F(n)$. It often happens that such a family defines a $\mathcal{D}$-diagram of groups for some strict monoidal category $\mathcal{D}$ over $\mathcal{I}$ and if the induced maps of classifying spaces define a $\mathcal{D}$-space over BF we get an associated $\mathcal{D}$-Thom spectrum. We begin by fixing notation for some of the relevant categories. For each $k \geq 1$ we have the strict symmetric monoidal faithful functor

$$\psi_k\colon \mathcal{I} \to \mathcal{I}, \quad \mathbf{n} \mapsto \underbrace{\mathbf{k} \sqcup \cdots \sqcup \mathbf{k}}_{n}.$$

and we write $\mathcal{I}[k]$ for its image in $\mathcal{I}$. Thus, $\mathcal{I}[k]$ is a strict symmetric monoidal category whose objects have cardinality a multiple of k, and whose morphisms permute blocks of k letters simultaneously. Let us write $\mathcal{M}$ for the subcategory of injective order preserving morphisms in $\mathcal{I}$. This inherits a strict monoidal (but not symmetric monoidal) structure from $\mathcal{I}$ and we similarly define monoidal subcategories $\mathcal{M}[k]$ for $k \geq 1$.

Example 8.1 (The classical groups). The orthogonal groups $O(n)$ and the special orthogonal groups $SO(n)$ define the commutative $\mathcal{I}$-space monoids BO and BSO that give rise to the commutative symmetric ring spectra MO and MSO. The unitary groups $U(n)$ and the special unitary groups $SU(n)$ define the commutative $\mathcal{I}[2]$-space monoids BU and BSU that give rise to the commutative $\mathcal{I}[2]$-ring spectra MU and MSU. The symplectic groups $Sp(n)$ define the commutative $\mathcal{I}[4]$-space monoid BSp that gives rise to the commutative $\mathcal{I}[4]$-spectrum MSp.

Example 8.2 (Discrete groups and $\mathcal{I}$-spaces). The symmetric groups Σ_n define the commutative $\mathcal{I}$-space monoid $B\Sigma$ in which the monoid structure is induced by concatenation of permutations. This gives rise to the commutative symmetric ring spectrum $M\Sigma$ whose associated bordism theory has been studied by Bullett [9]. Other systems of discrete groups that give rise to symmetric ring spectra include the general linear groups $GL_n(\mathbb{Z})$, the groups $(\mathbb{Z}/2)^n$ of diagonal matrices with entries ± 1, and the groups $\Sigma_n \wr \mathbb{Z}/2$ of permutation matrices with entries ± 1. For details and more examples, see [9] and [12].

Example 8.3 (Braid groups and $\mathcal{M}$-spaces). The family of braid groups $\mathfrak{B}(n)$ defines an $\mathcal{M}$-space monoid $B\mathfrak{B}$ in a natural way. We refer to [5] for the definition and basic properties of the braid groups. If we view an element of $\mathfrak{B}(n)$ as a system of n strings in the usual way, then the monoid structure on

$B\mathfrak{B}$ is induced by concatenation of such systems. Let ρ denote the sequence of monoid homomorphisms

$$\rho_n\colon \mathfrak{B}(n) \to \Sigma_n \to F(n)$$

where the first map takes a system of strings to the induced permutation of the endpoints and the second map in the canonical inclusion. This defines a map $B\rho\colon B\mathfrak{B} \to BF$ of $\mathcal{M}$-space monoids and we write $M\mathfrak{B}$ for the associated $\mathcal{M}$-Thom ring spectrum. The underlying spectrum of $M\mathfrak{B}$ has been analyzed in [9] and [10], where it is shown to be equivalent to the Eilenberg-Mac Lane spectrum $H\mathbb{Z}/2$. Suppose that G is an $\mathcal{M}$-diagram of groups over the monoids $F(n)$ and that the homomorphisms ρ_n can be factored as $\mathfrak{B}(n) \to G(n) \to F(n)$. If the $\mathcal{M}$-space BG admits a monoid structure such that the induced map $B\mathfrak{B} \to BG$ is a map of $\mathcal{M}$-space monoids over BF it then follows that MG is an $\mathcal{M}$-module spectrum over $M\mathfrak{B}$. For example, this applies to $M\Sigma$ and MO but not to MSO. We show how to symmetrize the constructions so as to get symmetric Thom spectra in Section 8.3. Again we refer to [9], [10], [12] for further examples.

Example 8.4 (Maps to $BF(k)$ and $\mathcal{I}[k]$-spaces). For our next class of examples we need some preliminary definitions. Let X be a based space and let $X^\bullet$ be the $\mathcal{I}$-space defined by $\mathbf{n} \mapsto X^n$. Given a morphism $\alpha\colon \mathbf{m} \to \mathbf{n}$, the induced map $\alpha_*\colon X^m \to X^n$ is defined by

$$\alpha_*(x_1, \dots, x_m) = (x_{\alpha^{-1}(1)}, \dots, x_{\alpha^{-1}(n)}),$$

with the convention that $x_\emptyset$ is the base point in X. We give $X^\bullet$ the structure of a commutative $\mathcal{I}$-space monoid using the identifications $X^m \times X^n = X^{m+n}$. Suppose now that $f\colon X \to BF(k)$ is a based map. Then we view $X^\bullet$ as an $\mathcal{I}[k]$-space via the isomorphism $\psi_k\colon \mathcal{I} \to \mathcal{I}[k]$, and the maps

$$X^n \to BF(k)^n \xrightarrow{\mu} BF(\underbrace{\mathbf{k} \sqcup \dots \sqcup \mathbf{k}}_{n})$$

define a map of $\mathcal{I}[k]$-space monoids $X^\bullet \to BF$, where we view BF as an $\mathcal{I}[k]$-space by restriction. We write $MX^{\wedge\infty}$ for the associated commutative $\mathcal{I}[k]$-ring spectrum, the function f being understood. In the cases $X = BO(1)$ and $X = BU(1)$, we get the Thom spectra $MO(1)^{\wedge\infty}$ and $MU(1)^{\wedge\infty}$ that represent the bordism theories of manifolds with stable normal bundle given as an ordered sum of real or complex line bundles. These Thom spectra have been analyzed by Arthan and Bullett [1], [9]. Letting $X = BO(k)$ or $X = BU(k)$, we similarly get the $\mathcal{I}[k]$-spectrum $MO(k)^{\wedge\infty}$ and the $\mathcal{I}[2k]$-spectrum $MU(k)^{\wedge\infty}$.

8.3. Symmetrization of diagram Thom spectra. As demonstrated in the last section, many Thom spectra naturally arise as $\mathcal{D}$-Thom spectra associated to maps of $\mathcal{D}$-spaces $f\colon X \to BF$ for suitable monoidal categories $\mathcal{D}$ over $\mathcal{I}$. In the applications it is often convenient to replace such a $\mathcal{D}$-Thom spectrum by a symmetric Thom spectrum and our preferred way of doing is to first transform f to a map of $\mathcal{I}$-spaces and then evaluate the symmetric Thom spectrum functor on this transformed map. We shall discuss two ways of performing this "symmetrization" procedure: in this section we consider symmetrizations using the $\mathcal{I}$-space lifting functor R and in the next section we consider symmetrizations via (homotopy) Kan extension.

For simplicity we shall from now on assume that $\mathcal{D}$ is a monoidal subcategory of $\mathcal{I}$ such that the intersection $\mathcal{D}\cap\mathcal{N}$ is a cofinal subcategory of $\mathcal{N}$. Thus, any $\mathcal{D}$-spectrum has an underlying $\mathcal{D}\cap\mathcal{N}$-spectrum and we define the spectrum homotopy groups in the usual way by evaluating the colimit of the associated $\mathcal{D}\cap\mathcal{N}$-diagram of homotopy groups. Given a map $f\colon X\to BF$, we write (by abuse of notation) $f_{h\mathcal{D}}$ for the composite map

$$f_{h\mathcal{D}}\colon X_{h\mathcal{D}}\xrightarrow{f_{h\mathcal{D}}} BF_{h\mathcal{D}}\to BF_{h\mathcal{I}}.$$

Applying the $\mathcal{I}$-space lifting functor R to this map we get a functor

$$\text{(8.5)}\qquad \mathcal{DU}/BF\to\mathcal{IU}/BF,\qquad (X\xrightarrow{f}BF)\mapsto(R_{f_{h\mathcal{D}}}(X_{h\mathcal{D}})\xrightarrow{R(f_{h\mathcal{D}})}BF).$$

We say that a map of $\mathcal{D}$-spaces $X\to Y$ is a *$\mathcal{D}$-equivalence* if the induced map $X_{h\mathcal{D}}\to Y_{h\mathcal{D}}$ is a weak homotopy equivalence.

Lemma 8.6. *The restriction of $R_{f_{h\mathcal{D}}}(X_{h\mathcal{D}})$ to a $\mathcal{D}$-space is related to X by a chain of $\mathcal{D}$-equivalences over BF.*

Proof. In analogy with the case of $\mathcal{I}$-spaces considered in Section 5, there is a diagram of $\mathcal{D}$-equivalences $X\leftarrow\overline{X}\to R_{f_{h\mathcal{D}}}(X_{h\mathcal{D}})$ over BF. □

It follows from the $\mathcal{D}$-space analogue of Bökstedt's approximation Lemma 4.4 that if $X\to Y$ is a $\mathcal{D}$-equivalence of convergent $\mathcal{D}$-spaces X and Y, then the connectivity of the maps $X(d)\to Y(d)$ tends to infinity with d. The previous lemma therefore has the following consequence.

Proposition 8.7. *If X is a convergent $\mathcal{D}$-space and $f\colon X\to BF$ is a map of $\mathcal{D}$-spaces which is level-wise T-good, then the restriction of the symmetric spectrum $T(R(f_{h\mathcal{D}}))$ to a $\mathcal{D}$-spectrum is π_*-equivalent to $T(f)$.* □

This construction preserves multiplicative structures in the sense that if X is a $\mathcal{D}$-space monoid and $f\colon X\to BF$ a map of $\mathcal{D}$-space monoids, then $f_{h\mathcal{D}}$ is a map of topological monoids and $T(f_{h\mathcal{D}})$ is a symmetric ring spectra by Lemma 4.19. In the following we consider the effect of applying the construction to the examples considered in the previous section.

Example 8.8. Let X be an $\mathcal{I}[k]$-space with an action of an operad $\mathcal{C}$ that is augmented over the Barratt-Eccles operad. If $f\colon X\to BF$ is a map of $\mathcal{C}$-$\mathcal{I}[k]$-spaces, then the induced map

$$f_{h\mathcal{I}[k]}\colon X_{h\mathcal{I}[k]}\to BF_{h\mathcal{I}[k]}\to BF_{h\mathcal{I}}$$

is map of $\mathcal{C}$-spaces and it follows from Corollary 6.9 that the symmetric spectrum $T(f_{h\mathcal{I}[k]})$ inherits a $\mathcal{C}$-action. Here we use the canonical isomorphism of categories $\psi_k\colon\mathcal{I}\to\mathcal{I}[k]$ to identify $X_{h\mathcal{I}[k]}$ with $(\psi_k^*X)_{h\mathcal{I}}$, and we transfer the $\mathcal{C}$-action on $(\psi_k^*X)_{h\mathcal{I}}$ defined in Corollary 6.6 to $X_{h\mathcal{I}[k]}$ via this identification. This applies in particular to the map of commutative $\mathcal{I}[2]$-spaces $BU\to BF$

to give a model of the Thom spectrum MU as a symmetric ring spectrum with an action of the Barratt-Eccles operad. We shall see how to realize MU as a strictly commutative symmetric ring spectrum in Example 8.17.

Example 8.9. Let as before $B\mathfrak{B}$ denote the $\mathcal{M}$-space monoid defined by the braid groups. We shall identify the map $B\mathfrak{B}_{h\mathcal{N}} \to B\mathfrak{B}_{h\mathcal{M}}$ in terms of Quillen's plus construction. Firstly, it follows from the homological stability of the braid groups (see [11], III, Appendix) and the homological version of Lemma 4.4, that this map is a homology isomorphism. Secondly, the monoidal structure of $\mathcal{M}$ gives $B\mathfrak{B}_{h\mathcal{M}}$ the structure of a topological monoid, which in particular implies that its fundamental group is abelian. Thus, the map in question has the effect of abelianizing the fundamental group. The space $B\mathfrak{B}_{h\mathcal{N}}$ may be identified with the classifying space of the infinite braid group $\mathfrak{B}(\infty)$. Since the commutator subgroup of the latter is perfect, it follows from the above that we may identify $B\mathfrak{B}_{h\mathcal{M}}$ with Quillen's plus construction $B\mathfrak{B}^+_{h\mathcal{N}}$. It is proved in [10] that there is a homotopy commutative diagram

$$\begin{array}{ccc} B\mathfrak{B}_{h\mathcal{N}} & \xrightarrow{\theta} & \Omega^2(S^3) \\ \big\downarrow{\scriptstyle (B\rho)_{h\mathcal{N}}} & & \big\downarrow{\scriptstyle \eta} \\ BF_{h\mathcal{N}} & = & BF_{h\mathcal{N}}, \end{array}$$

where η denotes the "Mahowald orientation", that is, the extension of the non-trivial map $S^1 \to BF_{h\mathcal{N}}$ to a 2-fold loop map. It is a theorem of Mahowald [24], that the Thom spectrum of η is stably equivalent to the Eilenberg-Mac Lane spectrum $H\mathbb{Z}/2$. By the universal property of the plus construction, we conclude from the above that there is a homotopy commutative diagram

$$\begin{array}{ccc} B\mathfrak{B}_{h\mathcal{M}} & \xrightarrow{\sim} & \Omega^2(S^3) \\ \big\downarrow{\scriptstyle (B\rho)_{h\mathcal{M}}} & & \big\downarrow{\scriptstyle \eta} \\ BF_{h\mathcal{I}} & = & BF_{h\mathcal{I}}, \end{array}$$

where the upper map is a homotopy equivalence as indicated. Consequently, the symmetric ring spectrum $T((B\rho)_{h\mathcal{M}})$ is a model of $H\mathbb{Z}/2$.

Example 8.10. Let $BGL(\mathbb{Z})$ be the commutative $\mathcal{I}$-space monoid associated to the general linear groups $GL_n(\mathbb{Z})$. As in the case of the braid groups, we may identify $BGL(\mathbb{Z})_{h\mathcal{N}} \to BGL(\mathbb{Z})_{h\mathcal{I}}$ in terms of Quillen's plus construction. Indeed, by the homological stability of the groups $GL_n(\mathbb{Z})$, it follows that this map is a homology equivalence. Since $BGL(\mathbb{Z})_{h\mathcal{I}}$ is a topological monoid it has abelian fundamental group, hence it may be identified with $BGL_\infty(\mathbb{Z})^+$; the base point component of Quillen's algebraic K-theory space. In similar fashion, starting with the $\mathcal{I}$-space $B\Sigma$, we may identify $B\Sigma_{h\mathcal{I}}$ with $B\Sigma^+_\infty$, which by the Barratt-Priddy-Quillen Theorem is equivalent to the base point component of $Q(S^0)$.

Example 8.11. Let $f\colon X \to BF(k)$ be a based map and consider the associated map of $\mathcal{I}[k]$-spaces $X^\bullet \to BF$. It is proved in [34] that if X is well-based and connected, then $X^\bullet_{h\mathcal{I}}$ is a model of the infinite loop space $Q(X)$. Identifying $X^\bullet_{h\mathcal{I}}$ with $X^\bullet_{h\mathcal{I}[k]}$ via the isomorphism ψ_k, it follows as in Example 8.8 that the induced map $X^\bullet_{h\mathcal{I}[k]} \to BF_{h\mathcal{I}}$ is a map of $\mathcal{E}$-spaces which models the usual extension of f to a map of infinite loop spaces. If we instead think of $X^\bullet$ as an $\mathcal{M}$-space by restriction, then one can show that $X_{h\mathcal{M}}$ is homotopy equivalent to the colimit $X^\bullet_{\mathcal{M}}$, that is, to the free based monoid generated by X. By a theorem of James [18] the latter is a model of $\Omega\Sigma(X)$, and the map $X^\bullet_{h\mathcal{M}} \to X^\bullet_{h\mathcal{I}}$ corresponds to the inclusion of $\Omega\Sigma(X)$ in $Q(X)$.

Example 8.12. Let $\mathcal{E}_k$ be the kth stage of the Smith filtration of the Barratt-Eccles operad $\mathcal{E}$ and write $X \mapsto E_k(X)$ for the associated monad on based spaces, see [3], [38]. Thus, $\mathcal{E}_k$ is equivalent to the little k-cubes operad, and if X is a well-based connected space, then $E_k(X)$ is a combinatorial model of $\Omega^k\Sigma^k(X)$. Given a based map $f\colon X \to BF_{h\mathcal{I}}$, we use that $\mathcal{E}_k$ is augmented over $\mathcal{E}$ to extend f to a map of $\mathcal{E}_k$-spaces

$$E_k(f)\colon E_k(X) \to E_k(BF_{h\mathcal{I}}) \to BF_{h\mathcal{I}},$$

which for connected X is a model of the usual extension of f to a k-fold loop map. It follows that the associated symmetric Thom spectrum $T(E_k(f))$ is equipped with an $\mathcal{E}_k$-action.

8.4. Symmetrization via Kan Extension. Let again $\mathcal{D}$ be a monoidal subcategory of $\mathcal{I}$ such that $\mathcal{D} \cap \mathcal{N}$ is cofinal in $\mathcal{N}$ and let us write $j\colon \mathcal{D} \to \mathcal{I}$ for the inclusion. We first consider homotopy Kan extensions along j. Recall that given a $\mathcal{D}$-space X, the homotopy Kan extension is the $\mathcal{I}$-space $j^h_*(X)$ defined by

$$j^h_*(X)(n) = \operatorname*{hocolim}_{(j\downarrow\mathbf{n})} X \circ \pi_n,$$

see Appendix A.1. The functor $j^h_*(-)$ induces a functor

$$j^h_*\colon \mathcal{DU}/BF \to \mathcal{IU}/BF,\ (f\colon X \to BF) \mapsto (j^h_*(f)\colon j^h_*(X) \to j^h_*(BF) \to BF)$$

which is $\mathcal{I}$-equivalent to the functor (8.5) in the sense of the following lemma.

Lemma 8.13. *There is a natural $\mathcal{I}$-equivalence $j^h_*(X) \to R_{f_{h\mathcal{D}}}(X_{h\mathcal{D}})$ of $\mathcal{I}$-spaces over BF.*

Proof. Notice first that there is a commutative diagram

$$\begin{tikzcd} j^h_*(X) \arrow[r] \arrow[d] & \overline{BF} \arrow[d] \\ X_{h\mathcal{D}} \arrow[r] & BF_{h\mathcal{I}}, \end{tikzcd}$$

inducing a map of $\mathcal{I}$-spaces $j^h_*(X) \to R_{f_{h\mathcal{D}}}(X_{h\mathcal{D}})$ over BF. Since this is also a map over $X_{h\mathcal{D}}$, the result follows from the fact that $j^h_*(X) \to X_{h\mathcal{D}}$ and $R_{f_{h\mathcal{D}}}(X_{h\mathcal{D}}) \to X_{h\mathcal{D}}$ are $\mathcal{I}$-equivalences, see Lemma A.2 and the proof of Proposition 4.9. $\square$

We now turn to (categorical) Kan extensions. Given a $\mathcal{D}$-space, the Kan extension $j_*(X)$ is defined as the homotopy Kan extension except that we use the colimit instead of the homotopy colimit, that is,

$$j_*(X)(n) = \operatorname*{colim}_{(j\downarrow\mathbf{n})} X \circ \pi_{\mathbf{n}}.$$

The functor j_* is left adjoint to the functor that pulls an $\mathcal{I}$-space back to a $\mathcal{D}$-space via j and it induces a functor

$$j_*\colon \mathcal{D}\mathcal{U}/BF \to \mathcal{I}\mathcal{U}/BF, \qquad (X \to BF) \mapsto (j_*(X) \to j_*(BF) \to BF)$$

where the map $j_*(BF) \to BF$ is the counit of the adjunction. This functor is strong monoidal and is symmetric monoidal if $\mathcal{D}$ and j are. Thus, in the latter case it takes commutative $\mathcal{D}$-space monoids to commutative $\mathcal{I}$-space monoids. The drawback of using the categorical Kan extension is of course that it is homotopically well-behaved only under suitable cofibration conditions on the $\mathcal{D}$-space X and the main purpose of this section is to formulate such conditions. More precisely, we shall consider an inclusion $j\colon \mathcal{D} \to \mathcal{I}$ of a (not necessarily monoidal) subcategory $\mathcal{D}$ of $\mathcal{I}$ and we shall formulate conditions on $\mathcal{D}$ and X which ensure that the canonical map $j_*^h(X) \to j_*(X)$ is a level-wise equivalence. Given an object $\mathbf{d}_0$ of $\mathcal{D}$, consider the category $(\mathcal{D} \downarrow \mathbf{d}_0)$ of objects in $\mathcal{D}$ over $\mathbf{d}_0$ and let $\partial\mathbf{d}_0$ be the subcategory obtained by excluding the terminal objects.

Lemma 8.14. *Let $j\colon \mathcal{D} \to \mathcal{I}$ be the inclusion of a subcategory $\mathcal{D}$ and suppose that X is a $\mathcal{D}$-space such that the map*

$$\operatorname*{colim}_{\partial\mathbf{d}_0} X \circ \pi_{\mathbf{d}_0} \to \operatorname*{colim}_{(\mathcal{D}\downarrow\mathbf{d}_0)} X \circ \pi_{\mathbf{d}_0} = X(d_0)$$

is a cofibration for all objects $\mathbf{d}_0$ in $\mathcal{D}$. Then $j_*^h(X) \to j_*(X)$ is a level-wise equivalence.

Proof. Notice first that the category $(j \downarrow \mathbf{n})$ is a preorder in the sense that the morphism sets have at most one element. Choosing a representative for each isomorphism class we get an equivalent skeleton subcategory $\mathcal{A}(\mathbf{n})$ (in fact a partially ordered set), and it suffices to show that the map

$$\operatorname*{hocolim}_{\mathcal{A}(\mathbf{n})} X \circ \pi_{\mathbf{n}} \to \operatorname*{colim}_{\mathcal{A}(\mathbf{n})} X \circ \pi_{\mathbf{n}}$$

is a weak homotopy equivalence. The advantage of this is that the category $\mathcal{A}(\mathbf{n})$ is *very small* in the sense that its nerve only has finitely many non-degenerate simplices. In this situation there is a general model categorical criterion for comparing the homotopy colimit to the colimit, see [13], Section 10. Working in the Strøm model category on $\mathcal{U}$ [41], we must check that for each object a in $\mathcal{A}(n)$ the map

$$\operatorname*{colim}_{\partial a} X \circ \pi_{\mathbf{n}} \circ \pi_a \to \operatorname*{colim}_{(\mathcal{A}(n)\downarrow a)} X \circ \pi_{\mathbf{n}} \circ \pi_a = X(\pi_{\mathbf{n}}(a))$$

is a cofibration. Here we use the notation ∂a for the subcategory of $(\mathcal{A}(\mathbf{n}) \downarrow a)$ obtained by excluding the terminal object. It remains to see that if a is an object of the form $\mathbf{d}_0 \to \mathbf{n}$, then this criterion is the same as that stated in the lemma. On the one hand we may view $(\mathcal{A}(n) \downarrow a)$ as a skeleton subcategory of $((j \downarrow \mathbf{n}) \downarrow a)$ and on the other hand we may identify the latter category with $(\mathcal{D} \downarrow \mathbf{d}_0)$. Taken together this gives a homeomorphism

$$\operatorname*{colim}_{\partial a} X \circ \pi_{\mathbf{n}} \circ \pi_a \cong \operatorname*{colim}_{\partial \mathbf{d}_0} X \circ \pi_{\mathbf{d}_0}$$

and the conclusion follows. □

The criterion in Lemma 8.14 is not very practical and in order to have a more convenient formulation we impose conditions on the subcategory $\mathcal{D}$ of $\mathcal{I}$. We say that $\mathcal{D}$ has the *intersection property* if each diagram in $\mathcal{D}$ of the form

$$\mathbf{d}_1 \xrightarrow{\delta_1} \mathbf{d}_{12} \xleftarrow{\delta_2} \mathbf{d}_2$$

can be completed to a commutative square

$$\begin{tikzcd} \mathbf{d}_0 \arrow[r] \arrow[d] & \mathbf{d}_1 \arrow[d, "\delta_1"] \\ \mathbf{d}_2 \arrow[r, "\delta_2"] & \mathbf{d}_{12} \end{tikzcd} \tag{8.15}$$

in $\mathcal{D}$ such that the image of the composite morphism equals the intersection of the images of δ_1 and δ_2. For example, the monoidal subcategories $\mathcal{I}[k]$ and $\mathcal{J}[k]$ have the intersection property for all $k \geq 1$. We say that a $\mathcal{D}$-space X is *intersection cofibrant* if for any diagram of the form (8.15), such that the intersection of the images of δ_1 and δ_2 equals the image of the composite morphism, the induced map

$$X(d_1) \cup_{X(d_0)} X(d_2) \to X(d_{12})$$

is a cofibration. By Lillig's union theorem [21] for cofibrations, this is equivalent to the requirement that (i) any morphism $\mathbf{d}_1 \to \mathbf{d}_2$ in $\mathcal{D}$ induces a cofibration $X(d_1) \to X(d_2)$, and (ii) that the intersection of the images of $X(d_1)$ and $X(d_2)$ in $X(d_{12})$ equals the image of $X(d_0)$.

Proposition 8.16. *Let $j\colon \mathcal{D} \to \mathcal{I}$ be the inclusion of a subcategory $\mathcal{D}$ which has the intersection property and let X be a $\mathcal{D}$-space which is intersection cofibrant. Then the map $j_*^h(X) \to j_*(X)$ is a level-wise equivalence.*

Proof. We show that the assumptions on $\mathcal{D}$ and X imply that the criterion in Lemma 8.14 is satisfied. Given an object $\mathbf{d}_0$ in $\mathcal{D}$, consider the *range functor*

$$r\colon (\mathcal{D} \downarrow \mathbf{d}_0) \to (\mathcal{I} \downarrow \mathbf{d}_0) \to \mathcal{P}(\mathbf{d}_0), \quad r(\mathbf{d} \xrightarrow{\delta} \mathbf{d}_0) = \delta(\mathbf{d}) \subseteq \mathbf{d}_0,$$

where $\mathcal{P}(\mathbf{d}_0)$ denotes the category of subsets and inclusions in $\mathbf{d}_0$. The assumption that $\mathcal{D}$ has the intersection property implies that the image of r is a full subcategory of $\mathcal{P}(\mathbf{d}_0)$ that is closed under inclusions and that r defines an equivalence of categories between $(\mathcal{D} \downarrow \mathbf{d}_0)$ and its image. Thus, we might as well view $X \circ \pi_{\mathbf{d}_0}$ as a diagram $U \mapsto X(U)$ indexed on the objects U in a full subcategory $\mathcal{A}$ of $\mathcal{P}(\mathbf{d}_0)$ that is closed under intersections. By assumption (i) above we may view $X(U)$ as a closed subspace of $X(d_0)$ for all $U \in \mathcal{A}$ and by assumption (ii) we have the equality

$$X(U) \cap X(V) = X(U \cap V)$$

for all pairs of objects U and V in $\mathcal{A}$. It therefore follows from the gluing lemma for continuous functions on a union of closed subspaces that $\operatorname{colim}_{\partial \mathbf{d}_0} X \circ \pi_{\mathbf{d}_0}$ is homeomorphic to the union of the subspaces $X(U)$ of $X(d_0)$ for $U \neq \mathbf{d}_0$. The conclusion then follows from Lillig's union theorem for cofibrations [21]. □

Example 8.17. Let $j\colon \mathcal{I}[2] \to \mathcal{I}$ be the inclusion of the symmetric monoidal subcategory $\mathcal{I}[2]$. Since $\mathcal{I}[2]$ has the intersection property and the commutative $\mathcal{I}[2]$-space monoid BU is intersection cofibrant, it follows from Lemma 8.13 and Proposition 8.16 that there is a chain of $\mathcal{I}$-equivalences

$$j_*(BU) \xleftarrow{\sim} j_*^h(BU) \xrightarrow{\sim} R(BU_{h\mathcal{I}})$$

over BF. Thus, it follows from Proposition 8.7 together with Theorem 1.4 and Lemma 2.3 that applying the symmetric Thom spectrum functor to the commutative $\mathcal{I}$-space monoid $j_*(BU)$ gives a commutative symmetric ring spectrum which is a model of MU.

8.5. Orthogonal Thom spectra and diagram lifting. Recall from [27] that an orthogonal spectrum is a spectrum X such that the nth space $X(n)$ has an action of the orthogonal group $O(n)$, and such that the iterated structure maps $S^m \wedge X(n) \to X(m+n)$ are $O(m) \times O(n)$-equivariant. We write Sp^O for the category of orthogonal spectra. Let $\mathcal{V}$ be the topological category whose objects are the vector spaces $\mathbb{R}^n$, and whose morphisms are the linear isometries. A $\mathcal{V}$-space is a continuous functor $\mathcal{V} \to \mathcal{U}$, and we write $\mathcal{V}\mathcal{U}$ for the category of such functors. The symmetric monoidal structure of $\mathcal{V}$ induces a symmetric monoidal structure on $\mathcal{V}\mathcal{U}$ in the usual way, and the $\mathcal{I}$-space BF extends to a commutative $\mathcal{V}$-space monoid. Applying the Thom space construction level-wise as in the definition of the symmetric Thom spectrum functor, we get the symmetric monoidal orthogonal Thom spectrum functor

$$T\colon \mathcal{V}\mathcal{U}/BF \to Sp^O.$$

In order to construct orthogonal Thom spectra from space level data, we need a $\mathcal{V}$-space version

$$R\colon \mathcal{U}/BF_{h\mathcal{V}} \to \mathcal{V}\mathcal{U}/BF$$

of the $\mathcal{I}$-space lifting functor. Here the homotopy colimit $BF_{h\mathcal{V}}$ denotes the realization of the simplicial space

$$[k] \mapsto \coprod_{n_0,\dots,n_k} \mathcal{V}(\mathbb{R}^{n_1}, \mathbb{R}^{n_0}) \times \cdots \times \mathcal{V}(\mathbb{R}^{n_k}, \mathbb{R}^{n_{k-1}}) \times BF(n_k).$$

The statement in Lemma 4.4 remains true with $\mathcal{V}$ instead of $\mathcal{I}$, and we conclude from this that the canonical map $BF_{h\mathcal{N}} \to BF_{h\mathcal{V}}$ is a weak homotopy equivalence. The definition of the $\mathcal{V}$-space lifting functor is then completely analogous to the definition of the $\mathcal{I}$-space lifting functor in Section 4.2: Let $\overline{BF}$ be the $\mathcal{V}$-space defined by the homotopy Kan extension along the identity functor on

$\mathcal{V}$. Given a map $f\colon X \to BF_{h\mathcal{V}}$, we define $R_f(X)$ to be the homotopy pullback of the diagram of $\mathcal{V}$-spaces

$$X \xrightarrow{f} BF_{h\mathcal{V}} \xleftarrow{\pi} \overline{BF},$$

and we define $R(f)$ to be the composition

$$R(f)\colon R_f(X) \to \overline{BF} \to BF.$$

The Barratt-Eccles operad acts on $BF_{h\mathcal{V}}$ and the results on preservation of operad actions from Section 6 carry over to this setting.

Appendix A. Homotopy colimits

We here collect the facts about homotopy colimits needed in the paper. We shall adapt the definitions of Bousfield and Kan [7], except that we work with topological spaces instead of simplicial sets. Thus, given a small category $\mathcal{A}$ and an $\mathcal{A}$-diagram $X\colon \mathcal{A} \to \mathcal{U}$, the homotopy colimit $\operatorname{hocolim}_{\mathcal{C}} X$ is defined to be the realization of the simplicial space

$$[k] \mapsto \coprod_{a_0 \leftarrow \cdots \leftarrow a_k} X(a_k), \tag{A.1}$$

where the coproduct is over the k-simplices of the nerve $N_\bullet\mathcal{C}$. It is sometimes convenient to view this as the classifying space of the topological category $\mathcal{A}(X)$ whose space of objects is the disjoint union of the spaces $X(a)$ where a runs through the objects of $\mathcal{A}$. A morphism $(a,x) \to (a',x')$ in $\mathcal{A}(X)$ is specified by a morphism $\alpha\colon a \to a'$ in $\mathcal{A}$ such that $\alpha_* x = x'$. If X is a based $\mathcal{A}$-diagram, that is, a functor $X\colon \mathcal{A} \to \mathcal{T}$, then the inclusion of the base points gives a map $B\mathcal{A} \to B\mathcal{A}(X)$ and we define the based homotopy colimit to be the quotient space. Equivalently, this is the realization of the simplicial space obtained by replacing the disjoint union in (A.1) by the wedge product.

A.1. Homotopy Kan extensions. Let $\phi\colon \mathcal{A} \to \mathcal{B}$ be a functor between small categories. Given an $\mathcal{A}$-diagram X, the (left) homotopy Kan extension $\phi_*^h X$ is the $\mathcal{B}$-diagram defined by

$$\phi_*^h X(b) = \operatorname*{hocolim}_{(\phi\downarrow b)} X \circ \pi_b.$$

The homotopy colimit is over the category $(\phi \downarrow b)$ whose objects are pairs (a,β) in which a is an object of $\mathcal{A}$ and $\beta\colon \phi(a) \to b$ is a morphism in $\mathcal{B}$. A morphism $(a,\beta) \to (a',\beta')$ is given by a morphism $\alpha\colon a \to a'$ in $\mathcal{A}$ such that $\beta = \beta' \circ \phi(\alpha)$. The functor $\pi_b\colon (\phi \downarrow b) \to \mathcal{A}$ is defined by $(a,\beta) \mapsto a$. We recall that the categorical Kan extension $\phi_* X$ is defined using the categorical colimit instead of the homotopy colimit, see [22]. If $\mathcal{B}$ is the terminal category $*$ and $p\colon \mathcal{A} \to *$ the projection, then $p_* X$ and $p_*^h X$ are respectively the colimit and the homotopy colimit of the $\mathcal{A}$-diagram X. Notice, that the functors π_b define a map of $\mathcal{B}$-diagrams from $\phi_*^h X$ to the constant $\mathcal{B}$-diagram $\operatorname{hocolim}_{\mathcal{A}} X$. A proof of the following well-known lemma can be found in [34].

Lemma A.2. *The induced map*

$$\pi\colon \operatorname*{hocolim}_{\mathcal{B}} \phi^h_* X \xrightarrow{\sim} \operatorname*{hocolim}_{\mathcal{A}} X.$$

is a weak homotopy equivalence. □

This lemma may be viewed as a statement about the composition of two derived functors. Given an additional functor $\psi\colon \mathcal{B} \to \mathcal{C}$, one can more generally show that there is a natural equivalence of functors $\psi^h_* \phi^h_* \xrightarrow{\sim} (\psi\phi)^h_*$. In the lemma below, we shall consider the case where X has the form $\phi^* Y$ for a $\mathcal{B}$-diagram Y, and we shall relate π to the map of homotopy colimits induced by the natural transformation of $\mathcal{B}$-diagrams

$$t\colon \phi^h_* \phi^* Y \to \phi_* \phi^* Y \to Y,$$

where the first arrow is the canonical projection from the homotopy colimit to the colimit and the second arrow is given by the universal property of the categorical Kan extension.

Lemma A.3. *Given a $\mathcal{B}$-diagram Y, the diagram*

$$\begin{tikzcd} \operatorname*{hocolim}_{\mathcal{B}} \phi^h_* \phi^* Y \arrow[rr, "\pi"] \arrow[dr, "t"'] & & \operatorname*{hocolim}_{\mathcal{A}} \phi^* Y \arrow[dl, "\phi"] \\ & \operatorname*{hocolim}_{\mathcal{B}} Y & \end{tikzcd}$$

is homotopy commutative by a canonical choice of a natural homotopy.

Proof. We may view the homotopy colimit of the $\mathcal{B}$-diagram $\phi^h_* \phi^* Y$ as the realization of the bisimplicial space

$$([i],[j]) \mapsto \coprod_{\left\{\substack{b_0 \leftarrow \dots \leftarrow b_i \leftarrow \phi(a_0) \\ a_0 \leftarrow \dots \leftarrow a_j}\right\}} \phi^* Y(a_j),$$

and it is well-known that this is homeomorphic to the realization of the diagonal simplicial space. Restricting to this simplicial space, the two maps in the diagram are induced by the simplicial maps that map a simplex

$$(b_0 \leftarrow \dots \leftarrow b_i \xleftarrow{\gamma} \phi(a_0),\ a_0 \xleftarrow{\alpha_1} \dots \xleftarrow{\alpha_i} a_i,\ y)$$

with y in $\phi^* Y(a_i)$, to

$$(b_0 \leftarrow \dots \leftarrow b_i,\ \gamma_* \phi(\alpha_1 \dots \alpha_i)_* y),$$

$$\text{respectively} \quad (\phi(a_0) \xleftarrow{\phi(\alpha_0)} \dots \xleftarrow{\phi(\alpha_i)} \phi(a_i),\ y).$$

The required homotopy between the topological realizations of these maps is then defined by

$$[(b_0 \leftarrow \dots \leftarrow b_i \leftarrow \phi(a_0) \leftarrow \dots \leftarrow \phi(a_i),\ y); (su, (1-s)u)],$$

for $u \in \Delta^i$ and $s \in I$. Here I denotes the unit interval and

$$\Delta^i = \{(u_0, \dots, u_i) \in I^{i+1}\colon\ u_0 + \dots + u_i = 1\}$$

is the standard i-simplex. $\square$

The following lemma is needed to ensure that the functor R defined in Section 4.2 takes values in the subcategory of level-wise T-good objects in $\mathcal{IU}/BF$.

Lemma A.4. *Let $\mathcal{A}$ be a small category and let $f_a\colon X_a \to BF(n)$ be an $\mathcal{A}$-diagram in $\mathcal{U}/BF(n)$. If each f_a classifies a well-based quasifibration, then the induced map*

$$f\colon \operatorname*{hocolim}_{\mathcal{A}} X_a \to BF(n)$$

also classifies a well-based quasifibration.

Proof. Let $W_a = f_a^* V(n)$, and notice that $f^*V(n)$ is homeomorphic to $\operatorname{hocolim}_{\mathcal{A}} W_a$ since topological realization preserves pullback diagrams. It follows that the pullback of $V(n) \to BF(n)$ along f is homeomorphic to the realization of the simplicial map

$$\coprod_{a_0 \leftarrow \cdots \leftarrow a_k} W_{a_k} \to \coprod_{a_0 \leftarrow \cdots \leftarrow a_k} X_{a_k}.$$

These are good simplicial spaces in the sense of [36], Appendix A, and the map is a degree-wise quasifibration by assumption. The result then follows from standard results on realization of simplicial quasifibrations and simplicial cofibrations, see e.g. [36], Proposition 1.6 and [19]. $\square$

References

[1] R. D. Arthan and S. R. Bullett, *The homology of $MO(1)^{\wedge\infty}$ and $MU(1)^{\wedge\infty}$*, J. Pure Appl. Algebra 26 (1982), no. 3, 229–234.

[2] M. G. Barratt and P. J. Eccles, *Γ^+-structures-I: A free group functor for stable homotopy theory*, Topology 13 (1974), 25–45.

[3] C. Berger, *Combinatorial models for real configuration spaces and E_n-operads*, Contemp. Math., vol. 202, 37–52, Amer. Math. Soc., Providence, RI, 1997.

[4] A. Blumberg, R. Cohen, and C. Schlichtkrull, *Topological Hochschild homology of Thom spectra and the free loop space*, Preprint, 2008.

[5] J. S. Birman, *Braids, links, and mapping class groups*, Annals of Mathematics Studies, No 82, Princeton University Press, Princeton, N.J., 1975.

[6] F. Borceux, *Handbook of categorical algebra 2: Categories and structures*, Encyclopedia of Mathematics and its Applications 51, Cambridge University Press, Cambridge, 1994.

[7] A. K. Bousfield and D. M. Kan, *Homotopy limits, completions and localizations*, Springer Lecture Notes in Math. 304, Springer Verlag, 1972.

[8] M. Bökstedt, *Topological Hochschild homology*, Preprint 1985, Bielefeld University.

[9] S. R. Bullett, *Permutations and braids in cobordism theory*, Proc. London Math. Soc. (3) 38 (1979), no. 3, 517–531.

[10] F. R. Cohen, *Braid orientations and bundles with flat connection*, Inventiones math. 46, (1978), 99–110.

[11] F. R. Cohen, T. Lada, and J. P. May, *Homology of iterated loop spaces*, Springer Lecture Notes in Math. 533, Springer Verlag, 1976.

[12] F. R. Cohen and V. V. Vershinin, *Thom spectra which are wedges of Eilenberg-Mac Lane spectra*, Stable and unstable homotopy (Toronto, ON, 1996), 43–65, Fields Inst. Commun., 19, Amer. Math. Soc., Providence, RI, 1998.

[13] W. G. Dwyer and J. Spalinski, *Homotopy theories and model categories*, Handbook of algebraic topology, North-Holland, Amsterdam, 1995, 73–126.

[14] T. Goodwillie, *Calculus II: Analytic functors*, K-theory 5, (1992), no. 4, 295–332.

[15] P. S. Hirschhorn, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, American Mathematical Society, Providence, R.I, 2003.

[16] M. Hovey, B. Shipley and J. Smith, *Symmetric spectra*, J. Amer. Math. Soc. 13, (2000), no. 1, 149–208.

[17] J. Hollender and R. M. Vogt, *Modules of topological spaces, applications to homotopy limits and E_∞ structures*, Arch. Math. 59 (1992), no. 2, pp. 115–129.

[18] I. M. James, *Reduced product spaces*, Ann. of Math. (2) 62, (1955), 170–197.

[19] L. G. Lewis, *When is the natural map $X \to \Omega\Sigma X$ a cofibration?*, Trans. Amer. Math. Soc. 273 (1982), no. 1, 147–155.

[20] L. G. Lewis, J. P. May and M. Steinberger, *Equivariant stable homotopy theory*, Springer Lecture Notes in Math. 1213, Springer Verlag, 1986.

[21] J. Lillig, *A union theorem for cofibrations*, Arch. Math 24 (1973), 410–415.

[22] S. Mac Lane, *Categories for the working mathematician.* Second Edition, Graduate Texts in Mathematics, Springer Verlag, New York, 1998. Springer-Verlag, 1971.

[23] I. Madsen, *Algebraic K-theory and traces*, from: "Current Developments in Mathematics, 1995, (Cambridge, MA), International Press, 191–321 (1996).

[24] M. Mahowald, *A new infinite family in ${}_2\pi_*^s$*, Topology 16 (1977), 249–256.

[25] M. Mahowald, *Ring spectra which are Thom spectra*, Duke Math. J. 46 (1979), 549–559.

[26] M. Mahowald and N. Ray, *A note on the Thom isomorphism*, Proc. Amer. Math. Soc. 82 (1981), no. 2, 307–308.

[27] M. A. Mandell, J. P. May, S. Schwede and B. Shipley, *Model categories of diagram spectra*, Proc. London Math. Soc. (3), 82 (2001), no. 2, 441–512.

[28] J. P. May, *The geometry of iterated loop spaces*, Springer Lecture Notes in Math. 271, Springer Verlag, 1972.

[29] J. P. May, *E_∞ spaces, group completions, and permutative categories*, New developments in topology (Proc. Sympos. Algebraic Topology, Oxford, 1972), pp. 61–93. London Math. Soc. Lecture Note Ser., No. 11, Cambridge Univ. Press, London, 1974.

[30] J. P. May, *Classifying spaces and fibrations*, Mem. Amer. Math. Soc. 1 (1975), 1, no. 155.
[31] J. P. May, *E_∞ ring spaces and E_∞ ring spectra*, Springer Lecture Notes in Math. 577, Springer Verlag, 1977.
[32] J. P. May and J. Sigurdsson, *Parametrized homotopy theory*, Mathematical Surveys and Monographs, 132, American Mathematical Society, Providence, RI, 2006.
[33] S. Sagave and C. Schlichtkrull, *Diagram spaces and symmetric spectra*, in preparation.
[34] C. Schlichtkrull, *The homotopy infinite symmetric product represents stable homotopy*, Algebr. Geom. Topol. 7 (2007), 1963–1977.
[35] S. Schwede, *Symmetric spectra*, in preparation, online version accessible at `http://www.math.uni-bonn.de/people/schwede/SymSpec.pdf`.
[36] G. Segal, *Categories and cohomology theories*, Topology 13 (1974), 293–312.
[37] B. Shipley, *Symmetric spectra and topological Hochschild homology*, K-theory 19, (2000), no. 2, 155–183.
[38] J. H. Smith, *Simplicial group models for $\Omega^n S^n X$*, Israel J. Math. 66 (1989), no. 1–3, 330–350.
[39] N. Steenrod, *A convenient category of topological spaces*, Michigan Math. J. 14 (1967), 133–152.
[40] A. Strøm, *Note on cofibrations II*, Math. Scand. 22 (1968), 130–142.
[41] A. Strøm, *The homotopy category is a homotopy category*, Arch. Math. 23 (1972), 435–441.
[42] F. Waldhausen, *Algebraic K-theory of spaces*, Lecture Notes in Math. 1126, 318–419, Springer-Verlag, New York, 1985.

Christian Schlichtkrull
Department of Mathematics
University of Bergen
Johannes Brunsgate 12
5008 Bergen
Norway
krull@math.uib.no

Twisted Cohomology of the Hilbert Schemes of Points on Surfaces

Marc A. Nieper-Wisskirchen[1]

Received: February 1, 2008
Revised: April 23, 2009

Communicated by Thomas Peternell

Abstract. We calculate the cohomology spaces of the Hilbert schemes of points on surfaces with values in local systems. For that purpose, we generalise I. Grojnoswki's and H. Nakajima's description of the ordinary cohomology in terms of a Fock space representation to the twisted case. We make further non-trivial generalisations of M. Lehn's work on the action of the Virasoro algebra to the twisted and the non-projective case.

Building on work by M. Lehn and Ch. Sorger, we then give an explicit description of the cup-product in the twisted case whenever the surface has a numerically trivial canonical divisor.

We formulate our results in a way that they apply to the projective and non-projective case in equal measure.

As an application of our methods, we give explicit models for the cohomology rings of the generalised Kummer varieties and of a series of certain even dimensional Calabi–Yau manifolds.

2000 Mathematics Subject Classification: 14C05; 14C15
Keywords and Phrases: Hilbert schemes of points on surfaces, rational cohomology ring, locally constant systems, generalised Kummer varieties

[1]The author would like to thank the Max Planck Institute for Mathematics in Bonn for its hospitality and support during the preparation of this paper. He would also like to thank the referee for some very helpful suggestions.

Contents

1. Introduction and results

Let X be a quasi-projective smooth surface over the complex numbers. We denote by $X^{[n]}$ the Hilbert scheme of n points on X, parametrising zero-dimensional subschemes of X of length n. It is a quasi-projective variety ([Gro61]) and smooth of dimension $2n$ ([Fog68]). Recall that the Hilbert scheme $X^{[n]}$ can be viewed as a resolution of the n-th symmetric power $X^{(n)} := X^n/\mathfrak{S}_n$ of the surface X by virtue of the Hilbert–Chow morphism $\rho\colon X^{[n]} \to X^{(n)}$, which maps each zero-dimensional subscheme ξ of X to its support $\operatorname{supp}\xi$ counted with multiplicities.

Let L be a local system (always over the complex numbers and of rank 1) over X. We can view it as a functor from the fundamental groupoid Π of X to the category of one-dimensional complex vector spaces.

The fundamental groupoid $\Pi^{(n)}$ of $X^{(n)}$ is the quotient groupoid of Π^n by the natural $\mathfrak{S}_n$-action by [Bro88]. (Recall from [Bro88] that the quotient groupoid of a groupoid P on which a group G is acting (by functors) is a groupoid P/G together with a functor $p\colon P \to P/G$ that is invariant under the G-action and so that $p\colon P \to P/G$ is universal with respect to this property.)

Readers who prefer to think in terms of the fundamental group (as opposed to the fundamental groupoid) can find a description of the fundamental group of $L^{(n)}$ in [Bea83].

By the universal property of $\Pi^{(n)}$, we can thus construct from L a local system $L^{(n)}$ on $X^{(n)}$ by setting

$$L^{(n)}(x_1, \dots, x_n) := \bigotimes_i L(x_i),$$

for each $(x_1, \dots, x_n) \in X^{(n)}$ (for the notion of the tensor product over an unordered index set see, e.g., [LS03]). This induces the locally free system $L^{[n]} := \rho^* L^{(n)}$ on $X^{[n]}$.

We are interested in the calculation of the direct sum of cohomology spaces $\bigoplus_{n\geq 0} H^*(X^{[n]}, L^{[n]}[2n])$. Besided the natural grading given by the cohomological degree it carries weighting (see remark 1.1 below) given by the number of points n. Likewise, the symmetric algebra $S^*(\bigoplus_{\nu\geq 1} H^*(X, L^\nu[2]))$ carries a grading by cohomological degree and a weighting, which is defined so that $H^*(X, L[2]^\nu)$ is of pure weight ν.

Remark 1.1. Here, a *weighting* is just another name for a second grading. A *weight space* is a homogeneous subspace to a given degree with respect to this second grading. Being of pure weight means being homogeneous with respect to the second grading.
In the context of super vector spaces, however, me make a difference between a grading and a weighting: Write $V = V^0 \oplus V^1$ for the decomposition of a super vector space into its even and odd part. Recall that for a grading $V = \bigoplus_{n\in\mathbf{Z}} V_n$ on V we have $V^i = \bigoplus_n V_n^{i+n \pmod 2}$.
For a weighting, on the contrary, we want to adopt the following convention: If $V = \bigoplus_{n\in\mathbf{Z}} V(n)$ is the decomposition of a weighted super vector space into its weight spaces, one has $V^i = \bigoplus_n V(n)^i$, i.e. the weighting does not interfere with the $\mathbf{Z}/(2)$-grading.
This difference is important, for example, for the notion of (super-)commutativity.

The first result of this paper is the following:

Theorem 1.2. *There is a natural vector space isomorphism*

$$\bigoplus_{n\geq 0} H^*(X^{[n]}, L^{[n]}[2n]) \to S^*\left(\bigoplus_{\nu\geq 1} H^*(X, L^\nu[2])\right)$$

that respects the grading and weighting.

For $L = \mathbf{C}$, the trivial system, this result has already appeared in [Gro96] and [Nak97]).
Theorem 1.2 is proven by defining a Heisenberg Lie algebra $\mathfrak{h}_{X,L}$, whose underlying vector space is given by

$$\bigoplus_{n\geq 0} H^*(X, L^n[2]) \oplus \bigoplus_{n\geq 0} H^*_c(X, L^{-n}[2]) \oplus \mathbf{C}\boldsymbol{c} \oplus \mathbf{C}\boldsymbol{d}$$

and by showing that $\bigoplus_{n\geq 0} H^*(X^{[n]}, L^{[n]}[2n])$ is an irreducible lowest weight representation of this Lie algebra, as is done in [Nak97] for the untwisted case.
Let $p\colon \hat{X} \to X$ be a finite abelian Galois covering over the surface X with Galois group G. The direct image $M := p_*\mathbf{C}$ of the trivial local system on $\hat{X}$ is a local system on X of rank $|G|$, the order of G. Note that G acts naturally on M. As G is abelian, there is a decomposition $M \cong \bigoplus_{\chi\in G^\vee} L_\chi$, where $G^\vee = \operatorname{Hom}(G, \mathbf{C}^\times)$ is the character group of G and L_χ is the subsystem of M on which G acts via χ. In fact, each L_χ is a local system of rank one.
Consider $M^{[n]} := \bigoplus_{\chi\in G^\vee} L_\chi^{[n]}$. This is a local system of rank $|G|$ on $X^{[n]}$. Let $q\colon \widehat{X^{[n]}} \to X^{[n]}$ be a finite abelian Galois covering of $X^{[n]}$ such that $q_*\mathbf{C} = M^{[n]}$. Using the Leray spectral sequence for q, which already degenerates at the E_2-term, the cohomology of $\widehat{X^{[n]}}$ can be computed by Theorem 1.2:

Corollary 1.3. *There is a natural vector space isomorphism*

$$\bigoplus_{n\geq 0} H^*(\widehat{X^{[n]}}, \mathbf{C}[2n]) \to \bigoplus_{\chi\in G^\vee} S^*\left(\bigoplus_{\nu\geq 1} H^*(X, L_\chi^\nu[2])\right)$$

that respects the grading and weighting.

We then proceed in the paper by defining a twisted version $\mathfrak{v}_{X,L}$ of the Virasoro Lie algebra, whose underlying vector space will be given by

$$\bigoplus_{n\geq 0} H^*(X, L^n) \oplus \bigoplus_{n\geq 0} H_c^*(X, L^{-n}) \oplus \mathbf{C}\boldsymbol{c} \oplus \mathbf{C}\boldsymbol{d}.$$

(Note the different grading compared to $\mathfrak{h}_{X,L}$.) We define an action of $\mathfrak{v}_{X,L}$ on $\bigoplus_{n\geq 0} H^*(X^{[n]}, L^{[n]}[2n])$ by generalising results of [Leh99] to the twisted, not necessarily projective case. As in [Leh99], we calculate the commutators of the operators in $\mathfrak{h}_{X,L}$ with the boundary operator ∂ that is given by multiplying with $-\frac{1}{2}$ of the exceptional divisor class of the Hilbert–Chow morphism. It turns out that the same relations as in the untwisted, projective case hold.
The next main result of the paper is a decription of the ring structure whenever X has a numerically trivial canonical divisor. Following ideas in [LS03], we introduce a family of explicitely described graded unital algebras $H^{[n]}$ associated to a G-weighted (non-counital) graded Frobenius algebra H of degree d. For example, $H = \bigoplus_{L\in G^\vee} H^*(X, L_\chi[2])$ is such a Frobenius algebra of degree 2 where G is as above. The following holds for each $n \geq 0$:

Theorem 1.4. *Assume that X has a numerically trivial canonical divisor. Then there is a natural isomorphism*

$$\bigoplus_{\chi\in G^\vee} H^*(X^{[n]}, L_\chi^{[n]}[2n]) \to \left(\bigoplus_{\chi\in G^\vee} H^*(X, L_\chi[2])\right)^{[n]}$$

of (G-weighted) graded algebras of degree $2n$.

For G the trivial group, and X projective, this theorem is the main result in [LS03].
The idea of the proof of Theorem 1.4 is not to reinvent the wheel but to study how everything can already be deduced from the more special case considered in [LS03].
Again by the Leray spectral sequence, Theorem 1.4 also has a natural application to the cohomology ring of coverings of $X^{[n]}$:

Corollary 1.5. *There is a natural isomorphism*

$$H^*(\widehat{X^{[n]}}, \mathbf{C}[2n]) \to \left(\bigoplus_{\chi\in G^\vee} H^*(X, L_\chi[2])\right)^{[n]}$$

of graded unital algebras of degree $2n$.

We want to point out at least two applications of our results. The first one is the computation of the cohomology ring of certain families of Calabi–Yau manifolds of even dimension: Let X be an Enriques surface. Let $\hat{X}$ be its universal covering, which is a K3 surface. Then $G \simeq \mathbf{Z}/(2)$. We denote the local system corresponding to the non-trivial element in G by L. The Hodge diamonds of $H^*(X, \mathbf{C}[2])$ and $H^*(X, L[2])$ are given by

$$\begin{array}{ccccc} & & 1 & & \\ & 0 & & 0 & \\ 0 & & 10 & & 0 \\ & 0 & & 0 & \\ & & 1 & & \end{array} \quad \text{and} \quad \begin{array}{ccccc} & & 0 & & \\ & 0 & & 0 & \\ 1 & & 10 & & 1, \\ & 0 & & 0 & \\ & & 0 & & \end{array}$$

respectively.

Denote by $X^{\{n\}}$ the (two-fold) universal cover of $X^{[n]}$. By Remark 2.7, the isomorphism of Corollary 1.3 is in fact an isomorphism of Hodge structures. It follows that

$$H^{k,0}(X^{\{n\}}, \mathbf{C}) = \begin{cases} \mathbf{C} & \text{for } k = 0 \text{ or } k = 2n, \text{ and} \\ 0 & \text{for } 0 < k < 2n. \end{cases}$$

In conjunction with Corollary 1.5, we have thus proven:

Proposition 1.6. *For $n > 1$, the manifold $X^{\{n\}}$ is a Calabi–Yau manifold in the strict sense. Its cohomology ring $H^*(X^{\{n\}}, \mathbf{C}[2n])$ is naturally isomorphic to $(H^*(X, \mathbf{C}[2]) \oplus H^*(X, L[2]))^{[n]}$.*

Our second application is the calculation of the cohomology ring of the generalised Kummer varieties $X^{[[n]]}$ for an abelian surface X. (A slightly less explicit description of this ring has been obtained by more special methods in [Bri02].)

Recall from [Bea83] that the generalised Kummer variety $X^{[[n]]}$ is defined as the fibre over 0 of the morphism $\sigma\colon X^{[n]} \to X$, which is the Hilbert–Chow morphism followed by the summation morphism $X^{(n)} \to X$ of the abelian surface. The generalised Kummer surface is smooth and of dimension $2n-2$ ([Bea83]).

As above, let H be a G-weighted graded Frobenius algebra of degree d. Assume further that H is equipped with a compatible structure of a Hopf algebra of degree d. For each $n > 0$, we associate to such an algebra an explicitely described graded unital algebra $H^{[[n]]}$ of degree n.

In the following Theorem, we view $H^*(X, \mathbf{C}[2])$ as such an algebra (the Hopf algebra structure is given by the group structure of X), where we give $H^*(X, \mathbf{C}[2])$ the trivial G-weighting for the group $G := X[n]$, the character group of the group of n-torsion points on X. We prove the following:

Theorem 1.7. *There is a natural isomorphism*

$$H^*(X^{[[n]]}, \mathbf{C}[2n]) \to (H^*(X, \mathbf{C}[2]))^{[[n]]}$$

of graded unital algebras of degree $2n$.

We should remark that most of the "hard work" that is hidden behind the scenes is the work of [Gro96], [Nak97], [Leh99], [LQW02], [LS03], etc. Our own contribution is to generalise and apply the ideas and results in the cited papers to the twisted and to the non-projective case.

Remark 1.8. Let us finally mention that the restriction to algebraic, i.e. quasi-projected surfaces, is just a matter of convenience. Our methods work equally well when we replace X by any complex surface. In this case, the Hilbert schemes become the Douady spaces ([Dou66]).

2. The Fock space description

In this section, we prove Theorem 1.2 for a local system L on X by the method that is used in [Nak97] for the untwisted case, i.e. by realising the cohomology space of the Hilbert schemes (with coefficients in a local system) as an irreducible representation of a Heisenberg Lie algebra.
Let $l \geq 0$ and $n \geq 1$ be two natural numbers. Set

$$X^{(l,n)} := \left\{ (\underline{x}', x, \underline{x}) \in X^{(n+l)} \times X \times X^{(l)} \mid \underline{x}' = \underline{x} + nx \right\}$$

(we write the union of unordered tuples additively). We further define the reduced subvariety

$$X^{[n,l]} := \left\{ (\xi', x, \xi) \in X^{[n+l]} \times X \times X^{[l]} \mid \xi \subset \xi', (\rho(\xi'), x, \rho(\xi)) \in X^{(l,n)} \right\}$$

in $X^{[n+l]} \times X \times X^{[l]}$. This incidence variety has already been considered in [Nak97]. In contrast to the Hilbert schemes, these incidence varieties are almost never smooth. Its image under the Hilbert–Chow morphism is again $X^{(l,n)}$.
We denote the projections of $X^{(l+n)} \times X \times X^{(l)}$ onto its three factors by $\tilde{p}$, $\tilde{q}$ and $\tilde{r}$, respectively. Likewise, we denote the three projections of $X^{[l+n]} \times X \times X^{[l]}$ by p, q and r.

LEMMA 2.1. *We have a natural isomorphism* $q^* L^n \otimes r^* L^{[l]}|_{X^{[n,l]}} \cong p^* L^{[l+n]}|_{X^{[n,l]}}$.

Proof. Firstly, we have a natural isomorphism $\tilde{q}^* L^n \otimes \tilde{r}^* L^{(l)}|_{X^{(n,l)}} \cong \tilde{p}^* L^{(l+n)}|_{X^{(n,l)}}$. This follows from

$$\begin{aligned}(\tilde{q}^* L^n \otimes \tilde{r}^* L^{(l)})(\underline{x} + nx, x, \underline{x}) \\ = L(x)^{\otimes n} \otimes \bigotimes_{x' \in \underline{x}} L(x') = \bigotimes_{x' \in \underline{x} + nx} L(x') = \tilde{p}^* L^{(l+n)}(\underline{x} + nx, x, \underline{x})\end{aligned}$$

for every $(\underline{x} + nx, x, \underline{x}) \in X^{(l,n)}$. By pulling back everything to the Hilbert schemes, the Lemma follows. $\square$

Due to Lemma 2.1 and the fact that $p|_{X^{[l,n]}}$ is proper ([Nak97]), the operator (a correspondence, see [Nak97])

$$N\colon H^*(X, L^n[2]) \times H^*(X^{[l]}, L^{[l]}[2l]) \to H^*(X^{[l+n]}, L^{[l+n]}[2(l+n)]),$$
$$(\alpha, \beta) \mapsto \mathrm{PD}^{-1} p_*((q^*\alpha \cup r^*\beta) \cap [X^{[l,n]}])$$

is well-defined. Here,

$$\mathrm{PD}\colon H^*(X^{[l+n]}, L^{[n+l]}[2(l+n)]) \to H_*^{\mathrm{BM}}(X^{[l+n]}, L^{[n+l]}[-2(l+n)])$$

is the Poincar -duality isomorphism between the cohomology and the Borel–Moore homology. (The degree shifts are chosen in a way that N is an operator of degree 0, see [LS03].)

Remark 2.2. Note that although the variety $X^{[l,n]}$ is not smooth in general, it nevertheless possesses a fundamental class $[X^{[l,n]}] \in H_*^{\mathrm{BM}}(X^{[l,n]}, \mathbf{C})$. (This is actually true for every analytic variety, see e.g. the appendices of [PS08].)

Furthermore, $q \times r|_{X \times X^{[l]}}$ is proper ([Nak97]). Thus we can also define an operator the other way round:

$$N^\dagger\colon H_c^*(X, L^{-n}[2]) \times H^*(X^{[n+l]}, L^{[l+n]}[2]) \to H^*(X^{[l]}, L^{[l]}[2l]),$$
$$(\alpha, \beta) \mapsto (-1)^n \mathrm{PD}^{-1} r_*(q^*\alpha \cup p^*\beta \cap [X^{[l,n]}])$$

As in [Nak97], we will use these operators to define an action of a Heisenberg Lie algebra on

$$V_{X,L} := \bigoplus_{n \geq 0} H^*(X^{[n]}, L^{[n]}[2n]).$$

For this, let A be a weighted, graded Frobenius algebra of degree d (over the complex numbers), that is a weighted and graded vector space over $\mathbf{C}$ with a (graded) commutative and associative multiplication of degree d and weight 0 and a unit element 1 (necessarily of degree $-d$ and weight 0) together with a linear form $\int\colon A \to \mathbf{C}$ of degree $-d$ and weight 0 such that for each weight $\nu \in \mathbf{Z}$ the induced bilinear form $\langle\cdot,\cdot\rangle\colon A(\nu) \times A(-\nu) \to \mathbf{C}, (a, a') \mapsto \int_A aa'$ is non-degenerate (of degree 0). Here $A(\nu)$ denotes the weight space of weight ν. In particular, all weight spaces are finite-dimensional. In the case of a trivial weighting, this notion of a graded Frobenius algebra has already appeared in [LS03].

Example 2.3. Consider the vector space

$$A_{X,L} := \bigoplus_{\nu \geq 0} H^*(X, L^\nu[2]) \oplus \bigoplus_{\nu \geq 0} H_c^*(X, L^{-\nu}[2]).$$

It inherits a grading from the cohomological grading of its pieces $H^*(X, L^\nu[2])$. We endow $A_{X,L}$ also with a weighting by defining $H^*(X, L^\nu[2])$ to be of pure weight ν for $\nu \geq 0$ and $H_c^*(X, L^{-\nu}[2])$ to be of pure weight $-\nu$.
Recall that there is a natural linear map $\phi\colon H_c^*(X, M) \to H^*(X, M)$ for every local system M on X. (In the de Rham-model of cohomology, it is induced by the inclusion of the (co-)complex of forms with compact support into the

complex of forms with arbitrary support.) This linear map is compatible with the — non-unitary in the case of compact support — algebra structures on $H^*_c(X,\cdot)$ and $H^*(X,\cdot)$ and the module structure of $H^*_c(X,\cdot)$ over $H^*(X,\cdot)$. With this, we mean that

$$\phi(am) = a\phi(m), \qquad \phi(mn) = \phi(m)\phi(n), \qquad \phi(m)n = mn \tag{1}$$

for all $a \in H^*(X, M)$ and $m, n \in H^*_c(X, M)$.
This allows us to define a commutative multiplication map of degree 2 ($= \dim X$) and weight 0 on $A_{X,L}$ as follows: For elements $\alpha, \beta \in A_{X,L}$ of pure weight, we set

$$\alpha \cdot \beta := \begin{cases} \alpha \cup \beta & \text{for } \alpha \in H^*(X, L^\nu[2]), \beta \in H^*(X, L^\mu[2]) \\ \alpha \cup \beta & \text{for } \alpha \in H^*_c(X, L^{-\nu}[2]), \beta \in H^*_c(X, L^{-\mu}[2]) \\ \alpha \cup \beta & \text{for } \alpha \in H^*(X, L^\nu[2]), \beta \in H^*_c(X, L^{-\mu}[2]) \text{ and } \nu \leq \mu \\ \phi(\alpha \cup \beta) & \alpha \in H^*(X, L^\nu[2]), \beta \in H^*_c(X, L^{-\mu})[2] \text{ and } \nu > \mu \end{cases}$$

for $\nu, \mu \geq 0$. By (1), it follows immediately that this multiplication map is associative, i.e. defines on $A_{X,L}$ the structure of a weighted, graded, unital, commutative and associative algebra of degree 2.
We proceed by extending the linear form $\int_X \colon H^*_c(X, \mathbf{C}[2]) \to \mathbf{C}$ of degree -2 given by evaluating a class of compact support on the fundamental class of X trivially (that is by extending by zero) on $A_{X,L}$ and call the resulting linear form $\int \colon A_{X,L} \to \mathbf{C}$.
We claim that this endows $A_{X,L}$ with the structure of a weighted, graded Frobenius algebra of degree 2: In fact, given a class $\alpha \in H^*(X, L^\nu)$, we can always find a class $\beta \in H^*_c(X, L^{-\nu})$ and vice versa so that $\int \alpha \cdot \beta = \int_X \alpha \cup \beta \neq 0$.

For any weighted, graded Frobenius algebra A we set

$$\mathfrak{h}_A := A \oplus \mathbf{C}\boldsymbol{c} \oplus \mathbf{C}\boldsymbol{d}.$$

We define the structure of a weighted, graded Lie algebra on $\mathfrak{h}_A$ by defining $\boldsymbol{c}$ to be a central element of weight 0 and degree 0, $\boldsymbol{d}$ an element of weight 0 and degree 0 and by setting the following commutator relations: $[\boldsymbol{d}, a] := n \cdot a$ for each element $a \in A$ of weight n, and $[a, a'] = \langle [\boldsymbol{d}, a], a' \rangle \boldsymbol{c}$ for elements $a, a' \in A$.

Definition 2.4. The Lie algebra $\mathfrak{h}_A$ the *Heisenberg algebra associated to A*.

For $A = A_{X,L}$, we set $\mathfrak{h}_{X,L} := \mathfrak{h}_A$. We define a linear map

$$q \colon \mathfrak{h}_{X,L} \to \operatorname{End}(V_{X,L})$$

as follows: Let $l \geq 0$ and $\beta \in V_{X,L}(l) = H^*(X^{[l]}, L^{[l]}[2l])$. We set $q(\boldsymbol{c})(\beta) := \beta$, and $q(\boldsymbol{d})(\beta) := l\beta$. For $n \geq 0$, and $\alpha \in A_{X,L}(\nu) = H^*(X, L^\nu[2])$, we set $q(\alpha)(\beta) := N(\alpha, \beta)$. For $\alpha \in A_{X,L}(-\nu) = H^*_c(X, L^{-\nu}[2])$, we set $q(\alpha)(\beta) := N^\dagger(\alpha, \beta)$. Finally, we set $q(\alpha)(\beta) = 0$ for $\alpha \in A_{X,L}(0) = H^*(X, \mathbf{C}) \oplus H^*_c(X, \mathbf{C})$.

Proposition 2.5. *The map q is a weighted, graded action of $\mathfrak{h}_{X,L}$ on $V_{X,L}$.*

Proof. This Proposition is proven in [Nak97] for the untwisted case, i.e. for $L = \mathbf{C}$. The proof there is based on calculating commutators on the level of cycles of the correspondences defined by the incidence schemes $X^{[l,n]}$. These commutators are independent of the local system used. Thus the proof in [Nak97] also applies to this more general case. $\square$

Example 2.6. Let $\alpha = \sum \alpha_{(1)} \otimes \cdots \otimes \alpha_{(n)} \in H^*(X^{(n)}, L^{(n)}[2n]) = S^n H^*(X, L[2]) = (H^*(X, L[2])^{\otimes n})^{\mathfrak{S}_n}$ (we use the Sweedler notation to denote elements in tensor products). The pull-back of α by the Hilbert–Chow morphism $\rho\colon X^{[n]} \to X^{(n)}$ is then given by

$$\rho^*\alpha = \frac{1}{n!}\sum q(\alpha_{(1)})\cdots q(\alpha_{(n)})|0\rangle,$$

where $|0\rangle$ is the unit $1 \in H^*(X^{[0]}, \mathbf{C}) = \mathbf{C}$.

We will use Proposition 2.5 to prove our first Theorem.

Proof of Theorem 1.2. The vector space $\tilde{V}_{X,L} := S^*(\bigoplus_{\nu\geq 1} H^*(X, L^\nu[2]))$ carries a unique structure of an $\mathfrak{h}_{X,L}$-module such that $\boldsymbol{c}$ acts as the identity, $\boldsymbol{d}$ acts by multiplying with the weight, $\alpha \in H^*(X, L^n)$ for $n \geq 1$ acts by multiplying with α, and $\alpha \in H^*(X, \mathbf{C}) \oplus H^*_c(X, \mathbf{C})$ acts by zero. By the representation theory of the Lie algebras of Heisenberg type, this is an irreducible lowest weight representation of $\mathfrak{h}_{V,L}$, which is generated by the lowest weight vector 1 of weight 0.
The $\mathfrak{h}_{V,L}$-module $V_{X,L}$ also has a vector of weight 0, namely $|0\rangle$. Thus, there is a unique morphism $\Phi\colon \tilde{V}_L \to V_L$ of $\mathbf{h}_L$-modules that maps 1 to $|0\rangle$. This will be the inverse of the isomorphism mentioned in Theorem 1.2. It remains to show that Φ is bijective. The injectivity follows from the fact that $\tilde{V}_{X,L}$ is irreducible as an $\mathfrak{h}_{X,L}$-module.
In order to prove the surjectivity, we will derive upper bounds on the dimensions of the weight spaces of the right hand side $V_{X,L}$ (see also [Leh04] about this proof method). By the Leray spectral sequence associated to the Hilbert–Chow morphism $\rho\colon X^{[n]} \to X^{(n)}$, such an upper bound is provided by the dimension of the spectral sequence's E_2-term $H^*(X^{(n)}, \mathbf{R}^*\rho_* L[2n])$. As shown in [GS93], it follows from the Beilinson–Bernstein–Deligne–Gabber decomposition theorem that

$$\mathbf{R}^*\rho_*\mathbf{Q}[2n] = \bigoplus_{\lambda\in\mathbf{P}(n)} (i_\lambda)_*\mathbf{Q}[2\ell(\lambda)].$$

Here, $P(n)$ is the set of all partitions of n, $\ell(\lambda) = r$ is the length of a partition $\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^r)$, $X^{(\lambda)} := \{\sum_{i=1}^r \lambda_i x_i \mid x_i \in X\} \subset X^{(n)}$, and $i_\lambda\colon X^{(\lambda)} \to X^{(n)}$ is the inclusion map.
Set $L^{(\lambda)} := i_\lambda^* L^{(n)}$. By the projection formula, it follows that $\mathbf{R}^*\rho_* L[2n] = \bigoplus_{\lambda\in\mathbf{P}(n)} (i_\lambda)_* L^{(\lambda)}[2\ell(\lambda)]$.
Thus, an upper bound on the dimension of $H^*(X^{[n]}, L^{[n]}[2n])$ is provided by the dimension of $\bigoplus_{\lambda\in\mathbf{P}(n)} H^*(X^{(\lambda)}, L^{(\lambda)}[2\ell(\lambda)])$. By [GS93], this can be seen

to be isomorphic to

$$\bigoplus_{\sum_{i\geq 1} i\nu_i = n} \bigotimes_{i\geq 1} S^{\nu_i} H^*(X, L^i[2]),$$

where each $\nu_i \geq 0$. It follows that the upper bound given by the E_2-term is exactly the dimension of the n-th weight space of $\tilde{V}_{X,L}$. Thus the dimension of the weight spaces of $V_{X,L}$ cannot be greater than the dimensions of the weight spaces of $\tilde{V}_{X,L}$. Thus the Theorem is proven. □

Remark 2.7. Assume that X is projective. In this case, the (twisted) cohomology spaces of X and its Hilbert schemes $X^{[n]}$ carry pure Hodge structures. As the isomorphism of Theorem 1.2 is defined by algebraic correspondences (i.e. by correspondences of Hodge type (p,p)), it follows that the isomorphism in Theorem 1.2 is compatible with the natural Hodge structures on both sides. In terms of Hodge numbers, the following equation encodes our result:

$$\sum_{n\geq 0} \prod_{i,j} h^{i,j}(X^{[n]}, L^{[n]}[2n]) p^i q^j z^n = \prod_{m\geq 1} \prod_{i,j} (1 - (-1)^{i+j} p^i q^j z^m)^{-(-1)^{i+j} h^{i,j}(X, L^m[2])}$$

3. The Virasoro algebra in the twisted case

To each weighted, graded Frobenius algebra A of degree d, we associate a skew-symmetric form $e\colon A \times A \to \mathbf{C}$ of degree d as follows:
Let $n \in \mathbf{Z}$. We note that $A(n)$ and $A(-n)$ are dual to each other via the linear form $\int$. Thus we can consider the linear map $\Delta(n)\colon \mathbf{C} \to A(n) \otimes A(-n)$ dual to the bilinear form $\langle\cdot,\cdot\rangle\colon A(n) \otimes A(-n) \to \mathbf{C}$. Write $\Delta(n)1 = \sum e_{(1)}(n) \otimes e_{(2)}(n)$ in Sweedler notation. Then we define e by setting

$$e(\alpha, \beta) := \sum_{\nu=0}^{n} \frac{\nu(n-\nu)}{2} \int \sum e_{(1)}(\nu) e_{(2)}(\nu) \alpha\beta$$

for all $\alpha \in A(n)$ whenever $n \geq 0$. We shall call this form the *Euler form of* A.

Example 3.1. Assume that $A(n) \equiv A(0)$ for all $n \in \mathbf{Z}$. In this case, we have

$$e(\alpha, \beta) = \frac{n^3 - n}{12} \int e\alpha\beta$$

for $\alpha \in A(n)$ with $e := \int \sum e_{(1)}(0) e_{(2)}(0)$ ([Leh99]).

We use the Euler form to define another Lie algebra associated to A. We set

$$\mathfrak{v}_A := A[-2] \oplus \mathbf{C}\boldsymbol{c} \oplus \mathbf{C}\boldsymbol{d}.$$

We define the structure of a weighted, graded Lie algebra on $\mathfrak{v}_A$ be defining $\boldsymbol{c}$ to be a central element or weight 0 and degree 0, $\boldsymbol{d}$ an alement of weight 0 and degree 0 and by introducing the following commutator relations: $[\boldsymbol{d}, a] := n \cdot a$ for each element $a \in A[-2]$ of weight n, and $[a, a'] := (\boldsymbol{d}a)a' - a(\boldsymbol{d}a') - e(a, a')$ for elements $a, a' \in A$.

Definition 3.2. The Lie algebra $\mathfrak{v}_A$ is the *Virasoro algebra associated to A*.

For $A = A_{X,L}$, we set $\mathfrak{v}_{X,L} := \mathfrak{v}_A$. The whole construction is a generalisation to the twisted case of the Virasoro algebra found in [Leh99].
We now define a linear map $L\colon \mathfrak{v}_{X,L} \to \mathrm{End}(V_{X,L})$ as follows: We define $L(\boldsymbol{c})$ to be the identity, $L(\boldsymbol{d})$ to be multiplication with the weight, and for $\alpha \in A[-2]$ we set

$$L(\alpha) := \frac{1}{2}\sum\sum_{\nu\in\mathbf{Z}} {:}q(e_{(1)}(\nu))q(e_{(2)}(\nu)\alpha){:},$$

where the *normal ordered product* $:aa':$ of two operators is defined to be aa' if the weight of a is greater or equal to the weight of a' and is defined to be $a'a$ if the weight of a' is greater than the weight of a.
The following Lemma is proven for the untwisted case in [Leh99].

Lemma 3.3. *For $\alpha \in A_{V,L}[-2]$ and $\beta \in A_{V,L}$, we have*

$$[L(\alpha), q(\beta)] = -q(\alpha[\boldsymbol{d}, \beta]).$$

Proof. Let $\alpha \in A_{V,L}[2](n)$ and $\beta \in A_{V,L}(m)$ with $n, m \in \mathbf{Z}$. In the following calculations we omit all Koszul signs arising from commuting the graded elements α and β. By definition, we have $[L(\alpha), q(\beta)] = \frac{1}{2}\sum\sum_\nu [{:}q(e_{(1)}(\nu))q(e_{(2)}(\nu)\alpha){:}, q(\beta)]$, where ν runs through all integers. As the commutator of two operators in $\mathfrak{h}_{V,L}$ is central, we do not have to pay attention to the order of the factors of the normally ordered product when calculating the commutator:

$$\begin{aligned}&[{:}q(e_{(1)}(\nu))q(e_{(2)}(\nu)\alpha){:}, q(\beta)]\\ &\qquad = \nu\langle e_{(1)}(\nu), \beta\rangle q(e_{(2)}(\nu)\alpha) + (n-\nu)\langle e_{(2)}(\nu)\alpha, \beta\rangle q(e_{(1)}(\nu)).\end{aligned}$$

As $\langle\cdot,\cdot\rangle$ is of weight zero, the first summand is only non-zero for $\nu = -m$, while the second summand is only non-zero for $\nu = n+m$. Thus we have

$$\begin{aligned}[L(\alpha), q(\beta)] = -\frac{m}{2}\sum \big(&\langle e_{(1)}(-m), \beta\rangle q(e_{(2)}(-m)\alpha)\\ &+\langle e_{(2)}(n+m)\alpha, \beta\rangle q(e_{(1)}(n+m))\big).\end{aligned}$$

As $e_{(1)}(\cdot)$ is the dual basis to $e_{(2)}(\cdot)$, the right hand side simplifies to $-mq(\alpha\beta)$, which proves the Lemma. □

We use Lemma 3.3 to prove the following Proposition, which has already appeared in [Leh99] for the untwisted, projective case:

Proposition 3.4. *The map L is a weighted, graded action of the Virasoro algebra $\mathfrak{v}_{X,L}$ on $V_{X,L}$.*

Proof. Let $\alpha \in A[-2](m)$ and $\beta \in A[-2](n)$ with $m, n \in \mathbf{Z}$. We have to prove that $[L(\alpha), L(\beta)] = (m-n)L(\alpha\beta) - e(\alpha, \beta)$. We follow ideas in [FLM88]. In all summations below, ν runs through all integers if not specified otherwise.

We begin with the case $n \neq 0$ and $m+n \neq 0$. In this case, by Lemma 3.3, it is

$$[L(\alpha), L(\beta)] = \frac{1}{2}\left[L(m), \sum\sum_{\nu} q(e_{(1)}(\nu))q(e_{(2)}(\nu)\beta)\right] =$$

$$= \frac{1}{2}\left(\sum\sum_{\nu}(-\nu)q(e_{(1)}(\nu)\alpha)q(e_{(2)}(\nu)\beta) + \sum\sum_{\nu}(\nu-n)q(e_{(1)}(\nu))q(e_{(2)}(\nu)\alpha\beta)\right).$$

As $\sum q(e_{(1)}(\nu)(\alpha)q(e_{(2)}(\nu)\beta) = q(e_{(1)}(\nu+m))q(e_{(2)}(\nu-m)\alpha\beta)$, the right hand side is equal to

$$\frac{1}{2}\sum\sum_{\nu}\big((-\nu)q(e_{(1)}(\nu+m))q(e_{(2)}(\nu+m)\alpha\beta) + \\ +(\nu-n)q(e_{(1)}(\nu))q(e_{(2)}(\nu)\alpha\beta)\big),$$

which is nothing else than $(m-n)L(\alpha\beta)$. Note that $e(\alpha,\beta) = 0$ in this case. The next case we study is $m > 0$ and $n = -m$. In order to ensure convergence in the following calculations we have to split up $L(\beta)$ as follows:

$$L(\beta) = \sum\sum_{\nu\geq m} q(e_{(1)}(\nu)\beta)q(e_{(2)}(\nu)) + \sum\sum_{\nu<m} q(e_{(2)}(\nu))q(e_{(1)}(\nu)\beta)$$

Calculating the commutator $[L(\alpha), L(\beta)]$ thus yields the four terms:

$$\frac{1}{2}\sum\sum_{\nu\geq m}(m-\nu)q(e_{(1)}(\nu)\alpha\beta)q(e_{(2)}(\nu)) + \frac{1}{2}\sum\sum_{\nu\geq m}\nu q(e_{(1)}(\nu)\beta)q(e_{(1)}(\nu)\alpha) +$$

$$+\frac{1}{2}\sum\sum_{\nu<m}\nu q(e_{(2)}(\nu)\alpha)q(e_{(1)}(\nu)\beta) + \frac{1}{2}\sum\sum_{\nu<m}(m-\nu)q(e_{(2)}(\nu))q(e_{(1)}(\nu)\alpha\beta).$$

As in the first case, we now move α and β rightwards. Then we can split off an infinite part given by a multiple of $L(\alpha\beta)$ and are left over with the finite sum

$$[L(\alpha), L(\beta)] - 2mL(\alpha\beta) \\ = \frac{1}{2}\sum\sum_{\nu=0}^{m}(m-\nu)\left(q(e_{(2)}(\nu))q(e_{(1)}(\nu)\alpha\beta) - q(e_{(1)}(\nu))q(e_{(2)}(\nu)\alpha\beta)\right).$$

The right side is exactly $e(\alpha,\beta)$.

The remaining cases either follow from the above by exchanging n and m or are trivial ($n = m = 0$). □

4. The boundary operator

We proceed as in [Leh99] by introducing a boundary operator on $V_{X,L}$. Recall the definition of the tautological classes of the Hilbert scheme [LQW02]: Let Ξ^n be the universal family over $X^{[n]}$, which is a subscheme of $X^{[n]} \times X$. We denote the projections of $X^{[n]} \times X$ onto its factors by p and q. To each $\alpha \in H^*(X, \mathbf{C})$ we associate the *tautological classes*

$$\alpha^{[n]} := p_*(\mathrm{ch}(\mathcal{O}_{\Xi^n}) \cup q^*(\mathrm{td}(X) \cup \alpha))$$

in $H^*(X^{[n]}, \mathbf{C})$.

Remark 4.1. Note that the tautological classes live in the cohomology with untwisted coefficients, and we have not generalised this concept to the twisted case.

Each $\alpha \in H^*(X, \mathbf{C})$ defines an operator $m(\alpha) \in \operatorname{End}(V_{X,L})$, which is given by $m(\alpha)(\beta) := \alpha^{[n]} \cup \beta$ for all $\beta \in H^*(X^{[n]}, L^{[n]})$. It is an operator of weight zero. As it does not respect the grading, we split it up into its homogeneous components $m(\alpha) = \sum m^*(\alpha)$ with respect to the grading. Following [Leh99], we set $\partial := m^2(1)$ and call it the *boundary operator*. It is an operator of weight 0 and degree 2. For each operator $p \in \operatorname{End}(V_{X,L})$, we set $p' := [\partial, p]$ and call it the *derivative of* p.
The main theorem in [Leh99] is the calculation of the derivatives of the Heisenberg operators in the untwisted, projective case. In the sequel, we will do this in our more general case:
Let K be the canonical divisor class of X. We make it into an operator $K\colon A_{X,L} \to A_{X,L}[-2]$ of weight zero by setting

$$K(\alpha) := \frac{|n|-1}{2} K\alpha$$

for $\alpha \in H^*(X, L^n[2])$.

Proposition 4.2. *For all $\alpha, \beta \in A_{X,L}$ the following holds:*

$$[q'(\alpha), q(\beta)] = -q([\boldsymbol{d}, \alpha][\boldsymbol{d}, \beta]) - \int K([\boldsymbol{d}, \alpha])[\boldsymbol{d}, \beta].$$

Proof. Let us first consider the case of $\alpha \in A(m)$ and $\beta \in A(n)$ with $n+m \neq 0$. We have to show that $[q'(\alpha), q(\beta)] = -nmq(\alpha\beta)$. This is proven in [Leh99] for the projective, untwisted case. The proof in [Leh99] is based on calculating the commutator on the level of cycles. As these calculations are local in X, the result remains true for non-projective X. Furthermore, the proof literally works in the twisted case.
The case $n + m = 0$ remains. Here we have to show that $[q'(\alpha), q(\beta)] = m^2 \frac{|m|-1}{2} \int K\alpha\beta$. In [Leh99] the following intermediate result is formulated for the projective, untwisted case: For all $m \in \mathbf{Z}$, there exists a class $K_m \in H^*(X, \mathbf{C})$ such that $[q'(\alpha), q(\beta)] = m^2 \mathrm{id} \int K_m \alpha\beta$. As above the proof for this intermediate result that is given in [Leh99] also works in the twisted and non-projective case. The classes K_m do not depend on the choice of L, i.e. are universal for the surface. In [Leh99], the classes K_m are computed for the projective case, namely $K_m = \frac{|m|-1}{2} K$, where K is the class of the canonical divisor. All that remains is to calculate the classes K_m for the non-projective (untwisted) case. As $[q'(\alpha), q(\beta)] = [q'(\beta), q(\alpha)]$ (up to Koszul signs), it is enough to calculate K_m for $m > 0$:
Let $\beta \in A_{X,\mathbf{C}}(-m) = H_c^*(X, \mathbf{C}[2])$. Consider an open embedding $j\colon X \to \hat{X}$ of X into a smooth, projective surface $\hat{X}$. We denote the corresponding embeddings $X^{[n]} \to \hat{X}^{[n]}$ also by the letter j. Denote the 1 in $A_{\hat{X},\mathbf{C}}(m) =$

$H^*(X, \mathbf{C}[2])$ by $1(m)$. As all constructions considered so far are functorial (in the appropriate senses) with respect to open embeddings, we have

$$j^*[q'(1(m)), q(j_*\beta)]|0\rangle = [q'(j^*1(m)), q(\beta)]|0\rangle.$$

The right hand side is given by $m^2 \int K_m \beta$, where K_m is the class corresponding to X. By the calculations in [Leh99], the left hand side is given by $m^2 \frac{|m|-1}{2} \int K_{\hat{X}} j_* \beta$, where $K_{\hat{X}}$ is the canonical divisor class of $\hat{X}$. As $j^* K_{\hat{X}} = K_X$, we see that $K_m = \frac{|m|-1}{2} K$ also holds in the non-projective case, which proves the Proposition. □

Corollary 4.3. *For all $\alpha \in A_{X,L}$, the following holds:*

$$q'(\alpha) = L([\boldsymbol{d}, \alpha]) + q(K([\boldsymbol{d}, \alpha])).$$

Proof. This can be deduced from 4.2 as the respective statement for the untwisted, projective case is proven in [Leh99]. □

5. The ring structure

From now on, we assume that the canonical divisor of X is numerically trivial.

Example 5.1. Let H be a graded Frobenius algebra of degree d. Recall the symmetric non-degenerate bilinear form $\langle \cdot, \cdot \rangle \colon H \otimes H \to \mathbf{C}, h \otimes h' \to \int hh'$. It defines an isomorphism between H and its dual $H^\vee$. We can use this to dualise the multiplication map $H \otimes H \to H$ to a map $\Delta \colon H \to H \otimes H, h \mapsto \sum h_{(1)} \otimes h_{(2)}$ (in Sweedler notation) of degree d. It is coassociative and cocommutative (this follows from the associativity and commutativity of the multiplication map of H). Further, this map is characterised by

$$\sum \langle h_{(1)}, e\rangle \, \langle h_{(2)}, f\rangle = \langle h, ef\rangle$$

for all $e, f \in H$. It follows that $\Delta(gh) = \sum (gh_{(1)}) \otimes h_{(2)}$ for all $g \in H$. Thus Δ is a homomorphisms of H-modules when we view $H \otimes H$ as a left H-algebra by scalar multiplication on the first factor. (By the cocommutativity of Δ we could have equally chosen the analogously defined right H-algebra structure on $H \otimes H$.)

The example leads us to the following definition when we forget about the linear form $\int$:

A *non-counital graded Frobenius algebra H of degree d (over the complex numbers)* is a graded vector space over $\mathbf{C}$ with a (graded) commutative and associative multiplication of degree d and a unit element 1 (of degree $-d$) together with a coassociative and cocommutative H-module homomorphism $\Delta \colon H \to H \otimes H$ of degree d where we regard $H \otimes H$ as a left H-algebra by multiplying on the left factor. The map Δ is called the *diagonal*.

By example 5.1, every graded Frobenius algebra is in particular a non-counital graded Frobenius algebra.

Let G be a finite abelian group. A *G-weighting on H* is an action of G on H compatible with the Frobenius structure on H. In other words, H comes together with a weight decomposition of the form $\bigoplus_{\chi \in G^\vee} H(\chi)$, where each $g \in G$ acts on $H(\chi)$ by multiplication with $\chi(g)$.

Example 5.2. Let $p\colon \hat{X} \to X$ be a finite abelian Galois covering of X with Galois group G. Then G acts on $M := p_* \mathbf{C}$. Write $M = \bigoplus_{\chi \in G^\vee} L_\chi$, where G acts on each L_χ by multiplication via the character χ. The multiplication on the local system on $\mathbf{C}$ induces a multiplication map $M \otimes M \to M$ and thus isomorphisms $L_\chi \otimes L_{\chi'} \cong L_{\chi\chi'}$ of local systems on X that are commutative and associative in a certain sense. Thus we may assume without loss of generality that these isomorphisms are in fact equalities.
The G-weighted vector space

$$H_{X,G} := \bigoplus_{\chi \in G^\vee} H^*(X, L_\chi[2])$$

is naturally a non-counital G-weighted, graded Frobenius algebra of degree 2 as follows: the grading is given by the cohomological grading. The multiplication is given by the cup product. The diagonal is given by the proper push-forward $\delta_*\colon H_{X,G} \to H_{X,G} \otimes H_{X,G}$ that is induced by the diagonal map $\delta\colon X \to X \times X$. (The map δ_* is indeed a module homomorphism with respect to the left (or, equivalently, right) module structure on $H_{X,G} \otimes H_{X,G}$ as one can see as follows: Let $\pi\colon X \times X \to X$ be the projection onto the left factor. Then one has by the projection formula that

$$\delta_*(\alpha \cup \beta) = \delta_*((\delta^* \pi^* \alpha) \cup \beta) = (\pi^* \alpha) \cup (\delta_* \beta)$$

for all $\alpha, \beta \in H_{X,G}$.)

By iterated application, Δ induce maps $\Delta\colon H \to H^{\otimes n}$ with $n \geq 1$. We denote the restriction of $\Delta\colon H \to H^{\otimes n}$ to $H(L_\chi^n)$, $\chi \in G^\vee$, followed by the projection onto $H(L_\chi)^{\otimes n}$ by $\Delta(\chi)\colon H(L_\chi^n) \to H(L_\chi)^{\otimes n}$. The element $e := (\nabla \circ \Delta(1))(1) \in H$ is called the *Euler class of* H, where $\nabla\colon H \otimes H \to H$ is the multiplication map.
There is a construction given in [LS03] that associates to each graded Frobenius algebra H of degree of d a sequence of graded Frobenius algebras $H^{[n]}$ (whose degrees are given by nd). We extend this construction to G-weighted not necessarily counital Frobenius algebras as follows: For each $\chi \in G^\vee$, set

$$H_n(\chi) := \bigoplus_{\sigma \in \mathfrak{S}_n} \left(\bigotimes_{B \in \sigma \backslash [n]} H(L_\chi^{|B|}) \right) \boldsymbol{\sigma} \quad \text{and} \quad H_n := \bigoplus_{\chi \in G^\vee} H_n(\chi),$$

where $[n] := \{1, \ldots, n\}$ and $\sigma \backslash [n]$ is the set of orbits of the action of the cyclic group generated by σ on the set $[n]$. (Note that $H_n(1) = H(1)\{\mathfrak{S}_n\}$ in the terminology of [LS03].) The symmetric group $\mathfrak{S}_n$ acts on H_n. The graded vector space of invariants, $H_n^{\mathfrak{S}_n}$, is denoted by $H^{[n]}$.

Let $f\colon I \to J$ a surjection of finite sets and $(n_i)_{i\in I}$ a tuple of integers. Fibrewise multiplication yields ring homomorphisms

$$\nabla^{I,J} := \nabla^f\colon \bigotimes_{i\in I} H(L_\chi^{n_i}) \to \bigotimes_{j\in J} H\left(L_\chi^{\sum_{f(i)=j} n_i}\right)$$

of degree $d(|I| - |J|)$. (These correspond to the ring homomorphism $f^{I,J}$ in [LS03].) Dually, by using the diagonal morphisms $\Delta(\chi)$ and relying on their coassociativity and cocommutativity, we can define ∇^f-module homomorphisms

$$\Delta_{J,I} := \Delta_f\colon \bigotimes_{j\in J} H\left(L_\chi^{\sum_{f(i)=j} n_i}\right) \to \bigotimes_{i\in I} H(L_\chi^{n_i}),$$

which are also of degree $d(|I| - |J|)$. (These correspond to the module homomorphisms $f_{J,I}$ in [LS03]).
Let $\sigma, \tau \in \mathfrak{S}_n$ be two permutations. By $\langle\sigma,\tau\rangle$ we denote the subgroup of $\mathfrak{S}_n$ generated by the two permutations. Note that there are natural surjections $\sigma\backslash[n] \to \langle\sigma,\tau\rangle\backslash[n]$, $\tau\backslash[n] \to \langle\sigma,\tau\rangle\backslash[n]$, and $(\sigma\tau)\backslash[n] \to \langle\sigma,\tau\rangle\backslash[n]$. The corresponding ring and module homomorphism are denoted by $\nabla^{\sigma,\langle\sigma,\tau\rangle}$, etc., and $\Delta_{\langle\sigma,\tau\rangle,\sigma}$, etc.
Let $\chi, \chi' \in G^\vee$. We define a linear map

$$m_{\sigma,\tau}\colon \bigotimes_{B\in\sigma\backslash[n]} H(L_\chi^{|B|}) \otimes \bigotimes_{B\in\tau\backslash[n]} H(L_{\chi'}^{|B|}) \to \bigotimes_{B\in(\sigma\tau)\backslash[n]} H(L_{\chi\chi'}^{|B|})$$

by

$$m_{\sigma,\tau}(\alpha\otimes\beta) = \Delta_{\langle\sigma,\tau\rangle,(\sigma\tau)}(\nabla^{\sigma,\langle\sigma,\tau\rangle}(\alpha)\nabla^{\tau,\langle\sigma,\tau\rangle}(\beta)e^{\gamma(\sigma,\tau)}),$$

where the expression $e^{\gamma(\sigma,\tau)}$ is defined as in [LS03] (we have to use our Euler class e, which is defined above). This defines a product $H_n \otimes H_n \to H_n$ which is given by

$$(\alpha\boldsymbol{\sigma}) \cdot (\beta\boldsymbol{\tau}) := m_{\sigma,\tau}(\alpha,\beta)\boldsymbol{\sigma\tau}$$

for $\alpha\boldsymbol{\sigma} \in H_n(L_\chi)$ and $\beta\boldsymbol{\tau} \in H_n(L_{\chi'})$. This product is associative, $\mathfrak{S}_n$-equivariant, and of degree nd, which can be proven exactly as the corresponding statements about the product of the rings $H\{\mathfrak{S}_n\}$, which are defined in [LS03]. The product becomes (graded) commutative when restricted to $H^{[n]}$. Thus we have made $H^{[n]}$ a graded commutative, unital algebra of degree nd.

Definition 5.3. The algebra $H^{[n]}$ is the *n-th Hilbert algebra of H.*

In case G is trivial, the n-th Hilbert algebra of H defined here is exactly the algebra $H^{[n]}$ of [LS03]. For non-trivial G, this is no longer true.
The underlying graded vector space of $\bigoplus_{n\geq 0} H^{[n]}(L_\chi)$ is naturally isomorphic to $S(L_\chi) := S^*(\bigoplus_{n\geq 1} H(L_\chi^n))$, namely as follows: Firstly, we introduce linear maps $H_n(L_\chi) \to S(L_\chi)$, which are defined by mapping an element of the form $\sum_{\sigma\in\mathfrak{S}_n} \bigotimes_{B\in\sigma\backslash[n]} \alpha_{\sigma,B}\boldsymbol{\sigma}$ to $\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} \prod_{B\in\sigma\backslash[n]} \alpha_{\sigma,B}$. The restrictions of these morphisms to the $\mathfrak{S}_*$-invariant parts define a linear map $\bigoplus_{n\geq 0} H^{[n]}(L_\chi) \to$

$S(L_\chi)$. This map is an isomorphism, which can be proven exactly as it is in [LS03] for trivial G.
Recall that $H^*(X, \mathbf{C}[2])$ is a (trivially weighted) graded non-counital Frobenius algebra of degree d.

Lemma 5.4. *There is a natural isomorphism $H^*(X, \mathbf{C}[2])^{[n]} \to H^*(X^{[n]}, \mathbf{C}[2n])$ of graded unital algebras of degree nd.*

Proof. Recall the just defined isomorphism between the spaces $\bigoplus_{n\geq 0} H^*(X, \mathbf{C}[2])^{[n]}$ and $S^*(\bigoplus_{\nu>0} H^*(X, \mathbf{C}[2]))$ (for the trivial character $\chi = 1$). The composition of this isomorphism with isomorphism between the spaces $S^*(\bigoplus_{\nu>0} H^*(X, \mathbf{C}[2]))$ and $\bigoplus_{n\geq 0} H^*(X^{[n]}, \mathbf{C}[2n])$ of Theorem 1.2 induces by restriction the claimed isomorphism of the Lemma on the level of graded vector spaces.
That this isomorphism is in fact an isomorphism of unital algebras, is proven in [LS03] for X being projective. The proof there does not use the fact that $H^*(X, \mathbf{C}[2])$ has a counit, in fact it only uses its diagonal map. It relies on the earlier work in [Leh99], which has been extended to the non-projective case above, and [LQW02], which can similarly be extended. Thus the proof in [LS03] also works in the non-projective case, when we replace the notion of a Frobenius algebra by the notion of a non-counital Frobenius algebra. □

We will now deduce Theorem 1.4 from Lemma 5.4:

Proof of Theorem 1.4. Let $\chi, \chi' \in G^\vee$. Set $L := L_\chi$ and $M := L_{\chi'}$.
Let $\lambda = (\lambda_1, \ldots, \lambda_l)$ be a partition of n. Let ν_i the multiplicity of i in λ, i.e. $\lambda = \sum_i \nu_i \cdot i$. Set $X^{(\lambda)} := \prod_i X^{(\nu_i)}$, and $L^{(\lambda)} := \prod_i \mathrm{pr}_i^* L^{(\nu_i)}$, where the pr_i denote the projections onto the factors $X^{(\nu_i)}$. Let $\alpha = \sum \alpha_{(1)} \cdots \alpha_{(r)} \in H^*(X^{(\lambda)}, L^{(\lambda)}[2l]) = \bigotimes_i S^{\nu_i} H^*(X, L^i[2])$.
We set

$$|\alpha\rangle := \sum q(\alpha_{(1)}) \cdots q(\alpha_{(r)})|0\rangle.$$

By Theorem 1.2, the cohomology space $H^*(X^{[n]}, L^{[n]}[2n])$ is linearly spanned by classes of the form $|\alpha\rangle$.
Let $\mu = (\mu_1, \ldots, \mu_m)$ be another partition of n and $\beta \in H^*(X^{(\mu)}, M^{(\mu)}[2m])$. In order to describe the ring structure of $H^*(X^{[n]}, L^{[n]}[2n])$, we have to calculate the classes $|\alpha \cup \beta\rangle := |\alpha\rangle \cup |\beta\rangle$ in terms of the vector space description given by Theorem 1.2.
This means that we have to calculate the numbers

$$\langle \gamma \mid \alpha \cup \beta\rangle := q(\gamma)|\alpha \cup \beta\rangle \in H^*(X^{[0]}, \mathbf{C}) = \mathbf{C}$$

for all $\gamma \in H_c^*(X^{(\kappa)}, ((LM)^{-1})^{(\kappa)}[2k])$ for all partitions $\kappa = (\kappa_1, \ldots, \kappa_k)$ of n, and we have to show that they are equal to the numbers that would come out if we calculated the product of α and β by the right hand side of the claimed isomorphism of the Theorem.
The class $|\alpha\rangle$ is given by applying a sequence of correspondences to the vacuum vector: Recall from [Nak97] how to compose correspondences. It follows that

$|\alpha\rangle$ is given by
$$\mathrm{PD}^{-1}(\mathrm{pr}_1)_*(\mathrm{pr}_2^*\alpha \cap \zeta_\lambda),$$
where the symbols have the following meaning: The maps pr_1 and pr_2 are the projections of $X^{[n]} \times X^{(\lambda)}$ onto its factors $X^{[n]}$ and $X^{(\lambda)}$. Further, ζ_λ is a certain class in $H_*^{\mathrm{BM}}(Z_\lambda)$, where Z_λ is the incidence variety
$$Z_\lambda := \left\{ (\xi, (\underline{x}_1, \underline{x}_2, \ldots)) \in X^{[n]} \times X^{(\lambda)} \mid \operatorname{supp} \xi = \sum_i i \underline{x}_i \right\}$$
in $X^{[n]} \times X^{(\lambda)}$. (Note that $\mathrm{pr}_1^* L^{[n]}|_{Z_\lambda} = \mathrm{pr}_2^* L^{(\lambda)}|_{Z_\lambda}$, and that $p|_{Z_\lambda}$ is proper.) For $|\beta\rangle$ and $|\gamma\rangle$ we get similar expressions. By definition of the cup-product (pull-back along the diagonal), it follows that $\langle \gamma \mid \alpha \cup \beta\rangle = \langle r^*\gamma \cup p^*\alpha \cup q^*\beta, \zeta_{\lambda,\mu,\kappa}\rangle$, where p, q, and r are the projections from $X^{(\lambda)} \times X^{(\mu)} \times X^{(\kappa)}$ onto its three factors, and $\zeta_{\lambda,\mu,\kappa}$ is a certain class in $H_*^{\mathrm{BM}}(Z_{\lambda,\mu,\kappa})$ with
$$Z_{\lambda,\mu,\kappa} := \\ = \left\{ ((\underline{x}_1, \underline{x}_2, \ldots), (\underline{y}_1, \underline{y}_2, \ldots), (\underline{z}_1, \underline{z}_2, \ldots)) \mid \sum_i i\underline{x}_i = \sum_j j \underline{y}_j = \sum_k k \underline{z}_k \right\}.$$
(The incidence variety is proper over any of the three factors, so everything is well-defined.) The main point is now that the incidence variety $Z_{\lambda,\mu,\kappa}$ and the homology class $\zeta_{\lambda,\mu,\kappa}$ are independent of the local systems L and M. In particular, we can calculate $\zeta_{\lambda,\mu,\kappa}$ once we know the cup-product in the case $L = M = \mathbf{C}$. But this is the case that is described in Lemma 5.4, which we will analyse now.

First of all, the incidence variety is given by
$$Z_{\lambda,\mu,\kappa} = \sum_{\sigma,\tau} Z_{\sigma,\tau}$$
where σ and τ run through all permutations with cycle type λ and μ, respectively, such that $\rho := \sigma\tau$ has cycle type κ. The varieties $Z_{\sigma,\tau}$ are defined as follows:

As the orbits of the group action of $\langle\sigma\rangle$ on $[n]$ correspond to the entries of the partition λ, there exists a natural map $X^{\sigma\backslash[n]} \to X^{(\lambda)}$, which is given by symmetrising. Furthermore the natural surjection $\sigma\backslash[n] \to \langle\sigma,\tau\rangle\backslash[n]$ induces a diagonal embedding $X^{\langle\sigma,\tau\rangle\backslash[n]} \to X^{\sigma\backslash[n]}$. Composing both maps, we get a natural map $X^{\langle\sigma,\tau\rangle\backslash[n]} \to X^{(\lambda)}$. Analoguously, we get maps from $X^{\langle\sigma,\tau\rangle\backslash[n]}$ to $X^{(\mu)}$ and $X^{(\kappa)}$. Together, these maps define a diagonal embedding
$$i_{\tau,\sigma} \colon X^{\langle\sigma,\tau\rangle\backslash[n]} \to X^{(\kappa)} \times X^{(\lambda)} \times X^{(\mu)}.$$
We define $Z_{\sigma,\tau}$ to be the image of this map.

By Lemma 5.4, the class $\zeta_{\lambda,\mu,\kappa}$ is given by $\sum_{\sigma,\tau}(i_{\sigma,\tau})_*\zeta_{\sigma,\tau}$, where each class $\zeta_{\sigma,\tau} \in H_*^{\mathrm{BM}}(X^{\langle\sigma,\tau\rangle\backslash[n]})$ is Poincaré dual to $c_{\sigma,\tau}\, e^{\gamma(\sigma,\tau)}$. Here, $c_{\sigma,\tau}$ is a certain combinatorial factor (possibly depending on σ and τ), whose precise value is of no concern for us.

Having derived the value of $\zeta_{\lambda,\mu,\kappa}$ from Lemma 5.4, we have thus calculated the value $\langle\gamma \mid \alpha \cup \beta\rangle$.
Now we have to compare this value with the one that is predicted by the description of the cup-product given by the right hand side of the claimed isomorphism of the Theorem. With the same analysis as above, we find this value is also given by a correspondence on $Z_{\lambda,\mu,\kappa}$ with the class $\sum_{\tau,\sigma}(i_{\sigma,\tau})_* c_{\sigma,\tau}\mathrm{PD}(e^{\gamma(\sigma,\tau)})$ with the same combinatorial factors $c_{\sigma,\tau}$ as above. We thus find that the claimed ring structure yields the correct value of $\langle\gamma \mid \alpha \cup \beta\rangle$. $\square$

Remark 5.5. One can also define a natural diagonal map for the Hilbert algebras $H^{[n]}$ making them into graded, non-counital Frobenius algebras of degree nd. The isomorphism of Theorem 1.4 then becomes an isomorphism of graded non-counital Frobenius algebras.

6. The generalised Kummer varieties

Finally, we want to use Theorem 1.4 to study the cohomology ring of the generalised Kummer varieties.
Let H be a non-counital graded Frobenius algebra of degree d that is moreover endowed with a compatible structure of a cocommutative Hopf algebra of degree d. The comultiplication δ of the Hopf algebra structure is of degree $-d$. The counit of the Hopf algebra structure is denoted by ϵ and is of degree d. We further assume that H is also equipped with a G-weighting for a finite abelian group G.

Example 6.1. Let X be an abelian surface. The group structure on X induces naturally a graded Hopf algebra structure of degree 2 on the graded Frobenius algebra $H^*(X, \mathbf{C}[2])$. This algebra is also trivially $X[n]$-weighted, where $G := X[n] \simeq (\mathbf{Z}/(n))^4$ is the group of n-torsion points on X. (Trivially weighted means that the only non-trivial $X[n]$-weight space of $H^*(X, \mathbf{C}[2])$ is the one corresponding to the identity element 0.)

Let n be a positive integer. Recall the definition of the (G-weighted) Hilbert algebra $H^{[n]}$. Repeated application of the comultiplication δ induces a map $\delta\colon H \to H^{\otimes n} = H^{\mathrm{id}\backslash[n]}$, which is of degree $-(n-1)d$. Its image lies in the subspace of symmetric tensors. Thus we can define a map $\phi\colon H \to H^{[n]}$ with $\phi(\alpha) := \delta(\alpha)\mathbf{id}$. One can easily check that this map is an algebra homomorphism of degree $-(n-1)d$, making $H^{[n]}$ into an H-algebra.
Define

$$H^{[[n]]} := H^{[n]} \otimes_H \mathbf{C},$$

where we view $\mathbf{C}$ as an H-algebra of degree d via the Hopf counit ϵ. It is $H^{[[n]]}$ a (G-weighted) graded Frobenius algebra of degree nd.

Definition 6.2. The algebra $H^{[[n]]}$ is the *n-th Kummer algebra of H*.

The reason of this naming is of course Theorem 1.7.

Proof of Theorem 1.7. Let $n\colon X \to X$ denote the morphism that maps x to $n \cdot x$. There is a natural cartesian square

$$
\begin{tikzcd}
X \times X^{[[n]]} \arrow[r, "\nu"] \arrow[d, "p"'] & X^{[n]} \arrow[d, "\sigma"] \\
X \arrow[r, "n"'] & X,
\end{tikzcd}
\tag{2}
$$

where p is the projection on the first factor and ν maps a pair (x, ξ) to $x + \xi$, the subscheme that is given by translating ξ by x ([Bea83]). Then G is the Galois group of n. Each element χ of $G^\vee$ corresponds to a local system L_χ on X, and we have $n_* \mathbf{C} = \bigoplus_{\chi \in G^\vee} L_\chi$. It follows that ν is an abelian Galois covering of $X^{[n]}$ with $\nu_* \mathbf{C} = \bigoplus_{\chi \in G^\vee} L_\chi^{[n]}$.
Together with Theorem 1.4, this leads to the claimed description of the cohomology ring of $X^{[[n]]}$: Firstly, there is a natural isomorphism

$$H^*(X^{[[n]]}, \mathbf{C}[2n]) \to H^*(X \times X^{[[n]]}, \mathbf{C}[2n]) \otimes_{H^*(X, \mathbf{C}[2])} \mathbf{C}$$

of unital algebras (the tensor product is taken with respect to the map p^* and the Hopf counit $H^*(X, \mathbf{C}[2]) \to \mathbf{C}$). By the Leray spectral sequence for ν and by (2), the right hand side is naturally isomorphic to

$$H^*(X^{[n]}, \nu_* \mathbf{C}[2n]) \otimes_{H^*(X, \mathbf{C}[2])} \mathbf{C} = \bigoplus_{\chi \in G^\vee} H^*(X^{[n]}, L_\chi^{[n]}[2n]) \otimes_{H^*(X, \mathbf{C}[2])} \mathbf{C}$$

(where the tensor product is taken with respect to the map σ^* and the Hopf counit).
By Theorem 1.4, the algebra $\bigoplus_{\chi \in G^\vee} H^*(X^{[n]}, L_\chi^{[n]}[2n])$ is naturally isomorphic to $\bigoplus_{\chi \in G^\vee} H^*(X, L_\chi[2])^{[n]}$. Now $H^*(X, L_\chi[2]) = 0$ unless χ is the trivial character, which follows from the fact that all classes in $H^*(X, \mathbf{C})$ are invariant under the action of the Galois group of n, i.e. correspond to the trivial character. Thus there is a natural isomorphism

$$\bigoplus_{\chi \in G^\vee} H^*(X^{[n]}, L_\chi^{[n]}[2n]) \to H^*(X, \mathbf{C}[2])^{[n]},$$

of G-weighted algebras, where we endow $H^*(X, \mathbf{C}[2])$ with the trivial G-weighting. Under this isomorphism, the map σ^* corresponds to the homomorphism ϕ defined below Example 6.1. Thus we have proven the existence of a natural isomorphism

$$H^*(X^{[[n]]}, \mathbf{C}[2n]) \to H^*(X, \mathbf{C}[2])^{[n]} \otimes_{H^*(X, \mathbf{C}[2])} \mathbf{C}$$

of unital, graded algebras. But the right hand side is nothing but $H^{[[n]]}$, thus the Theorem is proven. $\square$

References

[Bea83] Arnaud Beauville, *Variétés kähleriennes dont la première classe de Chern est nulle.*, J. Differ. Geom. 18 (1983), 755–782 (French).

[Bri02] Michael Britze, *On the Cohomology of Generalized Kummer Varieties*, Ph.D. thesis, Mathematisch-Naturwissenschaftliche Fakultät der Universität zu Köln, 2002.

[Bro88] Ronald Brown, *Topology: A geometric account of general topology, homotopy types and the fundamental groupoid. Rev., updated and expanded ed.*, Ellis Horwood Series in Mathematics and its Applications. Chichester (UK): Ellis Horwood Ltd.; New York etc.: Halsted Press., 1988.

[Dou66] A. Douady, *Le problème des modules pour les sous-espaces analytiques compacts d'un espace analytique donne*, Sovremen. Probl. Teor. Analit. Funktsij, Mezhdunarod. Konf. Teor. Analit. Funktsij, Erevan 1965, 141–143, 1966.

[FLM88] Igor Frenkel, James Lepowsky, and Arne Meurman, *Vertex operator algebras and the monster.*, Pure and Applied Mathematics, 134. Boston etc.: Academic Press, Inc., 1988 (English).

[Fog68] J. Fogarty, *Algebraic families on an algebraic surface*, Am. J. Math. (1968), 511–521.

[Gro61] Alexandre Grothendieck, *Techniques de construction et théoremes d'existence en géométrie algébrique. IV: Les schemas de Hilbert*, Sem. Bourbaki 13(1960/61), No.221, 1961.

[Gro96] I. Grojnowski, *Instantons and affine algebras. I: The Hilbert scheme and vertex operators*, Math. Res. Lett. 3 (1996), no. 2, 275–291.

[GS93] Lothar Göttsche and Wolfgang Soergel, *Perverse sheaves and the cohomology of Hilbert schemes of smooth algebraic surfaces.*, Math. Ann. 296 (1993), no. 2, 235–245 (English).

[Leh99] Manfred Lehn, *Chern classes of tautological sheaves on Hilbert schemes of points on surfaces*, Invent. Math. 136 (1999), no. 1, 157–207.

[Leh04] ———, *Lectures on Hilbert schemes.*, Hurtubise, Jacques (ed.) et al., Algebraic structures and moduli spaces. Proceedings of the CRM workshop, Montréal, Canada, July 14-20, 2003. Providence, RI: American Mathematical Society (AMS). CRM Proceedings &; Lecture Notes 38, 1-30 (2004)., 2004.

[LQW02] Wei-ping Li, Zhenbo Qin, and Weiqiang Wang, *Vertex algebras and the cohomology ring structure of Hilbert schemes of points on surfaces.*, Math. Ann. 324 (2002), no. 1, 105–133 (English).

[LS03] Manfred Lehn and Christoph Sorger, *The cup product of Hilbert schemes for K3 surfaces.*, Invent. Math. 152 (2003), no. 2, 305–329 (English).

[Nak97] Hiraku Nakajima, *Heisenberg algebra and Hilbert schemes of points on projective surfaces*, Ann. of Math. (2) 145 (1997), no. 2, 379–388.

[PS08] Chris A. M. Peters and Joseph H. M. Steenbring, *Mixed Hodge Structures*, Springer-Verlag, 2008 (English).

Marc A. Nieper-Wißkirchen
Institut für Mathematik
Universität Augsburg
86157 Augsburg
Germany
marc.nieper-wisskirchen@math.uni-augsburg.de

The Sato-Tate Conjecture for Modular Forms of Weight 3

Toby Gee[1]

Received: August 26, 2009
Revised: December 15, 2009

Communicated by Don Blasius

Abstract. We prove a natural analogue of the Sato-Tate conjecture for modular forms of weight 2 or 3 whose associated automorphic representations are a twist of the Steinberg representation at some finite place.

2000 Mathematics Subject Classification: 11F33.

Contents

[1]The author was partially supported by NSF grant DMS-0841491

1. Introduction

The Sato-Tate conjecture is a conjecture about the distribution of the number of points on an elliptic curve over finite fields. Specifically, if E is an elliptic curve over $\mathbb{Q}$ without CM, then for each prime l such that E has good reduction at l we set

$$a_l := 1 + l - \#E(\mathbb{F}_l).$$

Then the Sato-Tate conjecture states that the quantities $\cos^{-1}(a_l/2\sqrt{l})$ are equidistributed with respect to the measure

$$\frac{2}{\pi}\sin^2\theta d\theta$$

on $[0,\pi]$. Alternatively, by the Weil bounds for E, the polynomial

$$X^2 - a_l X + l = (X - \alpha_l l^{1/2})(X - \beta_l l^{1/2})$$

satisfies $|\alpha_l| = |\beta_l| = 1$, and there is a well-defined conjugacy class $x_{E,l}$ in $SU(2)$, the conjugacy class of the matrix

$$\begin{pmatrix} \alpha_l & 0 \\ 0 & \beta_l \end{pmatrix}.$$

The Sato-Tate conjecture is then equivalent to the statement that the classes $x_{E,l}$ are equidistributed with respect to the Haar measure on $SU(2)$.

Tate observed that the conjecture would follow from properties of the symmetric power L-functions of E, specifically that these L-functions (suitably normalised) should have nonvanishing analytic continuation to the region $\Re s \geq 1$. This would follow (given the modularity of elliptic curves) from the Langlands conjectures (specifically, it would be a consequence of the symmetric power functoriality from GL_2 to GL_n for all n). Unfortunately, proving this functoriality appears to be well beyond the reach of current techniques. However, Harris, Shepherd-Baron and Taylor observed that the required analytic properties would follow from a proof of the potential automorphy of the symmetric power L-functions (that is, the automorphy of the L-functions after base change to some extension of $\mathbb{Q}$), and were able to use Taylor's potential automorphy techniques to prove the Sato-Tate conjecture for all elliptic curves E with non-integral j-invariant (see [HSBT09]).

There are various possible generalisations of the Sato-Tate conjecture; if one wishes to be maximally ambitious, one could consider equidistribution results for the Satake parameters of rather general automorphic representations (see for example section 2 of [Lan79]). Again, such results appear to be well beyond the range of current technology. There is, however, one special case that does seem to be reasonable to attack, which is the case of Hilbert cuspidal eigenforms of regular weight. In this paper, we prove a natural generalisation of the Sato-Tate conjecture for modular newforms (over $\mathbb{Q}$) of weight 2 or 3, subject to the natural analogue of the condition that an elliptic curve has non-integral j-invariant. We note that previously the only modular forms for which the conjecture was known were those corresponding to elliptic curves; in particular, there were no examples of weight 3 modular forms for which the conjecture was known. After this paper was made available, the conjecture was proved for all modular forms of weight at least 2 in [BLGHT09], by rather different methods. Our approach is similar to that of [HSBT09], and we are fortunate in being able to quote many of their results. Indeed, it is straightforward to check that Tate's argument shows that the conjecture would follow from the potential automorphy of the symmetric powers of the l-adic Galois representations associated to a modular form. One might then hope to prove this potential automorphy in the style of [HSBT09]; one would proceed by realising the symmetric powers of the mod l Galois representation geometrically in such a way that their potential automorphy may be established, and then deduce the potential automorphy of the l-adic representations by means of the modularity lifting theorems of [CHT08] and [Tay08].

It turns out that this simple strategy encounters some significant obstacles. First and foremost, it is an unavoidable limitation of the known potential automorphy methods that they can only deduce that a mod l Galois representation is automorphic of minimal weight (which we refer to as "weight 0"). However, the symmetric powers of the Galois representations corresponding to modular forms of weight greater than 2 are never automorphic of minimal weight, so one has no hope of directly proving their potential automorphy in the fashion outlined above without some additional argument. If, for example, one knew the weight part of Serre's conjecture for GL_n (or even for unitary groups) one would be able to deduce the required results, but this appears to be an extremely difficult problem in general. There is, however, one case in which the analysis of the Serre weights is rather easier, which is the case that the l-adic Galois representations are ordinary. It is this observation that we exploit in this paper.

In general, it is anticipated that for a given newform f of weight $k \geq 2$, there is a density one set of primes l such that there is an ordinary l-adic Galois representation corresponding to f. Unfortunately, if $k > 3$ then it is not even known that there is an infinite set of such primes; this is the reason for our restriction to $k = 2$ or 3. In these cases, one may use the Ramanujan conjecture and Serre's form of the Cebotarev density theorem (see [Ser81]) to prove that the set of l which are "ordinary" in this sense has density one, via an argument that is presumably well-known to the experts (although we have not been able to find the precise argument that we use in the literature). We note that it is important for us to be able to choose l arbitrarily large in certain arguments (in order to satisfy the hypotheses of the automorphy lifting theorems of [Tay08]), so it does not appear to be possible to apply our methods to any modular forms of weight greater than 3. Similarly, we cannot prove anything for Hilbert modular forms of parallel weight 3 over any field other than $\mathbb{Q}$.

We now outline our arguments in more detail, and explain exactly what we prove. The early sections of the paper are devoted to proving the required potential automorphy results. In section 2 we recall some basic definitions and results from [CHT08] on the existence of Galois representations attached to regular automorphic representations of GL_n over totally real and CM fields, subject to suitable self-duality hypotheses and to the existence of finite places at which the representations are square integrable. Section 3 recalls some standard results on the Galois representations attached to modular forms, and proves the result mentioned above on the existence of a density one set of primes for which there is an ordinary Galois representation.

In section 4 we prove the potential automorphy in weight 0 of the symmetric powers of the residual Galois representations associated to a modular form, under the hypotheses that the residual Galois representation is ordinary and irreducible, and the automorphic representation corresponding to the modular form is an unramified twist of the Steinberg representation at some finite place. The latter condition arises because of restrictions of our knowledge as to when there are Galois representations associated to automorphic representations on unitary groups, and it is anticipated that it will be possible to remove it in the near future. That would then allow us to prove our main theorems for any modular forms of weights 2 or 3 which are not of CM type. (Note added in proof: such results are now available, cf. [Shi09], [CH09], [Gue09], and it is thus an easy exercise to deduce our main results without any Steinberg assumption.)

One approach to proving the potential automorphy result in weight 0 would be to mimic the proofs for elliptic curves in [HSBT09]. In fact we can do better than this, and are able to directly utilise their results. We are reduced to proving that after making a quadratic base change and twisting, the mod l representation attached to our modular form is, after a further base change, congruent to a mod l representation arising from a certain Hilbert-Blumenthal abelian variety. This is essentially proved in [Tay02], and we only need to make minor changes to the proofs in [Tay02] in order to deduce the properties we need. We can then directly apply one of the main results of [HSBT09] to deduce the automorphy of the even-dimensional symmetric powers of the Hilbert-Blumenthal abelian variety, and after twisting back we deduce the required potential automorphy of our residual representations. Note that apart from resulting in rather clean proofs, the advantage of making an initial congruence to a Galois representation attached to an abelian variety and then using the potential automorphy of the symmetric powers of this abelian variety is that we are able to obtain local-global compatibility at all finite places (including those dividing the residue characteristic). This compatibility is not yet available for automorphic representations on unitary groups in general, and is needed in our subsequent arguments. In particular, it tells us that the automorphic representations of weight 0 which correspond to the symmetric powers of the l-adic representations coming from our Hilbert-Blumenthal abelian variety are ordinary at l.

In section 5 we exploit this ordinarity to deduce that the even-dimensional symmetric powers of the mod l representations are potentially automorphic of the "correct" weight. This is a basic consequence of Hida theory for unitary groups, but we are not aware of any reference that proves the precise result we need. Accordingly, we provide a proof in the style of the arguments of [Tay88].

There is nothing original in this section, and as the arguments are somewhat technical the reader may wish to skip it on a first reading.

The results of the preceding sections are combined in section 6 to establish the required potential automorphy results for l-adic (rather than mod l) representations. This essentially comes down to checking the hypotheses of the modularity lifting theorem that we wish to apply from [Tay08], which follow from the analogous arguments in [HSBT09] together with the conditions that we have imposed in our potential automorphy arguments. It is here that we need the freedom to choose l to be arbitrarily large, which results in our restriction to weights 2 and 3.

Finally, in section 7 we deduce the form of the Sato-Tate conjecture mentioned above. As in [HSBT09] we have only proved the potential automorphy of the even-dimensional symmetric powers of the l-adic representations associated to our modular form, and we deduce the required analytic properties for the L-functions attached to odd-dimensional symmetric powers via an argument with Rankin-Selberg convolutions exactly analogous to that of [HSBT09]. In fact, we need to prove the same results for the L-functions of certain twists of our representations by finite-order characters, but this is no more difficult.

We now describe the form of the final result, which is slightly different from that for elliptic curves, because our modular forms may have non-trivial nebentypus (and indeed are required to do so if they have weight 3). Suppose that the newform f has level N, nebentypus χ_f and weight k; then the image of χ_f is precisely the m-th roots of unity for some m. Then if $p \nmid N$ is a prime, we know that if

$$X^2 - a_p X + p^{k-1}\chi_f(p) = (X - \alpha_p p^{(k-1)/2})(X - \beta_p p^{(k-1)/2})$$

where a_p is the eigenvalue of f for the Hecke operator T_p, then the matrix

$$\begin{pmatrix} \alpha_p & 0 \\ 0 & \beta_p \end{pmatrix}$$

defines a conjugacy class $x_{f,p}$ in $U(2)_m$, the subgroup of $U(2)$ of matrices with determinant an m-th root of unity. Then our main result is

Theorem. *If f has weight 2 or 3 and the associated automorphic representation is a twist of the Steinberg representation at some finite place, then the conjugacy classes $x_{f,p}$ are equidistributed with respect to the Haar measure on $U(2)_m$ (normalised so that $U(2)_m$ has measure 1).*

One can make this more concrete by restricting to primes p such that $\chi_f(p)$ is a specific m-th root of unity; see the remarks at the end of section 7.

We would like to thank Thomas Barnet-Lamb, David Geraghty and Richard Taylor for various helpful discussions during the writing of this paper.

2. Notation and assumptions

We let ϵ denote the l-adic cyclotomic character, regarded as a character of the absolute Galois group of a number field or of a completion of a number field at

a finite place. We sometimes use the same notation for the mod l cyclotomic character; it will always be clear from the context which we are referring to. We denote Tate twists in the usual way, i.e. $\rho(n) := \rho \otimes \epsilon^n$. We write $\bar{K}$ for a separable closure of a field K. If x is a finite place of a number field F, we will write I_x for the inertia subgroup of $\mathrm{Gal}(\overline{F}_x/F_x)$. We fix an algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$, and regard all finite extensions of $\mathbb{Q}$ as being subfields of $\overline{\mathbb{Q}}$. We also fix algebraic closures $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$ for all primes p, and embeddings $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$.
We need several incarnations of the local Langlands correspondence. Let K be a finite extension of $\mathbb{Q}_p$, and $l \neq p$ a prime. We have a canonical isomorphism

$$\mathrm{Art}_K : K^\times \to W_K^{ab}$$

normalised so that geometric Frobenius elements correspond to uniformisers. Let $\mathrm{Irr}(\mathrm{GL}_n(K))$ denote the set of isomorphism classes of irreducible admissible representations of $\mathrm{GL}_n(K)$ over $\mathbb{C}$, and let $\mathrm{WDRep}_n(W_K)$ denote the set of isomorphism classes of n-dimensional Frobenius semi-simple complex Weil-Deligne representations of the Weil group W_K of K. The main result of [HT01] is that there is a family of bijections

$$\mathrm{rec}_K : \mathrm{Irr}(\mathrm{GL}_n(K)) \to \mathrm{WDRep}_n(W_K)$$

satisfying a number of properties that specify them uniquely (see the introduction to [HT01] for a complete list). Among these properties are:

- If $\pi \in \mathrm{Irr}(\mathrm{GL}_1(K))$ then $\mathrm{rec}_K(\pi) = \pi \circ \mathrm{Art}_K^{-1}$.
- $\mathrm{rec}_K(\pi^\vee) = \mathrm{rec}_K(\pi)^\vee$.
- If $\chi_1, \ldots, \chi_n \in \mathrm{Irr}(\mathrm{GL}_1(K))$ are such that the normalised induction n-Ind$(\chi_1, \ldots, \chi_n)$ is irreducible, then

$$\mathrm{rec}_K(\text{n-Ind}(\chi_1, \ldots, \chi_n)) = \oplus_{i=1}^n \mathrm{rec}_K(\chi_i).$$

We will often just write rec for rec_K when the choice of K is clear from the context. After choosing an isomorphism $\iota : \overline{\mathbb{Q}}_l \to \mathbb{C}$ one obtains bijections rec_l from the set of isomorphism classes of irreducible admissible representations of $\mathrm{GL}_n(K)$ over $\overline{\mathbb{Q}}_l$ to the set of isomorphism classes of n-dimensional Frobenius semi-simple Weil-Deligne $\overline{\mathbb{Q}}_l$-representations of W_K. We then define $r_l(\pi)$ to be the l-adic representation of $\mathrm{Gal}(\overline{K}/K)$ associated to $\mathrm{rec}_l(\pi^\vee \otimes |\cdot|^{(1-n)/2})$ whenever this exists (that is, whenever the eigenvalues of $\mathrm{rec}_l(\pi^\vee \otimes |\cdot|^{(1-n)/2})(\phi)$ are l-adic units, where ϕ is a Frobenius element). We will, of course, only use this notation where it makes sense. It is useful to note that

$$r_l(\pi)^\vee(1-n) = r_l(\pi^\vee).$$

Let M denote a CM field with maximal totally real subfield F (by "CM field" we always mean "imaginary CM field"). We denote the nontrivial element of $\mathrm{Gal}(M/F)$ by c. Following [CHT08] we define a RACSDC (regular, algebraic, conjugate self dual, cuspidal) automorphic representation of $\mathrm{GL}_n(\mathbb{A}_M)$ to be a cuspidal automorphic representation π such that

- $\pi^\vee \cong \pi^c$, and

- π_∞ has the same infinitesimal character as some irreducible algebraic representation of $\mathrm{Res}_{M/\mathbb{Q}}\,\mathrm{GL}_n$.

We say that $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,\mathbb{C})}$ is a weight if

- $a_{\tau,1} \geq \cdots \geq a_{\tau,n}$ for all $\tau \in \mathrm{Hom}(M,\mathbb{C})$, and
- $a_{\tau c,i} = -a_{\tau,n+1-i}$.

For any weight a we may form an irreducible algebraic representation W_a of $\mathrm{GL}_n^{\mathrm{Hom}(M,\mathbb{C})}$, the tensor product over τ of the irreducible algebraic representations of GL_n with highest weight a_τ. We say that π has weight a if it has the same infinitesimal character as $W_a^\vee$; note that any RACSDC automorphic representation has some weight. Let S be a non-empty finite set of finite places of M. For each $v \in S$, choose an irreducible square integrable representation ρ_v of $\mathrm{GL}_n(M_v)$ (in this paper, we will in fact only need to consider the case where each ρ_v is the Steinberg representation). We say that an RACSDC automorphic representation π has type $\{\rho_v\}_{v\in S}$ if for each $v \in S$, π_v is an unramified twist of $\rho_v^\vee$. There is a compatible family of Galois representations associated to such a representation in the following fashion.

Proposition 2.1. *Let $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. Suppose that π is an RACSDC automorphic representation of $\mathrm{GL}_n(\mathbb{A}_M)$ of type $\{\rho_v\}_{v\in S}$ for some nonempty set of finite places S. Then there is a continuous semisimple representation*

$$r_{l,\iota}(\pi) : \mathrm{Gal}(\overline{M}/M) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$$

such that

(1) *For each finite place $v \nmid l$ of M, we have*

$$r_{l,\iota}(\pi)|^{ss}_{\mathrm{Gal}(\overline{M}_v/M_v)} = r_l(\iota^{-1}\pi_v)^\vee(1-n)^{ss}.$$

(2) $r_{l,\iota}(\pi)^c = r_{l,\iota}(\pi)^\vee \epsilon^{1-n}$.

Proof. This follows from Proposition 4.2.1 of [CHT08] (which in fact also gives information on $r_{l,\iota}|_{\mathrm{Gal}(\overline{M}_v/M_v)}$ for places $v|l$). □

The representation $r_{l,\iota}(\pi)$ may be conjugated to be valued in the ring of integers of a finite extension of $\mathbb{Q}_l$, and we may reduce it modulo the maximal ideal of this ring of integers and semisimplify to obtain a well-defined continuous representation

$$\bar{r}_{l,\iota}(\pi) : \mathrm{Gal}(\overline{M}/M) \to \mathrm{GL}_n(\overline{\mathbb{F}}_l).$$

Let $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,\overline{\mathbb{Q}}_l)}$, and let $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. Define $\iota_* a \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,\mathbb{C})}$ by $(\iota_* a)_{\iota\tau,i} = a_{\tau,i}$. Now let ρ_v be a discrete series representation of $\mathrm{GL}_n(M_v)$ over $\overline{\mathbb{Q}}_l$ for each $v \in S$. If $r : \mathrm{Gal}(\overline{M}/M) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$, we say that r is automorphic of weight a and type $\{\rho_v\}_{v\in S}$ if $r \cong r_{l,\iota}(\pi)$ for some RACSDC automorphic representation π of weight $\iota_* a$ and type $\{\iota\rho_v\}_{v\in S}$. Similarly, if $\bar{r} : \mathrm{Gal}(\overline{M}/M) \to \mathrm{GL}_n(\overline{\mathbb{F}}_l)$, we say that $\bar{r}$ is automorphic of weight a and type $\{\rho_v\}_{v\in S}$ if $\bar{r} \cong \bar{r}_{l,\iota}(\pi)$ for some RACSDC automorphic representation π with π_l unramified, of weight $\iota_* a$ and type $\{\iota\rho_v\}_{v\in S}$.

We now consider automorphic representations of $\mathrm{GL}_n(\mathbb{A}_F)$. We say that a cuspidal automorphic representation π of $\mathrm{GL}_n(\mathbb{A}_F)$ is RAESDC (regular, algebraic, essentially self dual, cuspidal) if

- $\pi^\vee \cong \chi\pi$ for some character $\chi : F^\times\backslash\mathbb{A}_F^\times \to \mathbb{C}^\times$ with $\chi_v(-1)$ independent of $v|\infty$, and
- π_∞ has the same infinitesimal character as some irreducible algebraic representation of $\mathrm{Res}_{F/\mathbb{Q}}\mathrm{GL}_n$.

We say that $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(F,\mathbb{C})}$ is a weight if

$$a_{\tau,1} \geq \cdots \geq a_{\tau,n}$$

for all $\tau \in \mathrm{Hom}(F,\mathbb{C})$. For any weight a we may form an irreducible algebraic representation W_a of $\mathrm{GL}_n^{\mathrm{Hom}(F,\mathbb{C})}$, the tensor product over τ of the irreducible algebraic representations of GL_n with highest weight a_τ. We say that an RAESDC automorphic representation π has weight a if it has the same infinitesimal character as $W_a^\vee$. In this case, by the classification of algebraic characters over a totally real field, we must have $a_{\tau,i} + a_{\tau,n+1-i} = w_a$ for some w_a independent of τ. Let S be a non-empty finite set of finite places of F. For each $v \in S$, choose an irreducible square integrable representation ρ_v of $\mathrm{GL}_n(M_v)$. We say that an RAESDC automorphic representation π has type $\{\rho_v\}_{v\in S}$ if for each $v \in S$, π_v is an unramified twist of $\rho_v^\vee$. Again, there is a compatible family of Galois representations associated to such a representation in the following fashion.

Proposition 2.2. *Let $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. Suppose that π is an RAESDC automorphic representation of $\mathrm{GL}_n(\mathbb{A}_F)$, of type $\{\rho_v\}_{v\in S}$ for some nonempty set of finite places S, with $\pi^\vee \cong \chi\pi$. Then there is a continuous semisimple representation*

$$r_{l,\iota}(\pi) : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$$

such that

(1) *For each finite place $v \nmid l$ of F, we have*

$$r_{l,\iota}(\pi)|^{ss}_{\mathrm{Gal}(\overline{F}_v/F_v)} = r_l(\iota^{-1}\pi_v)^\vee(1-n)^{ss}.$$

(2) $r_{l,\iota}(\pi)^\vee = r_{l,\iota}(\chi)\epsilon^{n-1}r_{l,\iota}(\pi)$.

Here $r_{l,\iota}(\chi)$ is the l-adic Galois representation associated to χ via ι (see Lemma 4.1.3 of [CHT08]*).*

Proof. This is Proposition 4.3.1 of [CHT08] (which again obtains a stronger result, giving information on $r_{l,\iota}|_{\mathrm{Gal}(\overline{F}_v/F_v)}$ for places $v|l$). □

Again, the representation $r_{l,\iota}(\pi)$ may be conjugated to be valued in the ring of integers of a finite extension of $\mathbb{Q}_l$, and we may reduce it modulo the maximal ideal of this ring of integers and semisimplify to obtain a well-defined continuous representation

$$\bar{r}_{l,\iota}(\pi) : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{F}}_l).$$

Let $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(F,\overline{\mathbb{Q}}_l)}$, and let $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. Define $\iota_* a \in (\mathbb{Z}^n)^{\mathrm{Hom}(F,\mathbb{C})}$ by $(\iota_* a)_{\iota\tau,i} = a_{\tau,i}$. Let ρ_v be a discrete series representation of $\mathrm{GL}_n(M_v)$ over $\overline{\mathbb{Q}}_l$ for each $v \in S$. If $r : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$, we say that r is automorphic of weight a and type $\{\rho_v\}_{v\in S}$ if $r \cong r_{l,\iota}(\pi)$ for some RAESDC automorphic representation π of weight $\iota_* a$ and type $\{\iota\rho_v\}_{v\in S}$. Similarly, if $\bar{r} : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{F}}_l)$, we say that $\bar{r}$ is automorphic of weight a and type $\{\rho_v\}_{v\in S}$ if $\bar{r} \cong \bar{r}_{l,\iota}(\pi)$ for some RAESDC automorphic representation π with π_l unramified, of weight $\iota_* a$ and type $\{\iota\rho_v\}_{v\in S}$.
As in [HSBT09] we denote the Steinberg representation of $\mathrm{GL}_n(K)$, K a nonarchimedean local field, by $\mathrm{Sp}_n(1)$.

3. Modular forms

3.1. Let f be a cuspidal newform of level $\Gamma_1(N)$, nebentypus χ_f, and weight $k \geq 2$. Suppose that for each prime $p \nmid N$ we have $T_p f = a_p f$. Then each a_p is an algebraic integer, and the set $\{a_p\}$ generates a number field K_f with ring of integers $\mathcal{O}_f$. We will view K_f as a subfield of $\mathbb{C}$. It is known that K_f contains the image of χ_f. For each place $\lambda | l$ of $\mathcal{O}_f$ there is a continuous representation

$$\rho_{f,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(K_{f,\lambda})$$

which is determined up to isomorphism by the property that for all $p \nmid Nl$, $\rho_{f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is unramified, and the characteristic polynomial of $\rho_{f,\lambda}(\mathrm{Frob}_p)$ is $X^2 - a_p X + p^{k-1}\chi_f(p)$ (where Frob_p is a choice of a geometric Frobenius element at p).
Assume from now on that f is not of CM type.

Definition 3.1. Let λ be a prime of $\mathcal{O}_f$ lying over a rational prime l. Then we say that f is ordinary at λ if $\lambda \nmid a_l$. We say that f is ordinary at l if it is ordinary at λ for some $\lambda | l$.

Lemma 3.2. *If $k = 2$ or 3, then the set of primes l such that f is ordinary at l has density one.*

Proof. The proof is based on an argument of Wiles (see the final lemma of [Wil88]). Let S be the finite set of primes which either divide N or which are ramified in $\mathcal{O}_f$. Suppose that f is not ordinary at $p \notin S$. By definition we have that $\lambda | a_p$ for each prime λ of $\mathcal{O}_f$ lying over p. Since p is unramified in $\mathcal{O}_f$, $(p) = \prod_{\lambda|p} \lambda$, so $p | a_p$. Write $a_p = p b_p$ with $b_p \in \mathcal{O}_f$.
Since $p \nmid N$, the Weil bounds (that is, the Ramanujan-Petersson conjecture) tell us that for each embedding $\iota : K_f \hookrightarrow \mathbb{C}$ we have $|\iota(a_p)| \leq 2p^{(k-1)/2}$. Since $k \leq 3$, this implies that $|\iota(b_p)| \leq 2$ for all ι. Let T be the set of $y \in \mathcal{O}_f$ such that $|\iota(y)| \leq 2$ for all ι. This is a finite set, because one can bound the absolute values of the coefficients of the characteristic polynomial of such a y.
From the above analysis, it is sufficient to prove that for each $y \in T$, the set of primes p for which $a_p = py$ has density zero. However, by Corollaire 1 to Théorème 15 of [Ser81], the number of primes $p \leq x$ for which $a_p = py$ is

$O(x/(\log x)^{5/4-\delta})$ for any $\delta > 0$, which immediately shows that the density of such primes is zero, as required. □

The following result is well known, and follows from, for example, [Sch90] and Theorem 2 of [Wil88].

Lemma 3.3. *If f is ordinary at a place $\lambda|l$ of $\mathcal{O}_f$, and $l \nmid N$, then the Galois representation $\rho_{f,\lambda}$ is crystalline, and furthermore it is ordinary; that is,*

$$\rho_{f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_l/\mathbb{Q}_l)} \cong \begin{pmatrix} \psi_1 & * \\ 0 & \psi_2\epsilon^{1-k} \end{pmatrix}$$

where ψ_1 and ψ_2 are unramified characters of finite order. In addition, ψ_1 takes Frob_l *to the unit root of $X^2 - a_l X + \chi_f(l)l^{k-1}$.*

3.2. Let $\overline{\rho}_{f,\lambda}$ denote the semisimplification of the reduction mod λ of $\rho_{f,\lambda}$; this makes sense because $\rho_{f,\lambda}$ may be conjugated to take values in $\mathrm{GL}_2(\mathcal{O}_{f,\lambda})$, and it is independent of the choice of lattice. It is valued in $\mathrm{GL}_2(k_{f,\lambda})$, where $k_{f,\lambda}$ is the residue field of $K_{f,\lambda}$.

Definition 3.4. We say that $\overline{\rho}_{f,\lambda}$ has *large image* if

$$\mathrm{SL}_2(k) \subset \overline{\rho}_{f,\lambda}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})) \subset k_{f,\lambda}^\times \mathrm{GL}_2(k)$$

for some subfield k of $k_{f,\lambda}$.

We will need to know that the residual Galois representations $\overline{\rho}_{f,\lambda}$ frequently have large image. The following result is essentially due to Ribet (see [Rib75], which treats the case $N = 1$; for a concrete reference, which also proves the corresponding result for Hilbert modular forms, see [Dim05]).

Lemma 3.5. *For all but finitely many primes λ of $\mathcal{O}_f$, $\overline{\rho}_{f,\lambda}$ has large image.*

3.3. We let $\pi(f)$ be the automorphic representation of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$ corresponding to f, normalised so that $\pi(f)$ is RAESDC of weight $(k-2, 0)$ (it is essentially self dual because

$$\pi(f)^\vee \cong \chi\pi(f)$$

where $\chi = |\cdot|^{k-2}\chi_f^{-1}$). Let $\lambda|l$ be a place of $\mathcal{O}_f$, and choose an isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$ and a compatible embedding $K_{f,\lambda} \hookrightarrow \overline{\mathbb{Q}}_l$; that is, an embedding such that the diagram

$$\begin{tikzcd} K_f \arrow[r] \arrow[d] & \mathbb{C} \\ K_{f,\lambda} \arrow[r] & \overline{\mathbb{Q}}_l \arrow[u, "\iota"'] \end{tikzcd}$$

commutes. Assume that $\pi_{f,v}$ is square integrable for some finite place v. Then by Proposition 2.2 there is a Galois representation

$$r_{l,\iota}(\pi(f)) : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{Q}}_l)$$

associated to π_f, and it follows from the definitions that

$$r_{l,\iota}(\pi(f)) \cong \rho_{f,\lambda} \otimes_{K_{f,\lambda}} \overline{\mathbb{Q}}_l.$$

DEFINITION 3.6. We say that f is Steinberg at a prime q if $\pi(f)_q$ is an unramified twist of the Steinberg representation.

DEFINITION 3.7. We say that f is potentially Steinberg at a prime q if $\pi(f)_q$ is a (possibly ramified) twist of the Steinberg representation.

Note that if f is (potentially) Steinberg at q for some q then it is not CM. Note also that if f is potentially Steinberg at q then there is a Dirichlet character θ such that $f \otimes \theta$ is Steinberg at q.

4. POTENTIAL AUTOMORPHY IN WEIGHT 0

4.1. Let l be an odd prime, and let f be a modular form of weight $2 \leq k < l$ and level N, $l \nmid N$. Assume that f is Steinberg at q. Suppose that $\lambda | l$ is a place of $\mathcal{O}_f$ such that f is ordinary at λ. Assume that $\overline{\rho}_{f,\lambda}$ is absolutely irreducible. By Lemma 3.3 we have

$$\overline{\rho}_{f,\lambda}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_l/\mathbb{Q}_l)} \cong \begin{pmatrix} \overline{\psi}_1 & * \\ 0 & \overline{\psi}_2 \epsilon^{1-k} \end{pmatrix}$$

where $\overline{\psi}_1$ and $\overline{\psi}_2$ are unramified characters. We wish to prove that the symmetric powers of $\overline{\rho}_{f,\lambda}$ are potentially automorphic of some weight. To do so, we use a potential modularity argument to realise $\overline{\rho}_{f,\lambda}$ geometrically, and then appeal to the results of [HSBT09].

The potential modularity result that we need is almost proved in [Tay02]; the one missing ingredient is that we wish to preserve the condition of being Steinberg at q. This is, however, easily arranged, and rather than repeating all of the arguments of [Tay02], we simply indicate the modifications required.

We begin by recalling some definitions from [Tay02]. Let N be a totally real field. Then an N-HBAV over a field K is a triple (A, i, j) where

- A/K is an abelian variety of dimension $[N : \mathbb{Q}]$,
- $i : \mathcal{O}_N \hookrightarrow \mathrm{End}(A/K)$, and
- $j : \mathcal{O}_N^+ \xrightarrow{\sim} \mathcal{P}(A, i)$ is an isomorphism of ordered invertible $\mathcal{O}_N$-modules.

For the definitions of ordered invertible $\mathcal{O}_N$-modules and of $\mathcal{O}_N^+$ and $\mathcal{P}(A, i)$, see page 133 of [Tay02].

Choose a totally real quadratic field F in which l is inert and q is unramified and which is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$, a finite extension $k/k_{f,\lambda}$ and a character $\overline{\theta} : \mathrm{Gal}(\overline{F}/F) \to k^\times$ which is unramified at q such that

$$\det \overline{\rho}_{f,\lambda}|_{\mathrm{Gal}(\overline{F}/F)} = \epsilon^{-1}\overline{\theta}^{-2}$$

and $(\overline{\rho}_{f,\lambda}|_{\mathrm{Gal}(\overline{F}/F)} \otimes \overline{\theta})(\mathrm{Frob}_w)$ has eigenvalues 1, $\#k(w)$, where $w|q$ is a place of F. This is possible as the obstruction to taking a square root of a character is in the 2-part of the Brauer group, and because any class in the Brauer group of a local field splits over an unramified extension. Let $\overline{\rho} = \overline{\rho}_{f,\lambda}|_{\mathrm{Gal}(\overline{F}/F)} \otimes \overline{\theta}$:

$\mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_2(k)$, so that $\det \overline{\rho} = \epsilon^{-1}$. If x is the place of F lying over l, then we may write (for some character $\overline{\chi}_x$ of $\mathrm{Gal}(\overline{F_x}/F_x)$)

$$\overline{\rho}|_{\mathrm{Gal}(\overline{F_x}/F_x)} \cong \begin{pmatrix} \overline{\chi}_x^{-1} & * \\ 0 & \overline{\chi}_x\epsilon^{-1} \end{pmatrix}$$

with $\overline{\chi}_x^2|_{I_x} = \epsilon^{2-k}$.

Theorem 4.1. *There is a finite totally real Galois extension E/F which is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$ and in which the unique prime of F dividing l splits completely, a totally real field N, a place $\lambda'|l$ of N, a place $v_q|q$ of E, and an N-HBAV $(A,i,j)/E$ with potentially good reduction at all places dividing l such that*

- *the representation of $\mathrm{Gal}(\bar{E}/E)$ on $A[\lambda']$ is equivalent to $(\overline{\rho}|_{\mathrm{Gal}(\bar{E}/E)})^\vee$,*
- *at each place $x|l$ of E, the action of $\mathrm{Gal}(\overline{E_x}/E_x)$ on $T_{\lambda'}A\otimes\mathbb{Q}_l$ is of the form*
$$\begin{pmatrix} \chi_x^{-1}\epsilon & * \\ 0 & \chi_x \end{pmatrix}$$
with χ_x a tamely ramified lift of $\overline{\chi}_x$, and
- *A has multiplicative reduction at v_q.*

Proof. As remarked above, this is essentially proved in [Tay02]. Indeed, if $k>2$ then with the exception of the fact that E can be chosen to be linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$, and the claim that A can be chosen to have multiplicative reduction at some place over q, the result is obtained on page 136 of [Tay02] (the existence of A with $A[\lambda']$ equivalent to $(\overline{\rho}|_{\mathrm{Gal}(\bar{E}/E)})^\vee$ is established in the second paragraph on that page, and the form of the action of $\mathrm{Gal}(\overline{E_x}/E_x)$ for $x|l$ follows from Lemma 1.5 of *loc. cit.*).

We now indicate the modifications needed to the arguments of [Tay02] to obtain the slight strengthening that we require. Suppose firstly that $k>2$. Rather than employing the theorem of Moret-Bailly stated as Theorem G of [Tay02], we use the variant given in Proposition 2.1 of [HSBT09]. This immediately allows us to assume that E is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$, so we only need to ensure that A has multiplicative reduction at some place dividing q. Let X be the moduli space defined in the first paragraph of page 136 of [Tay02]. Let v be a place of F lying over q. It is enough to check that there is a non-empty open subset Ω_v of $X(F_v)$ such that for each point of Ω_v, the corresponding N-HBAV has multiplicative reduction. Let Ω_v denote the set of *all* points of $X(F_v)$ such that the corresponding N-HBAV has multiplicative reduction; this is an open subset of $X(F_v)$, and it is non-empty (by the assumptions on $\bar{\theta}$ at places of F dividing q, and the assumption that $\pi(f)$ is an unramified twist of the Steinberg representation, we see that $\overline{\rho}(\mathrm{Frob}_v)$ has eigenvalues 1 and $\#k(v)$, and is congruent to a Galois representation attached to an unramified twist of a Steinberg representation, so any N-HBAV with multiplicative reduction suffices), as required.

If $k = 2$, then the only additional argument needed is one to ensure that if $\overline{\chi}_x^2 = 1$, then the abelian variety can be chosen to have good reduction rather than multiplicative reduction. This follows easily from the fact that $\overline{\rho}|_{\mathrm{Gal}(\overline{F_x}/F_x)}$ is finite flat (cf. the proof of Theorem 2.1 of [KW08], which establishes a very similar result). □

Let M be a totally real field, and let $(A, i, j)/M$ be an N-HBAV. Fix an embedding $N \subset \mathbb{R}$. We recall some definitions from section 4 of [HSBT09]. For each finite place v of M there is a two dimensional Weil-Deligne representation $\mathrm{WD}_v(A, i)$ defined over $\overline{N}$ such that if $\mathfrak{p}$ is a place of N of residue characteristic p different from the residue characteristic of v, we have

$$\mathrm{WD}(H^1(A \times \overline{M}, \mathbb{Q}_p)|_{\mathrm{Gal}(\overline{M}_v/M_v)} \otimes_{N_p} \overline{N}_{\mathfrak{p}}) \cong \mathrm{WD}_v(A, i) \otimes_{\overline{N}} \overline{N}_{\mathfrak{p}}.$$

DEFINITION 4.2. We say that $\mathrm{Sym}^m A$ is automorphic of type $\{\rho_v\}_{v \in S}$ if there is an RAESDC representation π of $\mathrm{GL}_{m+1}(\mathbb{A}_M)$ of weight 0 and type $\{\rho_v\}_{v \in S}$ such that for all finite places v of M,

$$\mathrm{rec}(\pi_v)|\,\mathrm{Art}_{M_v}^{-1}|^{-m/2} = \mathrm{Sym}^m \mathrm{WD}_v(A, i).$$

THEOREM 4.3. *Let E, A be as in the statement of Theorem 4.1. Let $\mathcal{N}$ be a finite set of even positive integers. Then there is a finite Galois totally real extension F'/E and a place $w_q|q$ of F' such that*

- *for any $n \in \mathcal{N}$, $\mathrm{Sym}^{n-1} A$ is automorphic over F' of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$,*
- *The primes of E dividing l are unramified in F', and*
- *F' is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$.*

Proof. This is essentially Theorem 4.1 of [HSBT09]. In particular, the proof in [HSBT09] establishes that there is a Galois totally real extension F'/E, and a place w_q of F' lying over q such that for any $n \in \mathcal{N}$, $\mathrm{Sym}^{n-1} A$ is automorphic over F' of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$. Note that the l used in their argument is *not* the l used here. To complete the proof, we need to establish that it is possible to obtain an F' in which l is unramified, and which is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$. The latter point causes no difficulty, but the first point requires some minor modifications of the arguments of [HSBT09]. We now outline the necessary changes.

To aid comparison to [HSBT09], for the rest of this proof we will refer to our l as s; all references to l will be to primes of that name in the proofs of various theorems in [HSBT09]. We begin by choosing a finite solvable totally real extension L of E, linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$, such that the base change of A to L has good reduction at all places dividing s. Choose a prime l as in the proof of Theorem 4.1 of [HSBT09]. We then apply a slight modification of Theorem 4.2 of *loc.cit.*, with the conclusion strengthened to include the hypothesis that s is unramified in F'. To prove this, in the proof of Theorem 4.2 of *loc.cit.*, note that $F_1 = E$. Choose all auxiliary primes not to divide s. Rather than constructing a moduli space X_W over E, construct the analogous space over L, and consider the restriction of scalars $Y = \mathrm{Res}_{L/E}(X_W)$. Applying Proposition 2.1 of [HSBT09] to Y, rather than X_W, we may find a finite totally real Galois extension $F^{(1)}/E$ in which s is unramified, such that Y has an $F^{(1)}$-point. Furthermore, we may assume that $F^{(1)}$ is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$. Note that an $F^{(1)}$-point of Y corresponds to an $F^{(1)}L$-point of X_W. We now make a similar modification to the proof of Theorem 3.1 of [HSBT09], replacing the schemes T_{W_i} over F with $\mathrm{Res}_{LF/F} T_{W_i}$. We conclude that there is a finite Galois totally real extension F'/E in which s is unramified, which is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$, such that for any $n \in \mathcal{N}$, $\mathrm{Sym}^{n-1} A$ is automorphic over $F'L$ of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$. Since the extension $F'L/F'$ is solvable, it follows from solvable base change (e.g. Lemma 1.3 of [BLGHT09]) that in fact for any $n \in \mathcal{N}$, $\mathrm{Sym}^{n-1} A$ is automorphic over F' of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$, as required. $\square$

We may now twist $\overline{\rho}$ by $\overline{\theta}^{-1}$ in order to deduce results about $\overline{\rho}_{f,\lambda}$. Let N and λ be as in the statement of Theorem 4.1. Fix an embedding $N_{\lambda'} \hookrightarrow \overline{\mathbb{Q}}_l$. Let θ be the Teichmüller lift of $\overline{\theta}$, and let ρ_n denote the action of $\mathrm{Gal}(\bar{E}/E)$ on

$$\mathrm{Sym}^{n-1}(H^1(A \times \overline{E}, \mathbb{Q}_l) \otimes_{N_l} N_{\lambda'} \otimes \theta^{-1}) \otimes_{N'_\lambda} \overline{\mathbb{Q}}_l.$$

By construction, ρ_n is a lift of $\mathrm{Sym}^{n-1} \overline{\rho}_{f,\lambda}|_{\mathrm{Gal}(\overline{E}/E)} \otimes_{k_{f,\lambda}} \bar{\mathbb{F}}_l$ (where the embedding $k_{f,\lambda} \hookrightarrow \overline{\mathbb{F}}_l$ is determined by the embedding $k \hookrightarrow \overline{\mathbb{F}}_l$ induced by the embedding $N_{\lambda'} \hookrightarrow \overline{\mathbb{Q}}_l$). Note also that (again by construction) at each place $x|l$ of E,

$$\rho_2|_{\mathrm{Gal}(\overline{E_x}/E_x)} \cong \begin{pmatrix} \psi_1 & * \\ 0 & \psi_2 \omega^{2-k} \epsilon^{-1} \end{pmatrix}$$

with ψ_1, ψ_2 unramified lifts of $\overline{\psi}_1|_{\mathrm{Gal}(\overline{E_x}/E_x)}$ and $\overline{\psi}_2|_{\mathrm{Gal}(\overline{E_x}/E_x)}$ respectively, and ω the Teichmüller lift of ϵ.

Corollary 4.4. *Let $\mathcal{N}$ be a finite set of even positive integers. Then there is a Galois totally real extension F'/E and a place $w_q|q$ of F' such that*

- *for any $n \in \mathcal{N}$, $\rho_n|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F')}$ is automorphic of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$,*
- *every prime of E dividing l is unramified in F' (so that l is unramified in F'), and*
- *F' is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$.*

Let $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$, and for $n \in \mathcal{N}$ let π_n be the RAESDC representation of $\mathrm{GL}_n(\mathbb{A}_{F'})$ with $r_{l,\iota}(\pi_n) \cong \rho_n|_{\mathrm{Gal}(\overline{\mathbb{Q}}_l/F')}$. If $k = 2$ then $\pi_{n,x}$ is unramified for each $x|l$, and if $k > 2$ then for each place $x|l$ of F', $\pi_{n,x}$ is a principal series representation $\text{n-Ind}_{B_n(F'_x)}^{\mathrm{GL}_n(F'_x)}(\chi_1, \ldots, \chi_n)$ with $\iota^{-1}\chi_i \circ \mathrm{Art}_{F'_x}^{-1}|_{I_x} = \omega^{(i-1)(2-k)}$ and $v_l(\iota^{-1}\chi_i(l)) = [F'_x : \mathbb{Q}_l]\left(i - 1 + \frac{1-n}{2}\right)$, where v_l is the l-adic valuation on $\overline{\mathbb{Q}}_l$ with $v_l(l) = 1$.

Proof. This is a straightforward consequence of Theorem 4.3. The only part that needs to be checked is the assertion about the form of $\pi_{n,x}$ for $x|l$ when $k > 2$. Without loss of generality, we may assume that $2 \in \mathcal{N}$. Note firstly that any principal series representation of the given form is irreducible, so that we need only check that

$$\iota^{-1}\operatorname{rec}(\pi_{n,x}) = \bigoplus_{i=1}^{n} \omega^{(i-1)(2-k)}\alpha_i,$$

where α_i is an unramified character with $v_l(\alpha_i(l)) = [F'_x : \mathbb{Q}_l]\left(i - 1 + \frac{1-n}{2}\right)$. By Definition 4.2 and Theorem 4.3 we see that $\operatorname{rec}(\pi_{n,x}) = \operatorname{Sym}^{n-1}\operatorname{rec}(\pi_{2,x})$, so it suffices to establish the result in the case $n = 2$, or rather (because of the compatibility of rec with twisting) it suffices to check the corresponding result for $\mathrm{WD}_v(A, i)$ at places $v|l$. This is now an immediate consequence of local-global compaitibility, and follows at once from, for example, Lemma B.4.1 of [CDT99], together with the computations of the Weil-Deligne representations associated to characters in section B.2 of *loc. cit.*

□

5. Changing weight

5.1. We now explain how to deduce from the results of the previous section that $\operatorname{Sym}^n \overline{\rho}_{f,\lambda}$ is potentially automorphic of the correct weight (that is, the weight of the conjectural automorphic representation corresponding to $\operatorname{Sym}^n \rho_{f,\lambda}$), rather than potentially automorphic of weight 0. We accomplish this as a basic consequence of Hida theory; note that we simply need a congruence, rather than a result about families, and the result follows from a straightforward combinatorial argument. This result is certainly known to the experts, but as we have been unable to find a reference which provides the precise result we need, we present a proof in the spirit of the arguments of [Tay88].

5.2. For each n-tuple of integers $a = (a_1, \dots, a_n)$ with $a_1 \geq \cdots \geq a_n$ there is an irreducible representation of the algebraic group GL_n defined over $\mathbb{Q}_l$, with highest weight (with respect to the Borel subgroup of upper-triangular matrices) given by

$$\operatorname{diag}(t_1, \dots, t_n) \mapsto \prod_{i=1}^{n} t_i^{a_i}.$$

We will need an explicit model of this representation, for which we follow section 2 of [Che04].

Let K be an algebraic extension of $\mathbb{Q}_l$, N the subgroup of $\mathrm{GL}_n(K)$ consisting of upper triangular unipotent matrices, $\overline{N}$ the subgroup of lower triangular unipotent matrices, and T the subgroup of diagonal matrices. Let $R := K[\mathrm{GL}_n] = K[\{X_{i,j}\}_{1\leq i,j\leq n}, \det(X_{i,j})^{-1}]$. We have commuting natural actions of $\mathrm{GL}_n(K)$ on R by left and right multiplication. For an element

$g \in \mathrm{GL}_n(K)$ we denote these actions by g_l and g_r respectively, so that if we let M denote the matrix $(X_{i,j})_{i,j} \in M_n(R)$, we have

$$(g_l.X)_{i,j} = g^{-1}M$$

and

$$(g_r.X)_{i,j} = Mg.$$

If $(t_1, \ldots, t_n) \in \mathbb{Z}^n$, we say that an element $f \in R$ is of left weight t (respectively of right weight t) if for all $d \in T$ we have $d_l f = t^{-1}(d)f$ (respectively $d_r f = t(d)f$) where

$$t(\mathrm{diag}(x_1, \ldots, x_n)) = \prod_{i=1}^{n} x_i^{t_i}.$$

For each $1 \leq i \leq n$ and each i-tuple $j = (j_1, \ldots, j_i)$, $1 \leq j_1 < \cdots < j_i \leq n$, we let $Y_{i,j}$ be the minor of order i of M obtained by taking the entries from the first i rows and columns $j_1,\ldots,j_i$. Let $R^{\overline{N}}$ denote the subalgebra of R of elements fixed by the g_l-action of $\overline{N}$; it is easy to check that $Y_{i,j} \in R^{\overline{N}}$. Because T normalises $\overline{N}$ it acts on $R^{\overline{N}}$ on the left, and we let $R_t^{\overline{N}}$ be the sub K-vector space of elements of left weight t; this has a natural action of $\mathrm{GL}_n(K)$ induced by g_r.

PROPOSITION 5.1. *Suppose that $t_1 \geq \cdots \geq t_n$. Then $R_t^{\overline{N}}$ is a model of the irreducible algebraic representation of $\mathrm{GL}_n(K)$ of highest weight t. Furthermore, it is generated as a K-vector space by the monomials in $Y_{i,j}$ of left weight t, and a highest weight vector is given by the unique monomial in $Y_{i,j}$ of left and right weight t.*

Proof. This follows from Proposition 2.2.1 of [Che04]. □

Assume that in fact $t_1 \geq \cdots \geq t_n \geq 0$, and let X_t denote the free $\mathcal{O}_K$-module with basis the monomials in $Y_{i,j}$ of left weight t. By Proposition 5.1, X_t is a $\mathrm{GL}_n(\mathcal{O}_K)$-stable lattice in $R_t^{\overline{N}}$. Let T^+ be the submonoid of T consisting of elements of the form

$$\mathrm{diag}(l^{b_1}, \ldots, l^{b_n})$$

with $b_1 \geq \cdots \geq b_n \geq 0$; then X_t is certainly also stable under the action of T^+. Let $\alpha = \mathrm{diag}(l^{b_1}, \ldots, l^{b_n}) \in T^+$. We wish to determine the action of α on X_t.

LEMMA 5.2. *If $Y \in X_t$ is a monomial in the $Y_{i,j}$, then $\alpha(Y) \subset l^{\sum_{i=1}^n b_i t_{n+1-i}} X_t$. If in fact $b_1 > \cdots > b_n$ then $\alpha(Y) \subset l^{1+\sum_{i=1}^n b_i t_{n+1-i}} X_t$ unless Y is the unique lowest weight vector.*

Proof. If Y has (right) weight $(v_1, \ldots, v_n)$, then $\alpha(Y) = l^{\sum_{i=1}^n b_i v_i} Y$. The unique lowest weight vector has weight $(t_n, \ldots, t_1)$, so it suffices to prove that for any other Y of weight $(v_1, \ldots, v_n)$ which occurs in $R_t^{\overline{N}}$, the quantity $\sum_{i=1}^n b_i v_i$ is at least as large, and is strictly greater if $b_1 > \cdots > b_n$. However, by standard weight theory we know that we may obtain $(v_1, \ldots, v_n)$ from $(t_n, \ldots, t_1)$ by successively adding vectors of the form

$(0,\dots,1,0\dots,0,-1,0,\dots,0)$, and it is clear that the addition of such a vector does not decrease the sum, and in fact increases it if $b_1 > \cdots > b_n$, as required. $\square$

We define a new action of T^+ on X_t, which we denote by $\cdot_{twist}$, by multiplying the natural action of $\mathrm{diag}(l^{b_1},\dots,l^{b_n})$ by $l^{-\sum_{i=1}^n b_i t_{n+1-i}}$; this is legitimate by Lemma 5.2.

5.3. Fix for the rest of this section a choice of isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. Assume for the rest of this section that F' is a totally real field in which each l is unramified, and π' is an RAESDC representation of $\mathrm{GL}_n(\mathbb{A}_{F'})$ of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$ for some place $w_q|q$ of F', with $(\pi')^\vee = \chi\pi'$. Suppose furthermore that there is an integer $k > 2$ such that

- for each place $x|l$, π'_x is a principal series $\text{n-Ind}_{B_n(F_x)}^{\mathrm{GL}_n(F_x)}(\chi_1,\dots,\chi_n)$ with $v_l(\iota^{-1}\chi_i(l)) = [F'_x : \mathbb{Q}_l]\left(i-1+\frac{1-n}{2}\right)$ and $\iota^{-1}\chi_i \circ \mathrm{Art}_{F_x}^{-1}|_{I_x} = \omega^{(i-1)(2-k)}$.

(See Corollary 4.4 for an example of such a representation.) We transfer to a unitary group, following section 3.3 of [CHT08]. Firstly, we make a quartic totally real Galois extension F/F', linearly disjoint from $\overline{\mathbb{Q}}^{\ker \overline{r}_{l,\iota}(\pi)}$ over $\mathbb{Q}$, such that w_q and all primes dividing l split in F. Let $S(B)$ be the set of places of F lying over w_q. Let E be a imaginary quadratic field in which l and q split, such that E is linearly disjoint from $\overline{\mathbb{Q}}^{\ker \overline{r}_{l,\iota}(\pi)}$ over $\mathbb{Q}$. Let $M = FE$. Let c denote the nontrivial element of $\mathrm{Gal}(M/F)$. Let S_l denote the places of F dividing l, and let $\tilde{S}_l$ denote a set of places of M dividing l such that the natural map $\tilde{S}_l \to S_l$ is a bijection. If $v|l$ is a place of F then we write $\tilde{v}$ for the corresponding place in $\tilde{S}_l$.

Lemma 5.3. *There is a finite order character $\phi : M^\times\backslash\mathbb{A}_M^\times \to \mathbb{C}^\times$ such that*

- *$\phi \circ N_{M/F} = \chi \circ N_{M/F}$, and*
- *ϕ is unramified at all places lying over $S(B)$ and at all places in $\tilde{S}_l$.*

Proof. By Lemma 4.1.1 of [CHT08] (or more properly its proof, which shows that the character produced may be arranged to have finite order) there is a finite order character $\psi : M^\times\backslash\mathbb{A}_M^\times \to \mathbb{C}^\times$ such that for each $v \in S_l$, $\psi|_{M_{\tilde{v}}{}^\times} = 1$ and $\psi|_{M_{c\tilde{v}}{}^\times} = \chi|_{F_v^\times}$, and such that ψ is unramified at each place in $S(B)$. It now suffices to prove the result for the character $\chi(\psi|_{\mathbb{A}_F^\times})^{-1}$, which is unramified at $S(B) \cup S_l$, and the result now follows from Lemma 4.1.4 of [CHT08]. $\square$

Now let $\pi = \pi'_M \otimes \phi$, which is an RACSDC representation of $\mathrm{GL}_n(\mathbb{A}_M)$, satisfying:

- π has weight 0.
- π has type $\{\mathrm{Sp}_n(1)\}_{w|w_q}$.
- for each place $x \in \tilde{S}_l$, π_x is a principal series $\text{n-Ind}_{B_n(M_x)}^{\mathrm{GL}_n(M_x)}(\chi_1,\dots,\chi_n)$ with $v_l(\iota^{-1}\chi_i(l)) = [F'_x : \mathbb{Q}_l]\left(i-1+\frac{1-n}{2}\right)$ and $\iota^{-1}\chi_i \circ \mathrm{Art}_{M_x}^{-1}|_{I_x} = \omega^{(i-1)(2-k)}$ with $k > 2$.

5.4. Choose a division algebra B with centre M such that

- B splits at all places not dividing a place in $S(B)$.
- If w is a place of M lying over a place in $S(B)$, then B_w is a division algebra.
- $\dim_M B = n^2$.
- $B^{op} \cong B \otimes_{M,c} M$.

For any involution $\ddagger$ on B with $\ddagger|_M = c$, we may define a reductive algebraic group $G_\ddagger/F$ by

$$G_\ddagger(R) = \{g \in B \otimes_F R : g^{\ddagger \otimes 1} g = 1\}$$

for any F-algebra R. Because $[F : \mathbb{Q}]$ is divisible by 4 and $\#S(B)$ is even, we may (by the argument used to prove Lemma 1.7.1 of [HT01]) choose $\ddagger$ such that

- If $v \notin S(B)$ is a finite place of F then $G_\ddagger(F_v)$ is quasi-split, and
- If $v|\infty$, $G_\ddagger(F_v) \cong U(n)$.

Fix such a choice of $\ddagger$, and write G for $G_\ddagger$. We wish to work with algebraic modular forms on G; in order to do so, we need an integral model for G. We obtain such a model by fixing an order $\mathcal{O}_B$ in B such that $\mathcal{O}_B^\ddagger = \mathcal{O}_B$ and $\mathcal{O}_{B,w}$ is a maximal order for all primes w which are split over M (see section 3.3 of [CHT08] for a proof that such an order exists). We now regard G as an algebraic group over $\mathcal{O}_F$, by defining

$$G(R) = \{g \in \mathcal{O}_B \otimes_{\mathcal{O}_F} R : g^{\ddagger \otimes 1} g = 1\}$$

for all $\mathcal{O}_F$-algebras R.
We may identify G with GL_n at places not in $S(B)$ which split in M in the following way. Let $v \notin S(B)$ be a place of F which splits in M. Choose an isomorphism $i_v : \mathcal{O}_{B,v} \xrightarrow{\sim} M_n(\mathcal{O}_{M_v})$ such that $i_v(x^\ddagger) = {}^t i_v(x)^c$ (where ^{t} denotes matrix transposition). Choosing a prime $w|v$ of M gives an isomorphism

$$\begin{aligned} i_w : G(F_v) &\xrightarrow{\sim} \mathrm{GL}_n(M_w) \\ i_v^{-1}(x, {}^t x^{-c}) &\mapsto x. \end{aligned}$$

This identification satisfies $i_w G(\mathcal{O}_{F,v}) = \mathrm{GL}_n(\mathcal{O}_{M,w})$. Similarly, if $v \in S(B)$ then v splits in M, and if $w|v$ then we obtain an isomorphism

$$i_w : G(F_v) \xrightarrow{\sim} B_w^\times$$

with $i_w G(\mathcal{O}_{F,v}) = \mathcal{O}_{B,w}^\times$.
Now let $K = \overline{\mathbb{Q}}_l$. Write $\mathcal{O}$ for the ring of integers of K, and k for the residue field $\overline{\mathbb{F}}_l$.
Let $I_l = \mathrm{Hom}(F, K)$, and let $\tilde{I}_l$ be the subset of elements of $\mathrm{Hom}(M, K)$ such that the induced place of M is in $\tilde{S}_l$. Let $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,K)}$; we assume that

- $a_{\tau,1} \geq \cdots \geq a_{\tau,n} \geq 0$ if $\tau \in \tilde{I}_l$, and
- $a_{\tau c,i} = -a_{\tau,n+1-i}$.

Consider the constructions of section 5.2 applied to our choice of K. Then we have an $\mathcal{O}$-module

$$Y_a = \otimes_{\tau \in \tilde{I}_l} X_{a_\tau}$$

which has a natural action of $G(\mathcal{O}_{F,l})$, where $g \in G(\mathcal{O}_{F,l})$ acts on X_{a_τ} by $\tau(i_\tau g_\tau)$. From now on, if $v|l$ is a place of F, we will identify $G(\mathcal{O}_{F_v})$ with $\mathrm{GL}_n(\mathcal{O}_{M_{\tilde{v}}})$ via the map $i_{\tilde{v}}$ without comment.

We say that an open compact subgroup $U \subset G(A_F^\infty)$ is sufficiently small if for some place v of F the projection of U to $G(F_v)$ contains no nontrivial elements of finite order. Assume from now on that U is sufficiently small, and in addition that we may write $U = \prod_v U_v$, $U_v \subset G(\mathcal{O}_{F_v})$, such that

- if $v \in S(B)$ and $w|v$ is a place of M, then $i_w(U_v) = \mathcal{O}_{B,w}^\times$, and
- if $v|l$ then U_v is the Iwahori subgroup of matrices which are upper-triangular mod l.

If $v|l$, let U'_v denote the pro-l subgroup of U_v corresponding to the group of matrices which are (upper-triangular) unipotent mod l, and let

$$\chi_v : U_v/U'_v \to \mathcal{O}^\times$$

be a character. Let $\chi = \otimes \chi_v : \prod_{v|l} U_v \to \mathcal{O}^\times$, and write

$$Y_{a,\chi} = Y_a \otimes_{\mathcal{O}} \chi,$$

a $\prod_{v|l} U_v$-module.

Let A be an $\mathcal{O}$-algebra. Then we define the space of algebraic modular forms

$$S_{a,\chi}(U, A)$$

to be the space of functions

$$f : G(F)\backslash G(\mathbb{A}_F^\infty) \to A \otimes_{\mathcal{O}} Y_{a,\chi}$$

satisfying

$$f(gu) = u^{-1} f(g)$$

for all $u \in U$, $g \in G(\mathbb{A}_F^\infty)$, where the action of U on $A \otimes_{\mathcal{O}} Y_{a,\chi}$ is inherited from the action of $\prod_{v|l} U_v$ on $Y_{a,\chi}$. Note that because U is sufficiently small we have

$$S_{a,\chi}(U, A) = S_{a,\chi}(U, \mathcal{O}) \otimes_{\mathcal{O}} A.$$

More generally, if V is any U''-module with U'' a sufficiently small compact open subgroup, we define the space of algebraic modular forms

$$S(U'', V)$$

to be the space of functions

$$f : G(F)\backslash G(\mathbb{A}_F^\infty) \to V$$

satisfying

$$f(gu) = u^{-1} f(g)$$

for all $u \in U''$, $g \in G(\mathbb{A}_F^\infty)$.

Let T_l^+ denote the monoid of elements of $G(\mathbb{A}_F^\infty)$ which are trivial outside of places dividing l, and at places dividing l correspond to matrices

$\mathrm{diag}(l^{b_1},\ldots,l^{b_n})$ with $b_1 \geq \cdots \geq b_n \geq 0$. In addition to the action of U on $Y_{a,\chi}$, we can also allow T_l^+ to act. We define the action of T_l^+ via the action $\cdot_{twist}$ on X_t defined above. This gives us an action of the monoid $\langle U, T_l^+\rangle$ on $Y_{a,\chi}$. Now suppose that g is an element of $G(\mathbb{A}_F^\infty)$ with either $g_l \in G(\mathcal{O}_{F,l})$ or $g \in T_l^+$; then we write

$$UgU = \coprod_i g_i U,$$

a finite union of cosets, and define a linear map

$$[UgU] : S_{a,\chi}(U,A) \to S_{a,\chi}(U,A)$$

by

$$([UgU]f)(h) = \sum_i g_i f(hg_i).$$

We now introduce some notation for Hecke algebras. Let v be a place of F which splits in M, and suppose that $v \notin S(B)$ and that $U_v = G(\mathcal{O}_{F_v})$ (so, in particular $v \nmid l$). Suppose that $w|v$ is a place of M, so that we may regard $G(\mathcal{O}_{F_v})$ as $\mathrm{GL}_n(\mathcal{O}_{M_w})$ via i_w. Then we let $T_w^{(j)}$, $1 \leq j \leq n$ denote the Hecke operator given by

$$[U\,\mathrm{diag}(\varpi_w,\ldots,\varpi_w,1,\ldots,1)U]$$

where ϖ_w is a uniformiser of M_w, and there are j occurrences of it in this matrix. We let $\mathbb{T}_{a,\chi}(U,A)$ denote the commutative A-subalgebra of $\mathrm{End}(S_{a,\chi}(U,A))$ generated by the operators $T_w^{(j)}$ and $(T_w^{(n)})^{-1}$ for all w, j as above. Note that $\mathbb{T}_{a,\chi}(U,A)$ commutes with $[UgU]$ for all $g \in T_l^+$. More generally, let $\mathbb{T}(U)$ denote the polynomial ring over $\mathcal{O}$ in the formal variables $T_w^{(j)}$ and $(T_w^n)^{-1}$, which we may think of as acting on $S_{a,\chi}(U,A)$ via the obvious map $\mathbb{T}(U) \to \mathbb{T}_{a,\chi}(U,A)$.

We also wish to consider the Hecke operator $U_l = [UuU]$, where $u \in T_l^+$ has $u_v = \mathrm{diag}(l^{n-1},\ldots,l,1)$ for each $v|l$. As usual, we can define a Hida idempotent

$$e_l = \lim_{n\to\infty} U_l^{n!},$$

which has the property that U_l is invertible on $e_l S_{a,\chi}(U,\mathcal{O})$ and is topologically nilpotent on $(1-e_l)S_{a,\chi}(U,\mathcal{O})$. We write

$$S_{a,\chi}^{ord}(U,A) := e_l S_{a,\chi}(U,A).$$

Let $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,K)}$ be a weight, and let $\chi_a = \otimes_{v|l}\chi_{a,v}$, where $\chi_{a,v} : U_v/U_v' \cong ((\mathcal{O}_{M_v}/\mathfrak{m}_{M_v})^\times)^n \to \mathcal{O}^\times$ is given by the character $(x_1,\ldots,x_n) \mapsto \prod_\tau \prod_i \tau(\tilde{x}_i)^{a_{\tilde{v},n+1-i}}$, where $\tilde{x}_i$ is the Teichmüller lift of x_i, and the product is over the embeddings $\tau \in \tilde{I}_l$ which give rise to v.

The main lemma we require is the following.

LEMMA 5.4. *Let a be a weight. Then there is a $\mathbb{T}(U)$-equivariant isomorphism*

$$S_{a,\chi}^{ord}(U,k) \to S_{0,\chi\chi_a}^{ord}(U,k).$$

Proof. Note firstly that there is a natural projection map j from $Y_{a,\chi}$ to the $\mathcal{O}$-module given by the tensor product $Z_{a,\chi}$ of the lowest weight vectors. This is a map of $\prod_{v|l} U_v$-modules, and by Lemma 5.2 we see that j induces an isomorphism

$$u \cdot_{twist} Y_{a,\chi} \otimes_{\mathcal{O}} k \to u \cdot_{twist} Z_{a,\chi} \otimes_{\mathcal{O}} k.$$

Note also that by definition we have an isomorphism of $\langle U, T_l^+\rangle$-modules $Z_{a,\chi} \to Y_{0,\chi\chi_a}$. It thus suffices to prove that the induced map

$$j : S^{ord}_{a,\chi}(U,k) \to S^{ord}(U, Z_{a,\chi} \otimes_{\mathcal{O}} k) \ (= S^{ord}_{0,\chi\chi_a}(U,k))$$

is an isomorphism.
We claim that there is a diagram

$$\begin{tikzcd}
S_{a,\chi}(U,k) \arrow[r, "j"] & S(U, Z_{a,\chi} \otimes_{\mathcal{O}} k) \arrow[r, "u\cdot_{twist}"] & S(U \cap uUu^{-1}, u \cdot_{twist} Z_{a,\chi} \otimes_{\mathcal{O}} k) \arrow[d, "j^{-1}"] \\
 & S_{a,\chi}(U \cap uUu^{-1}, k) \arrow[ul, "\mathrm{cor}"] & S(U \cap uUu^{-1}, u \cdot_{twist} Y_{a,\chi} \otimes_{\mathcal{O}} k) \arrow[l, "i"']
\end{tikzcd}$$

such that the maps

$$\mathrm{cor} \circ i \circ j^{-1} \circ u \cdot_{twist} \circ j : S_{a,\chi}(U,k) \to S_{a,\chi}(U,k)$$

and

$$j \circ \mathrm{cor} \circ i \circ j^{-1} \circ u \cdot_{twist} : S(U, Z_{a,\chi} \otimes_{\mathcal{O}} k) \to S(U, Z_{a,\chi} \otimes_{\mathcal{O}} k)$$

are both given by U_l. Since U_l is an isomorphism on $S^{ord}_{a,\chi}(U,k)$, the result will follow.
In fact, the construction of the diagram is rather straightforward. The maps j, j^{-1} are just the natural maps on the coefficients (note that both are maps of U-modules). The map $u \cdot_{twist}$ is given by

$$(u \cdot_{twist} f)(h) = u \cdot_{twist} f(hu).$$

The map i is given by the inclusion of U-modules $u \cdot_{twist} Y_{a,\chi} \otimes_{\mathcal{O}} k \hookrightarrow Y_{a,\chi} \otimes_{\mathcal{O}} k$. Finally, the map cor is defined in the following fashion. We may write

$$U = \coprod u_i (U \cap uUu^{-1}),$$

and we define

$$(\mathrm{cor}\, f)(h) = \sum u_i f(hu_i).$$

The claims regarding the compositions of these maps follow immediately from the observation that

$$UuU = \coprod u_i u U.$$

$\square$

5.5. We now recall some results on tamely ramified principal series representations of GL_n from [Roc98]. Let L be a finite extension of $\mathbb{Q}_p$ for some p, and let π_L be an irreducible smooth complex representation of $\mathrm{GL}_n(L)$. Let I denote the Iwahori subgroup of $\mathrm{GL}_n(\mathcal{O}_L)$ consisting of matrices which are upper-triangular mod $\mathfrak{m}_L$, and let I_1 denote its Sylow pro-l subgroup. Let l be the residue field of L, and let ϖ_L denote a uniformiser of L. Then there is a natural isomorphism $I/I_1 \cong (l^\times)^n$. If $\chi = (\chi_1, \ldots, \chi_n) : (l^\times)^n \to \mathbb{C}^\times$ is a character, then we let $\pi_L^{I,\chi}$ denote the space of vectors in π_L which are fixed by I_1 and transform by χ under the action of I/I_1. The space $\pi_L^{I,\chi}$ has a natural action of the Hecke algebra $\mathcal{H}(I,\chi)$ of compactly supported χ^{-1}-spherical functions on $\mathrm{GL}_n(L)$. We consider the commutative subalgebra $\mathbb{T}(I,\chi)$ of $\mathcal{H}(I,\chi)$ generated by double cosets $[I\alpha I]$ where $\alpha = \mathrm{diag}(\varpi_L^{b_1}, \ldots, \varpi_L^{b_n})$ with $b_1 \geq \cdots \geq b_n \geq 0$.
If $\chi : (\mathcal{O}_L^\times)^n \to \mathbb{C}^\times$ is tamely ramified, then we let $\pi_L^{I,\chi}$ denote $\pi_L^{I,\bar{\chi}}$, where $\bar{\chi}$ is the character $(l^\times)^n \to \mathbb{C}^\times$ determined by χ. Let δ denote the modulus character of $\mathrm{GL}_n(L)$, so that

$$\delta(\mathrm{diag}(a_1, \ldots, a_n)) = |a_1|^{n-1}|a_2|^{n-3}\ldots|a_n|^{1-n}$$

where $|\cdot|$ denotes the usual norm on L.

Proposition 5.5. (1) *If $\pi_L^I \neq 0$ then π is a subquotient of an unramified principal series representation.*

(2) *If $\pi_L^{I_1} \neq 0$ then π is a subquotient of a tamely ramified principal series representation. More precisely, if $\pi_L^{I,\chi} \neq 0$ then π_L is a subquotient of a tamely ramified principal series representation $\text{n-Ind}_{B_n(L)}^{\mathrm{GL}_n(L)}(\chi'_1, \ldots, \chi'_n)$ with χ'_i extending χ_i for each i.*

(3) *If $\pi_L = \text{n-Ind}_{B_n(L)}^{\mathrm{GL}_n(L)}(\chi)$ with χ tamely ramified, then*

$$\pi_L^{I,\chi} \cong \oplus_w \chi\delta^{-1/2}$$

as a $\mathbb{T}(I,\chi)$-module, where the sum is over the elements w of the Weyl group of GL_n with $\chi^w = \chi$; that is, the double coset $[I\alpha I]$ acts via $(\chi\delta^{-1/2})(\alpha)$ on $\pi_L^{I,\chi}$.

Proof. The first two parts follow from Lemma 3.1.6 of [CHT08] and its proof. All three parts follow at once from Theorem 7.7 and Remark 7.8 of [Roc98] (which are valid for GL_n without any restrictions on L - see the proof of Lemma 3.1.6 of [CHT08]), together with the standard calculation of the Jacquet module of a principal series representation, for which see for example Theorem 6.3.5 of [Cas95] (although note that there is a missing factor of $\delta^{1/2}$ (or rather $\delta_\Omega^{1/2}$ in the notation of *loc. cit.*) in the formula given there). □

5.6. Keep our running assumptions on π. Suppose that $U = \prod_v U_v$ is a sufficiently small subgroup of $G(\mathbb{A}_F)$. Assume further that U has been chosen such that if $v \notin S(B)$, $v = ww^c$ splits completely in M, and U_v is a maximal compact subgroup of $G(F_v)$, then π_w is unramified. Recall that we have fixed an isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. There is a maximal ideal $\mathfrak{m}_{\iota,\pi}$ of $\mathbb{T}(U)$ determined by π in the following fashion. For each place $v = ww^c$ as above the Hecke operators $T_w^{(i)}$ act via scalars $\alpha_{w,i}$ on $(\pi_w)^{\mathrm{GL}_n(\mathcal{O}_{M_w})}$. The $\alpha_{w,i}$ are all algebraic integers, so that $\iota^{-1}(\alpha_{w,i}) \in \mathcal{O}$. Then $\mathfrak{m}_{\iota,\pi}$ is the maximal ideal of $\mathbb{T}(U)$ containing all the $T_w^{(i)} - \iota^{-1}(\alpha_{w,i})$. Let $\sigma_k \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,K)}$ be the weight determined by $(\sigma_k)_{\tau,i} = (k-2)(n-i)$ for each $\tau \in \tilde{I}_l$.

LEMMA 5.6. *Suppose that π is a RACSDC representation of $\mathrm{GL}_n(\mathbb{A}_M)$ of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{S(B)}$. Suppose that for each place $x \in \tilde{S}_l$, π_x is a principal series* n-Ind$_{B_n(M_x)}^{\mathrm{GL}_n(M_x)}(\chi_{x,1}, \ldots, \chi_{x,n})$ *with $\iota^{-1}\chi_{x,i} \circ \mathrm{Art}_{F_x}^{-1}|_{I_x} = \omega^{(i-1)(2-k)}$. Then there is a sufficiently small compact open subgroup U of $G(\mathbb{A}_F)$ such that U satisfies the requirements above (in particular, $U = \prod_v U_v$ where U_v is an Iwahori subgroup of $\mathrm{GL}_n(F_v)$ for each $v|l$) and $S_{0,\chi_{\sigma_k}}(U, \mathcal{O})_{\mathfrak{m}_{\iota,\pi}} \neq 0$. If we assume furthermore that $v_l(\iota^{-1}\chi_{x,i}(l)) = [M_x : \mathbb{Q}_l]\left(i - 1 + \frac{1-n}{2}\right)$ for all i (and all $x \in \tilde{S}_l$) then $S^{ord}_{0,\chi_{\sigma_k}}(U, \mathcal{O})_{\mathfrak{m}_{\iota,\pi}} \neq 0$.*

Proof. This is a consequence of Proposition 3.3.2 of [CHT08]. The only issues are at places dividing l and places in $S(B)$. For the latter, it is enough to note that under the Jacquet-Langlands correspondence, $\mathrm{Sp}_n(1)$ corresponds to the trivial representation. For the first part, we also need to check that at each place $x \in \tilde{S}_l$, $\pi_x^{I_x,\chi_x} \neq 0$, where I_x is the standard Iwahori subgroup of $\mathrm{GL}_n(M_x)$, and $\chi_x = (\chi_{x,1}, \ldots, \chi_{x,n})$. This follows at once from Proposition 5.5.

For the second part, we must check in addition that if the Hecke operator $[I_x u_x I_x]$ (where $u_x = \mathrm{diag}(l^{n-1}, \ldots, 1)$) acts via the scalar α_x on $\pi_x^{I_x,\chi_x}$, then $\iota^{-1}(\alpha_x)$ is an l-adic unit. This is straightforward; by Proposition 5.5(3), $\alpha_x = \chi_x(u)\delta^{-1/2}(u)$. Thus

$$\begin{aligned}
v_l(\iota^{-1}(\alpha_x)) &= v_l(\iota^{-1}(\chi_x(u)\delta^{-1/2}(u))) \\
&= \sum_{i=1}^n (n-i) v_l(\iota^{-1}\chi_{x,i}(l)) + \sum_{i=1}^n (n-i) v_l((l^{-[M_x:\mathbb{Q}_l]})^{-(n+1-2i)/2})) \\
&= \sum_{i=1}^n (n-i)([M_x : \mathbb{Q}_l]\left(i - 1 + \frac{1-n}{2}\right)) + \\
&\qquad\qquad + \sum_{i=1}^n [M_x : \mathbb{Q}_l](n-i)(n+1-2i)/2 \\
&= \frac{[M_x : \mathbb{Q}_l]}{2} \sum_{i=1}^n (n-i)((2i-1-n) + (n+1-2i)) \\
&= 0,
\end{aligned}$$

as required. □

Lemma 5.7. *Keep (all) the assumptions of Lemma 5.6. Then there is an RACSDC representation π'' of $\mathrm{GL}_n(\mathbb{A}_M)$ of weight $\iota_*\sigma_k$, type $\{\mathrm{Sp}_n(1)\}_{\{S(B)\}}$ and with π''_l unramified such that $\overline{r}_{l,\iota}(\pi'') \cong \overline{r}_{l,\iota}(\pi)$.*

Proof. This is essentially a consequence of Lemma 5.6, Lemma 5.4, and Proposition 5.5, together with Proposition 3.3.2 of [CHT08]. Indeed, Lemma 5.4 and Lemma 5.6 show that $S^{ord}_{\sigma_k,1}(U,\mathcal{O})_{\mathfrak{m}_{\iota,\pi}} \neq 0$, which by Proposition 5.5(1) and Proposition 3.3.2 of [CHT08] gives us a π'' satisfying all the properties we claim, except that we only know that for each $x|l$, π''_x is a subquotient of an unramified principal series representation. We claim that this unramified principal series is irreducible, so that π''_x is unramified. To see this, note that the fact that we know that $S^{ord}_{\sigma_k,1}(U,\mathcal{O})_{\mathfrak{m}_{\iota,\pi}} \neq 0$ (rather than merely $S_{\sigma_k,1}(U,\mathcal{O})_{\mathfrak{m}_{\iota,\pi}} \neq 0$) means that we can choose π'' so that for each $x \in \tilde{S}_l$, π''_x is a subquotient of an unramified principal series representation $\text{n-Ind}^{\mathrm{GL}_n(M_x)}_{B_n(M_x)}(\chi_{x,1},\ldots,\chi_{x,n})$ with

$$v_l(\iota^{-1}\chi_{x,i}(l)) = [M_x:\mathbb{Q}_l]\,((i-1)(k-1)+(1-n)/2)$$

(this follows from the comparison of the Hecke actions on $(\pi''_x)^{I_x}$ and $S_{\sigma_k,1}(U,\mathcal{O})$, noting that the latter action is defined in terms of $\cdot_{twist}$). Now, if the principal series $\text{n-Ind}^{\mathrm{GL}_n(M_x)}_{B_n(M_x)}(\chi_{x,1},\ldots,\chi_{x,n})$ were reducible, there would be i, j with $\chi_{x,i} = \chi_{x,j}|\cdot|$, so that $\chi_{x,i}(l)l^{[M_x:\mathbb{Q}_l]} = \chi_{x,j}(l)$, which is a contradiction because $k > 2$. The result follows. □

Combining Corollary 4.4 with Lemma 5.7, we obtain

Proposition 5.8. *Let l be an odd prime, and let f be a modular form of weight $2 \le k < l$ and level coprime to l. Assume that f is Steinberg at q, and that for some place $\lambda|l$ of $\mathcal{O}_f$, f is ordinary at λ and $\overline{\rho}_{f,\lambda}$ is absolutely irreducible. Fix an embedding $K_{f,\lambda} \hookrightarrow \overline{\mathbb{Q}}_l$. Let $\mathcal{N}$ be a finite set of even positive integers. Then there is a Galois totally real extension $F/\mathbb{Q}$ and a quadratic imaginary field E, together with a place $w_q|q$ of $M = FE$ such that if we choose a set $\tilde{S}_l$ of places of M consisting of one place above each place of F dividing l, and define $\sigma_k \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,K)}$ by $(\sigma_k)_{\tau,i} = (k-2)(n-i)$, then*

- *for each $n \in \mathcal{N}$, there is a character $\bar{\phi}_n : \mathrm{Gal}(\overline{M}/M) \to \overline{\mathbb{F}}_l^\times$ which is unramified at all places in $\tilde{S}_l$, which satisfies*
$$\bar{\phi}_n\bar{\phi}_n^c = (\epsilon\det\overline{\rho}_{f,\lambda}\otimes\overline{\mathbb{F}}_l)^{1-n}|_{\mathrm{Gal}(\overline{M}/M)}$$
and $(\mathrm{Sym}^{n-1}\overline{\rho}_{f,\lambda}\otimes\overline{\mathbb{F}}_l)|_{\mathrm{Gal}(\overline{M}/M)}\otimes\bar{\phi}_n$ is automorphic of weight σ_k and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$.
- *l is unramified in M.*
- *M is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$.*

6. Potential automorphy

6.1. Assume as before that f is a cuspidal newform of level $\Gamma_1(N)$, weight $k \ge 2$, and nebentypus χ_f. Let $\pi(f)$ be the RAESDC representation of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$

corresponding to f. We will think of χ_f as an automorphic representation of $\mathrm{GL}_1(\mathbb{A}_\mathbb{Q})$, and write $\chi_f = \otimes_p \chi_{f,p}$. We now define what we mean by the claim that the symmetric powers of f are potentially automorphic. If F is a totally real field and $v|p$ is a place of F, we write $\mathrm{rec}(\pi_{f,p})|_{F_v}$ for the restriction of the Weil-Deligne representation $\mathrm{rec}(\pi_{f,p})$ to the Weil group of F_v. Then we say that $\mathrm{Sym}^{n-1} f$ is potentially automorphic over F if there is an RAESDC representation π_n of $\mathrm{GL}_n(\mathbb{A}_F)$ such that for all primes p and all places $v|p$ of F we have

$$\mathrm{rec}(\pi_{n,v}) = \mathrm{Sym}^{n-1}(\mathrm{rec}(\pi_{f,p})|_{F_v}).$$

By a standard argument (see for example section 4 of [HSBT09]) this is equivalent to asking that $\mathrm{Sym}^{n-1}\rho_{f,\lambda}|_{\mathrm{Gal}(\overline{F}/F)}$ be automorphic for some place (equivalently for all places) λ of K_f.

Similarly, we may speak of $\mathrm{Sym}^{n-1} f$ being potentially automorphic of a specific weight and type. We then define (for each $n \geq 1$ and each integer a) the L-series

$$L(\chi_f^a \otimes \mathrm{Sym}^{n-1} f, s) = \prod_p L((\chi_{f,p}^a \circ \mathrm{Art}_{\mathbb{Q}_p}^{-1}) \otimes \mathrm{Sym}^{n-1}\mathrm{rec}(\pi_{f,p}), s + (1-n)/2).$$

We now normalise the L-functions of RAESDC automorphic representations to agree with those of their corresponding Galois representations. Specifically, if π is an RAESDC representation of $\mathrm{GL}_n(\mathbb{A}_F)$, we define

$$L(\pi, s) = \prod_{v\nmid\infty} L(\pi_v, s + (1-n)/2).$$

If π is square integrable at some finite place, then for each isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$ there is a Galois representation $r_{l,\iota}(\pi)$, and by definition we have

$$\begin{aligned} L(\pi, s) &= \prod_{v\nmid\infty} L(\pi_v \otimes (|\cdot| \circ \det)^{(1-n)/2}, s) \\ &= \prod_{v\nmid\infty} L(\mathrm{rec}(\pi_v \otimes (|\cdot| \circ \det)^{(1-n)/2}), s) \\ &= \prod_{v\nmid\infty} L(r_l(\iota^{-1}\pi_v)^\vee(1-n), s) \\ &= L(r_{l,\iota}(\pi), s). \end{aligned}$$

Theorem 6.1. *Suppose that f is a cuspidal newform of level $\Gamma_1(N)$ and weight $k = 2$ or 3. Suppose that f is Steinberg at q. Let $\mathcal{N}$ be a finite set of even positive integers. Then there is a Galois totally real field F such that for any $n \in \mathcal{N}$ and any subfield $F' \subset F$ with F/F' soluble,* $\mathrm{Sym}^{n-1} f$ *is automorphic over F'.*

Proof. By Lemma 3.2 and Lemma 3.5 we may choose a prime $l > 3$ and a place λ of $\mathcal{O}_f$ lying over l such that

- $l \nmid N$.
- f is ordinary at λ.
- $l > \max(2n+1)_{n\in\mathcal{N}}$.

- $\overline{\rho}_{f,\lambda}$ has large image.

By Corollary 5.8 there is an embedding $K_{f,\lambda} \hookrightarrow \overline{\mathbb{Q}}_l$, a Galois totally real extension $F/\mathbb{Q}$ and a quadratic imaginary field E, together with a place $w_q|q$ of $M = FE$ such that if we choose a set $\tilde{S}_l$ of places of M consisting of one place above each place of F dividing l, and define $\sigma_k \in (\mathbb{Z}^n)^{\mathrm{Hom}(M,K)}$ by $(\sigma_k)_{\tau,i} = (k-2)(n-i)$, then

- for each $n \in \mathcal{N}$, there is a character $\bar{\phi}_n : \mathrm{Gal}(\overline{M}/M) \to \overline{\mathbb{F}}_l^\times$ which is unramified at all places in $\tilde{S}_l$ and satisfies
$$\bar{\phi}_n \bar{\phi}_n^c = (\epsilon \det \overline{\rho}_{f,\lambda} \otimes \overline{\mathbb{F}}_l)^{1-n}|_{\mathrm{Gal}(\overline{M}/M)},$$
and $(\mathrm{Sym}^{n-1} \overline{\rho}_{f,\lambda} \otimes \overline{\mathbb{F}}_l)|_{\mathrm{Gal}(\overline{M}/M)} \otimes \bar{\phi}_n$ is automorphic of weight σ_k and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$.
- l is unramified in M.
- M is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$.

Fix $n \in \mathcal{N}$, and let $\rho := \mathrm{Sym}^{n-1} \rho_{f,\lambda}|_{\mathrm{Gal}(\overline{F}/F)} \otimes \overline{\mathbb{Q}}_l$. There is a crystalline character $\chi : \mathrm{Gal}(\overline{F}/F) \to \mathcal{O}_{\overline{\mathbb{Q}}_l}^\times$ which is unramified above q such that
$$\rho^\vee \cong \rho\chi\epsilon^{n-1};$$
in fact,
$$\chi = (\epsilon \det \rho_{f,\lambda} \otimes \mathcal{O}_{\overline{\mathbb{Q}}_l})^{1-n}|_{\mathrm{Gal}(\overline{F}/F)}.$$
By Lemma 4.1.6 of [CHT08] we can choose an algebraic character
$$\psi : \mathrm{Gal}(\overline{M}/M) \to \mathcal{O}_{\overline{\mathbb{Q}}_l}^\times$$
such that

- $\chi|_{\mathrm{Gal}(\overline{M}/M)} = \psi\psi^c$,
- ψ is crystalline,
- ψ is unramified at each place in $\tilde{S}_l$.
- ψ is unramified above q,
- $\bar{\psi} = \bar{\phi}_n$.

Then $\rho' = \rho|_{\mathrm{Gal}(\overline{M}/M)}\psi$ satisfies
$$(\rho')^c \cong (\rho')^\vee \epsilon^{1-n}.$$
We claim that ρ' is automorphic of weight σ_k, level prime to l and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$. This follows from Theorem 5.2 of [Tay08]; we now check the hypotheses of that theorem. Certainly $\overline{\rho}' \cong (\mathrm{Sym}^{n-1} \overline{\rho}_{f,\lambda} \otimes \overline{\mathbb{F}}_l)|_{\mathrm{Gal}(\overline{M}/M)} \otimes \bar{\phi}_n$ is automorphic of weight σ_k, level prime to l and type $\{\mathrm{Sp}_n(1)\}_{\{w_q\}}$. The only non-trivial conditions to check are that:

- $\overline{M}^{\ker \mathrm{ad}\, \overline{\rho}'}$ does not contain $M(\zeta_l)$, and
- The image $\overline{\rho}'(\mathrm{Gal}(\overline{M}/M(\zeta_l)))$ is big in the sense of Definition 2.5.1 of [CHT08].

These both follow from the assumption that $\overline{\rho}_{f,\lambda}$ has large image, the fact that M is linearly disjoint from $\overline{\mathbb{Q}}^{\ker(\overline{\rho}_{f,\lambda})}$ over $\mathbb{Q}$, Corollary 2.5.4 of [CHT08], and the fact that $\mathrm{PSL}_2(k)$ is simple if k is a finite field of cardinality greater than 3.

It follows from Lemma 4.3.3 of [CHT08] that ρ is automorphic. Then from Lemma 4.3.2 of [CHT08] we see that for each F' with F/F' soluble, $\mathrm{Sym}^{n-1} \rho_{f,\lambda}|_{\mathrm{Gal}(\overline{F}/F')}$ is automorphic, as required. □

Corollary 6.2. *Suppose that f is a cuspidal newform of level $\Gamma_1(N)$ and weight $k = 2$ or 3. Suppose that f is potentially Steinberg at q. Let $\mathcal{N}$ be a finite set of even positive integers. Then there is a Galois totally real field F such that for any $n \in \mathcal{N}$ and any subfield $F' \subset F$ with F/F' soluble,* $\mathrm{Sym}^{n-1} f$ *is automorphic over F'.*

Proof. Let θ be a Dirichlet character such that $f' = f \otimes \theta$ is Steinberg at q. The result then follows from Theorem 6.1 applied to f'. □

7. The Sato-Tate Conjecture

7.1. Let f be a cuspidal newform of level $\Gamma_1(N)$, nebentypus χ_f, and weight $k \geq 2$. Suppose that χ_f has order m, so that the image of χ_f is precisely the group μ_m of m-th roots of unity. Let $U(2)_m$ be the subgroup of $U(2)$ consisting of matrices with determinant in μ_m. For each prime $l \nmid N$, if we write

$$X^2 - a_l X + l^{k-1}\chi_f(l) = (X - \alpha_l l^{(k-1)/2})(X - \beta_l l^{(k-1)/2})$$

then (by the Ramanujan conjecture) the matrix

$$\begin{pmatrix} \alpha_l & 0 \\ 0 & \beta_l \end{pmatrix}$$

defines a conjugacy class $x_{f,l}$ in $U(2)_m$. A natural generalisation of the Sato-Tate conjecture is

Conjecture 7.1. *If f is not of CM type, then the conjugacy classes $x_{f,l}$ are equidistributed with respect to the Haar measure on $U(2)_m$ (normalised so that $U(2)_m$ has measure 1).*

The group $U(2)_m$ is compact, and its irreducible representations are given by $\det^a \otimes \mathrm{Sym}^b \mathbb{C}^2$ for $0 \leq a < m$ and $b \geq 0$. By the corollary to Theorem 2 of section I.A.2 of [Ser68] (noting the different normalisations of L-functions in force there), Conjecture 7.1 follows if one knows that for each $b \geq 1$, the functions $L(\chi_f^a \otimes \mathrm{Sym}^b f, s)$ are holomorphic and non-zero for $\Re s \geq 1 + b(k-1)/2$ (the required results for $b = 0$ are classical).

Theorem 7.2. *Suppose that f is a cuspidal newform of level $\Gamma_1(N)$, character χ_f, and weight $k = 2$ or 3. Suppose that χ_f has order m. Suppose also that f is potentially Steinberg at q for some prime q. Then for all integers $0 \leq a < m$, $b \geq 1$ the function $L(\chi_f^a \otimes \mathrm{Sym}^b f, s)$ has meromorphic continuation to the whole* *complex plane, satisfies the expected functional equation, and is holomorphic and nonzero in* $\Re s \geq 1 + b(k-1)/2$.

Proof. The argument is very similar to the proof of Theorem 4.2 of [HSBT09]. We argue by induction on b; suppose that b is odd, and the result is known for all $1 \leq b' < b$. We will deduce the result for b and for $b+1$ simultaneously. Apply Corollary 6.2 with $\mathcal{N} = \{2, b+1\}$. Let F be as in the conclusion of Corollary 6.2. By Brauer's theorem, we may write

$$1 = \Sigma_j a_j \operatorname{Ind}_{\mathrm{Gal}(F/F_j)}^{\mathrm{Gal}(F/\mathbb{Q})} \chi_j$$

where $F \supset F_j$ with F/F_j soluble, χ_j a character $\mathrm{Gal}(F/F_j) \to \mathbb{C}^\times$, and $a_j \in \mathbb{Z}$. Then for each j, $\mathrm{Sym}^b f$ is automorphic over F_j, corresponding to an RAESDC representation π_j of $\mathrm{GL}_{b+1}(\mathbb{A}_{F_j})$. In addition, f is automorphic over F_j, corresponding to an RAESDC representation σ_j of $\mathrm{GL}_2(\mathbb{A}_{F_j})$.
Then we have

$$L(\chi_f^a \otimes \mathrm{Sym}^b f, s) = \prod_j L(\pi_j \otimes (\chi_j \circ \mathrm{Art}_{F_j}) \otimes (\chi_f^a \circ N_{F_j/\mathbb{Q}}), s)^{a_j},$$

$$L(\chi_f^a \otimes \mathrm{Sym}^2 f, s) = \prod_j L((\mathrm{Sym}^2 \sigma_j) \otimes (\chi_j \circ \mathrm{Art}_{F_j}) \otimes (\chi_f^a \circ N_{F_j/\mathbb{Q}}), s)^{a_j},$$

and

$$\begin{aligned} &L(\chi_f^a \otimes \mathrm{Sym}^{b+1} f, s) L(\chi_f^{a+1} \otimes \mathrm{Sym}^{b-1} f, s-k+1) = \\ &\qquad = \prod_j L((\pi_j \otimes (\chi_j \circ \mathrm{Art}_{F_j}) \otimes (\chi_f^a \circ N_{F_j/\mathbb{Q}})) \times \sigma_j, s + b(k-1)/2)^{a_j}. \end{aligned}$$

The result then follows from the main results of [CPS04] and [GJ78] (in the case $b = 1$) together with Theorem 5.1 of [Sha81]. $\square$

Corollary 7.3. *Suppose that f is a cuspidal newform of level $\Gamma_1(N)$ and weight $k = 2$ or 3. Suppose also that f is potentially Steinberg at q for some prime q. Then Conjecture 7.1 holds for f.*

Finally, we note that one can make this result more concrete, as one can easily explicitly determine the Haar measure on $U(2)_m$ from that of its finite index subgroup $SU(2)$. One finds that (as already follows from Dirchlet's theorem) the classes $x_{f,l}$ are equidistributed by determinant, and that furthermore the classes with fixed determinant are equidistributed with respect to the natural analogue of the usual Sato-Tate measure. That is, suppose that $\zeta \in \mu_m$, and fix a square root $\zeta^{1/2}$ of ζ. Then any conjugacy class $x_{f,l}$ in $U(2)_m$ with determinant ζ contains a representative of the form

$$\begin{pmatrix} \zeta^{1/2} e^{i\theta_l} & 0 \\ 0 & \zeta^{1/2} e^{-i\theta_l} \end{pmatrix}$$

with $\theta_l \in [0, \pi]$, and the θ_l are equidistributed with respect to the measure $\frac{2}{\pi} \sin^2 \theta d\theta$.

References

[BLGHT09] Tom Barnet-Lamb, David Geraghty, Michael Harris, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy II*, Preprint, 2009.

[Cas95] Bill Casselman, *Introduction to the theory of admissible representations of p-adic reductive groups*, 1995.

[CDT99] Brian Conrad, Fred Diamond, and Richard Taylor, *Modularity of certain potentially Barsotti-Tate Galois representations*, J. Amer. Math. Soc. 12 (1999), no. 2, 521–567.

[CH09] Gaëtan Chenevier and Michael Harris, *Construction of automorphic Galois representations, II*, preprint, 2009.

[Che04] Gaëtan Chenevier, *Familles p-adiques de formes automorphes pour* GL_n, J. Reine Angew. Math. 570 (2004), 143–217. MR MR2075765 (2006b:11046)

[CHT08] Laurent Clozel, Michael Harris, and Richard Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations*, Pub. Math. IHES 108 (2008), 1–181.

[CPS04] James W. Cogdell and Ilya I. Piatetski-Shapiro, *Remarks on Rankin-Selberg convolutions*, Contributions to automorphic forms, geometry, and number theory, Johns Hopkins Univ. Press, Baltimore, MD, 2004, pp. 255–278. MR MR2058610 (2005d:11075)

[Dim05] Mladen Dimitrov, *Galois representations modulo p and cohomology of Hilbert modular varieties*, Ann. Sci. École Norm. Sup. (4) 38 (2005), no. 4, 505–551. MR MR2172950 (2006k:11100)

[GJ78] Stephen Gelbart and Hervé Jacquet, *A relation between automorphic representations of* $\mathrm{GL}(2)$ *and* $\mathrm{GL}(3)$, Ann. Sci. École Norm. Sup. (4) 11 (1978), no. 4, 471–542. MR MR533066 (81e:10025)

[Gue09] L. Guerberoff, *Modularity lifting theorems for Galois representations of unitary type*, Arxiv preprint arXiv:0906.4189 (2009).

[HSBT09] Michael Harris, Nick Shepherd-Barron, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy*, to appear in Annals of Mathematics (2009).

[HT01] Michael Harris and Richard Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, vol. 151, Princeton University Press, Princeton, NJ, 2001, With an appendix by Vladimir G. Berkovich.

[KW08] Chandrashekhar Khare and Jean-Pierre Wintenberger, *On Serre's conjecture for 2-dimensional mod p representations of the absolute Galois group of the rationals*, to appear in Annals of Mathematics (2008).

[Lan79] R. P. Langlands, *Automorphic representations, Shimura varieties, and motives. Ein Märchen*, Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, pp. 205–246. MR MR546619 (83f:12010)

[Rib75] Kenneth A. Ribet, *On l-adic representations attached to modular forms*, Invent. Math. 28 (1975), 245–275. MR MR0419358 (54 #7379)

[Roc98] Alan Roche, *Types and Hecke algebras for principal series representations of split reductive p-adic groups*, Ann. Sci. École Norm. Sup. (4) 31 (1998), no. 3, 361–413. MR MR1621409 (99d:22028)

[Sch90] A. J. Scholl, *Motives for modular forms*, Invent. Math. 100 (1990), no. 2, 419–430. MR MR1047142 (91e:11054)

[Ser68] Jean-Pierre Serre, *Abelian l-adic representations and elliptic curves*, McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam, 1968.

[Ser81] ———, *Quelques applications du théorème de densité de Chebotarev*, Inst. Hautes Études Sci. Publ. Math. (1981), no. 54, 323–401. MR MR644559 (83k:12011)

[Sha81] Freydoon Shahidi, *On certain L-functions*, Amer. J. Math. 103 (1981), no. 2, 297–355. MR MR610479 (82i:10030)

[Shi09] Sug Woo Shin, *Galois representations arising from some compact Shimua varieties*, preprint, 2009.

[Tay88] Richard Taylor, *On congruences between modular forms*, PhD Thesis, Princeton, 1988.

[Tay02] ———, *Remarks on a conjecture of Fontaine and Mazur*, J. Inst. Math. Jussieu 1 (2002), no. 1, 125–143.

[Tay08] ———, *Automorphy for some l-adic lifts of automorphic mod l Galois representations. II*, Pub. Math. IHES 108 (2008), 183–239.

[Wil88] A. Wiles, *On ordinary λ-adic representations associated to modular forms*, Invent. Math. 94 (1988), no. 3, 529–573.

Toby Gee
Department of Mathematics
Harvard University
gee@math.harvard.edu

Subproduct Systems

Orr Moshe Shalit and Baruch Solel [1]

Received: July 1, 2009
Revised: August 12, 2009

Communicated by Joachim Cuntz

Abstract. The notion of a *subproduct system*, a generalization of that of a product system, is introduced. We show that there is an essentially 1 to 1 correspondence between *cp*-semigroups and pairs (X, T) where X is a subproduct system and T is an injective subproduct system representation. A similar statement holds for subproduct systems and units of subproduct systems. This correspondence is used as a framework for developing a dilation theory for *cp*-semigroups. Results we obtain:
(i) a $*$-automorphic dilation to semigroups of $*$-endomorphisms over quite general semigroups;
(ii) necessary and sufficient conditions for a semigroup of CP maps to have a $*$-endomorphic dilation;
(iii) an analogue of Parrot's example of three contractions with no isometric dilation, that is, an example of three commuting, contractive normal CP maps on $B(H)$ that admit no $*$-endomorphic dilation (thereby solving an open problem raised by Bhat in 1998).
Special attention is given to subproduct systems over the semigroup $\mathbb{N}$, which are used as a framework for studying tuples of operators satisfying homogeneous polynomial relations, and the operator algebras they generate. As applications we obtain a noncommutative (projective) Nullstellensatz, a model for tuples of operators subject to homogeneous polynomial relations, a complete description of all representations of Matsumoto's subshift C*-algebra when the subshift is of finite type, and a classification of certain operator algebras – including an interesting non-selfadjoint generalization of the noncommutative tori.

2000 Mathematics Subject Classification: 46L55, 46L57, 46L08, 47L30.

[1]B.S. was supported in part by the US-Israel Binational Foundation and by the Fund for the Promotion of Research at the Technion.

Keywords and Phrases: Product system, subproduct system, semigroups of completely positive maps, dilation, e_0-dilation, $*$-automorphic dilation, row contraction, homogeneous polynomial identities, universal operator algebra, q-commuting, subshift C*-algebra.

Contents

Introduction

Motivation: dilation theory of CP_0-semigroups. We begin by describing the problems that motivated this work.
Let H be a separable Hilbert space, and let $\mathcal{M} \subseteq B(H)$ be a von Neumann algebra. A *CP map* on $\mathcal{M}$ is a contractive, normal and completely positive map. A *CP_0-semigroup* on $\mathcal{M}$ is a family $\Theta = \{\Theta_t\}_{t\geq 0}$ of unital CP maps on $\mathcal{M}$ satisfying the semigroup property

$$\Theta_{s+t}(a) = \Theta_s(\Theta_t(a)) \ , \ s,t \geq 0, a \in \mathcal{M},$$

$$\Theta_0(a) = a \ , \ a \in B(H),$$

and the continuity condition

$$\lim_{t\to t_0} \langle \Theta_t(a)h, g\rangle = \langle \Theta_{t_0}(a)h, g\rangle \ , \ a \in \mathcal{M}, h, g \in H.$$

A CP_0-semigroup is called an *E_0-semigroup* if each of its elements is a $*$-endomorphism.

Let Θ be a CP_0-semigroup acting on $\mathcal{M}$, and let α be an E_0-semigroup acting on $\mathcal{R}$, where $\mathcal{R}$ is a von Neumann subalgebra of $B(K)$ and $K \supseteq H$. Denote the orthogonal projection of K onto H by p. We say that α is an *E_0-dilation* of Θ if for all $t \geq 0$ and $b \in \mathcal{R}$

$$\Theta_t(pbp) = p\alpha_t(b)p. \tag{0.1}$$

In the mid 1990's Bhat proved the following result, known today as "Bhat's Theorem" (see [9] for the case $\mathcal{M} = B(H)$, and also [40, 15, 29, 6] for different proofs and for the general case):

Theorem 0.1. (Bhat). *Every CP_0-semigroup has a unique minimal E_0-dilation.*

A natural question is then this: *given* two *commuting CP_0-semigroups, can one simultaneously dilate them to a pair of commuting E_0-semigroups?* In [43] the following partial positive answer was obtained[2]:

Theorem 0.2. [43, Theorem 6.6] *Let $\{\phi_t\}_{t\geq 0}$ and $\{\theta_t\}_{t\geq 0}$ be two strongly commuting CP_0-semigroups on a von Neumann algebra $\mathcal{M} \subseteq B(H)$, where H is a separable Hilbert space. Then there is a separable Hilbert space K containing H and an orthogonal projection $p : K \to H$, a von Neumann algebra $\mathcal{R} \subseteq B(K)$ such that $\mathcal{M} = p\mathcal{R}p$, and two commuting E_0-semigroups α and β on $\mathcal{R}$ such that*

$$\phi_s \circ \theta_t(pbp) = p\alpha_s \circ \beta_t(b)p$$

for all $s, t \geq 0$ and all $b \in \mathcal{R}$.

In other words: every two-parameter CP_0-semigroup that satisfies an additional condition of *strong commutativity* has a two-parameter E_0-dilation. The condition of strong commutativity was introduced in [48]. A precise definition will not be given here. The main tools in the proof of Theorem 0.2, and also in some of the proofs of Theorem 0.1, were product systems of W^*-correspondences and their representations. In fact, the only place in the proof of Theorem 0.2 where the assumption of strong commutativity is used, is in the construction of a certain product system. More about that later.

In [10], Bhat showed that given a pair of commuting CP maps Θ and Φ on $B(H)$, there is a Hilbert space $K \supseteq H$ and a pair of commuting normal $*$-endomorphisms α and β acting on $B(K)$ such that

$$\Theta^m \circ \Phi^n(pbp) = p\alpha^m \circ \beta^n(b)p \ , \ b \in B(K)$$

for all $m, n \in \mathbb{N}$ (here p denotes the projection of K onto H). Later on Solel, using a different method (using in fact product systems and their representations), proved this result for commuting CP maps on arbitrary von Neumann algebras [48]. Neither one of the above results requires strong commutativity.

[2]The same result was obtained in [42] for nonunital semigroups acting on $\mathcal{M} = B(H)$.

In light of the above discussion, and inspired by classical dilation theory [47, 49], it is natural to conjecture that *every* two commuting (not necessarily *strongly* commuting) CP_0-semigroups have an E_0-dilation, and in fact that the same is true for any k commuting CP_0-semigroups, for any positive integer k. However, the framework given by product systems seems to be too weak to prove this. Trying to bypass this stoppage, we arrived at the notion of a *subproduct system.*

Background: from product systems to subproduct systems. Product systems of Hilbert spaces over $\mathbb{R}_+$ were introduced by Arveson some 20 years ago in his study of E_0-semigroups [3]. In a few imprecise words, a product system of Hilbert spaces over $\mathbb{R}_+$ is a bundle $\{X(t)\}_{t\in\mathbb{R}_+}$ of Hilbert spaces such that

$$X(s+t) = X(s) \otimes X(t) \ , \ s,t \in \mathbb{R}_+.$$

We emphasize immediately that Arveson's definition of product systems required also that the bundle carry a certain Borel measurable structure, but we do not deal with these matters here. To every E_0-semigroup Arveson associated a product system, and it turns out that the product system associated to an E_0-semigroup is a complete cocycle conjugacy invariant of the E_0-semigroup.
Later, product systems of Hilbert C^*-correspondences over $\mathbb{R}_+$ appeared (see the survey [46] by Skeide). In [15], Bhat and Skeide associate with every semigroup of completely positive maps on a C^*-algebra A a product system of Hilbert A-correspondences. This product system was then used in showing that every semigroup of completely positive maps can be "dilated" to a semigroup of $*$-endomorphisms. Muhly and Solel introduced a different construction [29]: to every CP_0-semigroup on a von Neumann algebra $\mathcal{M}$ they associated a product system of Hilbert W^*-correspondences over $\mathcal{M}'$, the commutant of $\mathcal{M}$. Again, this product system is then used in constructing an E_0-dilation for the original CP_0-semigroup.
Product systems of C^*-correspondences over semigroups other than $\mathbb{R}_+$ were first studied by Fowler [21], and they have been studied since then by many authors. In [48], product systems over $\mathbb{N}^2$ (and their representations) were studied, and the results were used to prove that every pair of commuting CP maps has a $*$-endomorphic dilation. Product systems over $\mathbb{R}_+^2$ were also central to the proof of Theorem 0.2, where every pair of strongly commuting CP_0-semigroups is associated with a product system over $\mathbb{R}_+^2$. However, the construction of the product system is one of the hardest parts in that proof. Furthermore, that construction fails when one drops the assumption of strong commutativity, and it also fails when one tries to repeat it for k strongly commuting semigroups.
On the other hand there is another object that may be naturally associated with a semigroup of CP maps over *any* semigroup: this object is the *subproduct system*, which, when the CP maps act on $B(H)$, is the bundle of Arveson's "metric operator spaces" (introduced in [4]). Roughly, a subproduct system of correspondences over a semigroup $\mathcal{S}$ is a bundle $\{X(s)\}_{s\in\mathcal{S}}$ of correspondences such that

$$X(s+t) \subseteq X(s) \otimes X(t) \ , \ s,t \in \mathcal{S}.$$

See Definition 1.1 below. Of course, a difficult problem cannot be made easy just by introducing a new notion, and the problem of dilating k-parameter CP_0-semigroups remains unsolved. However, subproduct systems did already provide us with an efficient general framework for tackling various problems in operator algebras, and in particular it has led us to a progress toward the solution of the discrete analogue of the above unsolved problem.

While preparing this paper we learned that Bhat and Mukherjee have also considered subproduct systems over the semigroup $\mathbb{R}_+$ ([14]). They called it inclusion systems and used it to compute the index of certain product systems.

This paper consists of two parts. In the first part we introduce subproduct systems over general semigroups, show the connection between subproduct systems and *cp*-semigroups, and use this connection to obtain three main results in dilation theory of *cp*-semigroups. The first result is that every e_0-semigroup over a (certain kind of) semigroup $\mathcal{S}$ can be dilated to a semigroup of $*$-automorphisms on some type I factor. The second is some necessary conditions and sufficient conditions for a *cp*-semigroup to have a (minimal) $*$-endomorphic dilation. The third is an analogue of Parrot's example of three contractions with no isometric dilation, that is, an example of three commuting, contractive normal CP maps on $B(H)$ that admit no $*$-endomorphic dilation. The CP maps in the stated example can be taken to have *arbitrarily small norm*, providing the first example of a theorem in the classical theory of isometric dilations that cannot be generalized to the theory of e-dilations of *cp*-semigroups.

Having convinced the reader that subproduct systems are an interesting and important object, we turn in the second part of the paper to take a closer look at the simplest examples of subproduct systems, that is, subproduct systems of Hilbert spaces over $\mathbb{N}$. We study certain tuples of operators and operator algebras that can be naturally associated with every subproduct system, and explore the relationship between these objects and the subproduct systems that give rise to them.

Some preliminaries. $\mathcal{M}$ and $\mathcal{N}$ will denote von Neumann subalgebras of $B(H)$, where H is some Hilbert space.

In Sections 1 through 5, $\mathcal{S}$ will denote a sub-semigroup of $\mathbb{R}_+^k$. In fact, in large parts of the paper $\mathcal{S}$ can be taken to be any semigroup with unit, or at least any *Ore semigroup* (see [25] for a definition), but we prefer to avoid this distraction.

Definition 0.3. *A* cp-semigroup *is a semigroup of CP maps, that is, a family $\Theta = \{\Theta_s\}_{s\in\mathcal{S}}$ of completely positive, contractive and normal maps on $\mathcal{M}$ such that*

$$\Theta_{s+t}(a) = \Theta_s(\Theta_t(a)) \ , \ s,t \in \mathcal{S}, a \in \mathcal{M}$$

and

$$\Theta_0(a) = a \ , \ a \in \mathcal{M}.$$

A cp_0-semigroup *is a semigroup of unital CP maps. An* e-semigroup *is a semigroup of $*$-endomorphisms. An* e_0-semigroup *is a semigroup of unital $*$-endomorphisms.*

For concreteness, one should think of the case $\mathcal{S} = \mathbb{N}^k$, where a *cp*-semigroup is a k-tuple of commuting CP maps, or the case $\mathcal{S} = \mathbb{R}_+^k$, where a *cp*-semigroup is a k-parameter semigroup of CP maps, or k mutually commuting one-parameter *cp*-semigroups.

Definition 0.4. *Let $\Theta = \{\Theta_s\}_{s\in\mathcal{S}}$ be a cp-semigroup acting on a von Neumann algebra $\mathcal{M} \subseteq B(H)$. An* e-dilation of Θ *is a triple $(\alpha, K, \mathcal{R})$ consisting of a Hilbert space $K \supseteq H$ (with orthogonal projection $P_H : K \to H$), a von Neumann algebra $\mathcal{R} \subseteq B(K)$ that contains $\mathcal{M}$ as a corner $\mathcal{M} = P_H \mathcal{R} P_H$, and an e-semigroup $\alpha = \{\alpha_s\}_{s\in\mathcal{S}}$ on $\mathcal{R}$ such that for all $T \in \mathcal{R}$, $s \in \mathcal{S}$,*

$$\Theta_s(P_H T P_H) = P_H \alpha_s(T) P_H.$$

Definition 0.5. *Let $\mathcal{A}$ be a C^*-algebra. A* Hilbert C^*-correspondences over *$\mathcal{A}$ is a (right) Hilbert $\mathcal{A}$-module E which carries a non-degenerate, adjointable, left action of $\mathcal{A}$.*

Definition 0.6. *Let $\mathcal{M}$ be a W^*-algebra. A* Hilbert W^*-correspondence over *$\mathcal{M}$ is a self-adjoint Hilbert C^*-correspondence E over $\mathcal{M}$, such that the map $\mathcal{M} \to \mathcal{L}(E)$ which gives rise to the left action is normal.*

Definition 0.7. *Let E be a C^*-correspondence over $\mathcal{A}$, and let H be a Hilbert space. A pair (σ, T) is called a* completely contractive covariant representation *of E on H (or, for brevity, a* c.c. representation*) if*

(1) *$T : E \to B(H)$ is a completely contractive linear map;*
(2) *$\sigma : A \to B(H)$ is a nondegenerate $*$-homomorphism; and*
(3) *$T(xa) = T(x)\sigma(a)$ and $T(a\cdot x) = \sigma(a)T(x)$ for all $x \in E$ and all $a \in \mathcal{A}$.*

If $\mathcal{A}$ is a W^*-algebra and E is W^*-correspondence then we also require that σ be normal.

Given a C^*-correspondence E and a c.c. representation (σ, T) of E on H, one can form the Hilbert space $E \otimes_\sigma H$, which is defined as the Hausdorff completion of the algebraic tensor product with respect to the inner product

$$\langle x \otimes h, y \otimes g\rangle = \langle h, \sigma(\langle x, y\rangle) g\rangle.$$

One then defines $\widetilde{T} : E \otimes_\sigma H \to H$ by

$$\widetilde{T}(x \otimes h) = T(x)h.$$

Definition 0.8. *A c.c. representation (σ, T) is called* isometric *if for all $x, y \in E$,*

$$T(x)^* T(y) = \sigma(\langle x, y\rangle).$$

(This is the case if and only if $\widetilde{T}$ is an isometry). It is called fully coisometric *if $\widetilde{T}$ is a coisometry.*

Given two Hilbert C^*-correspondences E and F over $\mathcal{A}$, the *balanced* (or *inner*) tensor product $E \otimes F$ is a Hilbert C^*-correspondence over $\mathcal{A}$ defined to be the Hausdorff completion of the algebraic tensor product with respect to the inner product

$$\langle x \otimes y, w \otimes z\rangle = \langle y, \langle x, w\rangle \cdot z\rangle \ , \ x, w \in E, y, z \in F.$$

The left and right actions are defined as $a \cdot (x \otimes y) = (a \cdot x) \otimes y$ and $(x \otimes y)a = x \otimes (ya)$, respectively, for all $a \in A, x \in E, y \in F$. When working in the context of W^*-correspondences, that is, if E and F are W^*-correspondences and $\mathcal{A}$ is a W^*-algebra, then $E \otimes F$ is understood do be the *self-dual extension* of the above construction.

Detailed overview of the paper. Subproduct systems, their representations, and their units, are defined in the next section. The following two sections, 2 and 3, can be viewed as a reorganization and sharpening of some known results, including several new observations.
Section 2 establishes the correspondence between *cp*-semigroups and subproduct systems. It is shown that given a subproduct system X of $\mathcal{N}$ - correspondences and a subproduct system representation R of X on H, we may construct a *cp*-semigroup Θ acting on $\mathcal{N}'$. We denote this assignment as $\Theta = \Sigma(X, R)$. Conversely, it is shown that given a *cp*-semigroup Θ acting on $\mathcal{M}$, there is a subproduct system E (called the *Arveson-Stinespring subproduct system* of Θ) of $\mathcal{M}'$-correspondences and an *injective* representation T of E on H such that $\Theta = \Sigma(E, T)$. Denoting this assignment as $(E, T) = \Xi(\Theta)$, we have that $\Sigma \circ \Xi$ is the identity. In Theorem 2.6 we show that $\Xi \circ \Sigma$ is also, after restricting to pairs (X, R) with R an injective representation (and up to some "isomorphism"), the identity. This allows us to deduce (Corollary 2.8) that a subproduct system that is not a product system has no isometric representations. We introduce the *Fock spaces* associated to a subproduct system and the canonical *shift representations*. These constructs allow us to show that every subproduct system is the Arveson-Stinespring subproduct system of some *cp*-semigroup.
In Section 3 we briefly sketch the picture that is dual to that of Section 2. It is shown that given a subproduct system and a unit of that subproduct system one may construct a *cp*-semigroup, and that every *cp*-semigroup arises this way.
In Section 4, we construct for every subproduct system X and every fully coisometric subproduct system representation T of X on a Hilbert space, a semigroup $\hat{T}$ of contractions on a Hilbert space that captures "all the information" about X and T. This construction is a modification of the construction introduced in [41] for the case where X is a *product* system. It turns out that when X is merely a *sub*product system, it is hard to apply $\hat{T}$ to obtain new results about the representation T. However, when X is a true *product* system $\hat{T}$ is very handy, and we use it to prove that every e_0-semigroup has a $*$-automorphic dilation (in a certain sense).
Section 5 begins with some general remarks regarding dilations and pieces of subproduct system representations, and then the connection between the dilation theories of *cp*-semigroups and of representations of subproduct systems is made. We define the notion of a *subproduct subsystem* and then we define *dilations* and *pieces* of subproduct system representations. These notions generalize the notions of *commuting piece* or *q-commuting piece* of [12] and [19], and also generalizes the definition of *dilation* of a product system representation of [29]. Proposition 5.8, Theorem 5.12 and Corollary 5.13 show that the

1-1 correspondences Σ and Ξ between *cp*-semigroups and subproduct systems with representations take isometric dilations of representations to e-dilations and vice-versa. This is used to obtain an example of three commuting, unital and contractive CP maps on $B(H)$ for which there exists no e-dilation acting on a $B(K)$, and no *minimal* dilation acting on any von Neumann algebra (Theorem 5.14).

In Section 5 we also present a reduction of both the problem of constructing an e_0-dilation to a cp_0-semigroup, and the problem of constructing an e-dilation to a k-tuple of commuting CP maps with *small enough norm*, to the problem of embedding a subproduct system in a larger *product system*. We show that not every subproduct system can be embedded in a product system (Proposition 5.15), and we use this to construct an example of three commuting CP maps $\theta_1, \theta_2, \theta_3$ such that for *any* $\lambda > 0$ the three-tuple $\lambda\theta_1, \lambda\theta_2, \lambda\theta_3$ has no e-dilation (Theorem 5.16). This unexpected phenomenon has no counterpart in the classical theory of isometric dilations, and provides the first example of a theorem in classical dilation theory that cannot be generalized to the theory of e-dilations of *cp*-semigroups.

The developments described in the first part of the paper indicate that subproduct systems are worthwhile objects of study, but to make progress we must look at plenty of concrete examples. In the second part of the paper we begin studying subproduct systems of Hilbert spaces over the semigroup $\mathbb{N}$. In Section 6 we show that every subproduct system (of W*-correspondences) over $\mathbb{N}$ is isomorphic to a *standard* subproduct system, that is, it is a subproduct subsystem of the full product system $\{E^{\otimes n}\}_{n\in\mathbb{N}}$ for some W*-correspondence E. Using the results of the previous section, this gives a new proof to the discrete analogue of Bhat's Theorem: *every cp_0-semigroup over $\mathbb{N}$ has an e_0-dilation.* Given a subproduct system we define the *standard X-shift*, and we show that if X is a subproduct subsystem of Y, then the standard X-shift is the maximal X-piece of the standard Y-shift, generalizing and unifying results from [12, 19, 39].

In Section 7 we explain why subproduct systems are convenient for studying noncommutative projective algebraic geometry. We show that every homogeneous ideal I in the algebra $\mathbb{C}\langle x_1, \ldots, x_d\rangle$ of noncommutative polynomials corresponds to a unique subproduct system X_I, and vice-versa. The representations of X_I on a Hilbert space H are precisely determined by the d-tuples in the zero set of I,

$$Z(I) = \{\underline{T} = (T_1, \ldots, T_d) \in B(H)^d : \forall p \in I . p(\underline{T}) = 0\}.$$

A noncommutative version of the Nullstellensatz is obtained, stating that

$$\{p \in \mathbb{C}\langle x_1, \ldots, x_d\rangle : \forall \underline{T} \in Z(I) . p(\underline{T}) = 0\} = I.$$

Section 8 starts with a review of a powerful tool, Gelu Popescu's "Poisson Transform" [38]. Using this tool we derive some basic results (obtained previously by Popescu in [39]) which allow us to identify the operator algebra $\mathcal{A}_X$ generated by the X-shift as the universal unital operator algebra generated by a row contraction subject to homogeneous polynomial identities. We then prove that every completely bounded representation of a subproduct system X is a piece of a scaled inflation of the X-shift, and derive a related "von Neumann inequality".

In Section 9 we discuss the relationship between a subproduct system X and $\mathcal{A}_X$, the (non-selfadjoint operator algebra generated by the X-shift). The main result in this section is Theorem 9.7, which says that $X \cong Y$ if and only if $\mathcal{A}_X$ is completely isometrically isomorphic to $\mathcal{A}_Y$ by an isomorphism that preserves the vacuum state. This result is used in Section 10, where we study the universal norm closed unital operator algebra generated by a row contraction $(T_1, \ldots, T_d)$ satisfying the relations

$$T_i T_j = q_{ij} T_j T_i \ , \ 1 \leq i < j \leq d,$$

where $q = (q_{i,j})_{i,j=1}^d \in M_n(\mathbb{C})$ is a matrix such that $q_{j,i} = q_{i,j}^{-1}$. These nonselfadjoint analogues of the noncommutative tori, are shown to be classified by their subproduct systems when $q_{i,j} \neq 1$ for all i, j. In particular, when $d = 2$, we obtain the universal algebra for the relation

$$T_1 T_2 = q T_2 T_1,$$

which we call $\mathcal{A}_q$. It is shown that $\mathcal{A}_q$ is isomorphically isomorphic to $\mathcal{A}_r$ if and only if $q = r$ or $q = r^{-1}$.

In Section 11 we describe all standard maximal subproduct systems X with $\dim X(1) = 2$ and $\dim X(2) = 3$, and classify their algebras up to isometric isomorphisms.

In the closing section of this paper, Section 12, we find that subproduct systems are also closely related to *subshifts* and to the *subshift C^*-algebras* introduced by K. Matsumoto [28]. We show how every subshift gives rise to a subproduct system, and characterize the subproduct systems that come from subshifts. We use this connection together with the results of Section 8 to describe all representations of subshift C*-algebras that come from a subshift of *finite type* (Theorem 12.7).

Acknowledgments. The authors owe their thanks to Eliahu Levy for pointing out a mistake in a previous version of the paper, and to Michael Skeide, for several helpful remarks.

Part 1. Subproduct systems and *cp*-semigroups

1. Subproduct systems of Hilbert W^*-correspondences

Definition 1.1. *Let $\mathcal{N}$ be a von Neumann algebra. A* subproduct system of Hilbert W^*-correspondences *over $\mathcal{N}$ is a family $X = \{X(s)\}_{s\in\mathcal{S}}$ of Hilbert W^*-correspondences over $\mathcal{N}$ such that*

(1) *$X(0) = \mathcal{N}$,*
(2) *For every $s,t \in \mathcal{S}$ there is a coisometric mapping of $\mathcal{N}$-correspondences*
$$U_{s,t} : X(s) \otimes X(t) \to X(s+t),$$
(3) *The maps $U_{s,0}$ and $U_{0,s}$ are given by the left and right actions of $\mathcal{N}$ on $X(s)$,*
(4) *The maps $U_{s,t}$ satisfy the following associativity condition:*

$$U_{s+t,r}\left(U_{s,t} \otimes I_{X(r)}\right) = U_{s,t+r}\left(I_{X(s)} \otimes U_{t,r}\right). \tag{1.1}$$

The difference between a subproduct system and a product system is that in a subproduct system the maps $U_{s,t}$ are only required to be coisometric, while in a product system these maps are required to be unitaries. Thus, given the image $U_{s,t}(x \otimes y)$ of $x \otimes y$ in $X(s+t)$, one cannot recover x and y. Thus, subproduct systems may be thought of as *irreversible* product systems. The terminology is, admittedly, a bit awkward. It may be more sensible – however, impossible at present – to use the term *product system* for the objects described above and to use the term *full product system* for product system.

Example 1.2. The simplest example of a subproduct system $F = F_E = \{F(n)\}_{n\in\mathbb{N}}$ is given by
$$F(n) = E^{\otimes n},$$
where E is some W*-correspondence. F is actually a product system. We shall call this subproduct system *the full product system (over E)*.

Example 1.3. Let E be a fixed Hilbert space. We define a subproduct system (of Hilbert spaces) $SSP = SSP_E$ over $\mathbb{N}$ using the familiar symmetric tensor products (one can obtain a subproduct system from the anti-symmetric tensor products as well). Define
$$E^{\otimes n} = E \otimes \cdots \otimes E,$$
(n times). Let p_n be the projection of $E^{\otimes n}$ onto the symmetric subspace of $E^{\otimes n}$, given by
$$p_n k_1 \otimes \cdots \otimes k_n = \frac{1}{n!}\sum_{\sigma\in S_n} k_{\sigma^{-1}(1)} \otimes \cdots \otimes k_{\sigma^{-1}(n)}.$$
We define
$$SSP(n) = E^{\circledS n} := p_n E^{\otimes n},$$
the symmetric tensor product of E with itself n times ($SSP(0) = \mathbb{C}$). We define a map $U_{m,n} : SSP(m) \otimes SSP(n) \to SSP(m+n)$ by
$$U_{m,n}(x \otimes y) = p_{m+n}(x \otimes y).$$

The U's are coisometric maps because every projection, when considered as a map from its domain onto its range, is coisometric. A straightforward calculation shows that (1.1) holds (see [35, Corollary 17.2]). In these notes we shall refer to SSP (or SSP_E to be precise) as the *symmetric subproduct system (over E)*.

Definition 1.4. *Let X and Y be two subproduct systems over the same semigroup $\mathcal{S}$ (with families of coisometries $\{U^X_{s,t}\}_{s,t\in\mathcal{S}}$ and $\{U^Y_{s,t}\}_{s,t\in\mathcal{S}}$). A family $V = \{V_s\}_{s\in\mathcal{S}}$ of maps $V_s : X(s) \to Y(s)$ is called* a morphism *of subproduct systems if $V_0 : X(0) \to Y(0)$ is the identity, if for all $s \in \mathcal{S}\setminus\{0\}$ the map V_s is a coisometric correspondence map, and if for all $s, t \in \mathcal{S}$ the following identity holds:*

$$V_{s+t} \circ U^X_{s,t} = U^Y_{s,t} \circ (V_s \otimes V_t). \tag{1.2}$$

V is said to be an isomorphism *if V_s is a unitary for all $s \in \mathcal{S}\setminus\{0\}$. X is said to be isomorphic to Y if there exists an isomorphism $V : X \to Y$.*

There is an obvious extension of the above definition to the case where $X(0)$ and $Y(0)$ are $*$-isomorphic. The above notion of *morphism* is not optimized in any way. It is simply precisely what we need in order to develop dilation theory for *cp*-semigroups.

Definition 1.5. *Let $\mathcal{N}$ be a von Neumann algebra, let H be a Hilbert space, and let X be a subproduct system of Hilbert $\mathcal{N}$-correspondences over the semigroup $\mathcal{S}$. Assume that $T : X \to B(H)$, and write T_s for the restriction of T to $X(s)$, $s \in \mathcal{S}$, and σ for T_0. T (or (σ, T)) is said to be a* completely contractive covariant representation *of X if*

(1) *For each $s \in \mathcal{S}$, (σ, T_s) is a c.c. representation of $X(s)$; and*
(2) *$T_{s+t}(U_{s,t}(x \otimes y)) = T_s(x)T_t(y)$ for all $s, t \in \mathcal{S}$ and all $x \in X(s), y \in X(t)$.*

T is said to be an isometric (fully coisometric) representation if it is an isometric (fully coisometric) representation on every fiber $X(s)$.

Since we shall not be concerned with any other kind of representation, we shall call a completely contractive covariant representation of a subproduct system simply a *representation*.

Remark 1.6. Item 2 in the above definition of product system can be rewritten as follows:

$$\widetilde{T}_{s+t}(U_{s,t} \otimes I_H) = \widetilde{T}_s(I_{X(s)} \otimes \widetilde{T}_t).$$

Here $\widetilde{T}_s : X(s) \otimes_\sigma H \to H$ is the map given by

$$\widetilde{T}_s(x \otimes h) = T_s(x)h.$$

Example 1.7. We now define a representation T of the symmetric subproduct system SSP from Example 1.3 on the symmetric Fock space. Denote by $\mathfrak{F}_+$ the symmetric Fock space

$$\mathfrak{F}_+ = \bigoplus_{n\in\mathbb{N}} E^{\circledS n}.$$

For every $n \in \mathbb{N}$, the map $T_n : SSP(n) = E^{\circledS n} \to B(\mathfrak{F}_+)$ is defined on the m-particle space $E^{\circledS m}$ by putting

$$T_n(x)y = p_{n+m}(x \otimes y)$$

for all $x \in X(n), y \in E^{\circledS m}$. Then T extends to a representation of the subproduct system SSP on $\mathfrak{F}_+$ (to see that item 2 of Definition 1.5 is satisfied one may use again [35, Corollary 17.2]).

Definition 1.8. *Let $X = \{X(s)\}_{s \in \mathcal{S}}$ be a subproduct system of $\mathcal{N}$-correspondences over $\mathcal{S}$. A family $\xi = \{\xi_s\}_{s \in \mathcal{S}}$ is called a* unit *for X if*

$$\xi_s \otimes \xi_t = U_{s,t}^* \xi_{s+t}. \tag{1.3}$$

A unit $\xi = \{\xi_s\}_{s \in \mathcal{S}}$ is called unital *if $\langle \xi_s, \xi_s \rangle = 1_{\mathcal{N}}$ for all $s \in \mathcal{S}$, it is called* contractive *if $\langle \xi_s, \xi_s \rangle \leq 1_{\mathcal{N}}$ for all $s \in \mathcal{S}$, and it is called* generating *if $X(s)$ is spanned by elements of the form*

(1.4)
$$U_{s_1+\cdots+s_{n-1},s_n}(\cdots U_{s_1+s_2,s_3}(U_{s_1,s_2}(a_1\xi_{s_1} \otimes a_2\xi_{s_2}) \otimes a_3\xi_{s_3}) \otimes \cdots \otimes a_n\xi_{s_n}a_{n+1}),$$

where $s = s_1 + s_2 + \cdots + s_n$.

From (1.3) follows the perhaps more natural looking

$$U_{s,t}(\xi_s \otimes \xi_t) = \xi_{s+t}.$$

Example 1.9. A unital unit for the symmetric subproduct system SSP from Example 1.3 is given by defining $\xi_0 = 1$ and

$$\xi_n = \underbrace{v \otimes v \otimes \cdots \otimes v}_{n \text{ times}}$$

for $n \geq 1$. This unit is generating only if E is one dimensional.

2. Subproduct system representations and *cp*-semigroups

In this section, following Muhly and Solel's constructions from [29], we show that subproduct systems and their representations provide a framework for dealing with *cp*-semigroups, and allow us to obtain a generalization of the classical result of Wigner that any strongly continuous one-parameter group of automorphisms of $B(H)$ is given by $X \mapsto U_t X U_t^*$ for some one-parameter unitary group $\{U_t\}_{t \in \mathbb{R}}$.

2.1. All *cp*-semigroups come from subproduct system representations.

Proposition 2.1. *Let $\mathcal{N}$ be a von Neumann algebra and let X be a subproduct system of $\mathcal{N}$-correspondences over $\mathcal{S}$, and let R be completely contractive covariant representation of X on a Hilbert space H, such that R_0 is unital. Then the family of maps*

$$\Theta_s : a \mapsto \widetilde{R}_s(I_{X(s)} \otimes a)\widetilde{R}_s^* \ , \ a \in R_0(\mathcal{N})', \tag{2.1}$$

is a semigroup of CP maps on $R_0(\mathcal{N})'$. Moreover, if R is an isometric (a fully coisometric) representation, then Θ_s is a $*$-endomorphism (a unital map) for all $s \in \mathcal{S}$.

Proof. By Proposition 2.21 in [29], $\{\Theta_s\}_{s\in\mathcal{S}}$ is a family of contractive, normal, completely positive maps on $R_0(\mathcal{N})'$. Moreover, these maps are unital if R is a fully coisometric representation, and they are $*$-endomorphisms if R is an isometric representation. It remains is to check that $\Theta = \{\Theta_s\}_{s\in\mathcal{S}}$ satisfies the semigroup condition $\Theta_s \circ \Theta_t = \Theta_{s+t}$. Fix $a \in R_0(\mathcal{N})'$. For all $s, t \in \mathcal{S}$,

$$\begin{aligned}
\Theta_s(\Theta_t(a)) &= \widetilde{R}_s \left(I_{X(s)} \otimes \left(\widetilde{R}_t (I_{X(t)} \otimes a) \widetilde{R}_t^* \right) \right) \widetilde{R}_s^* \\
&= \widetilde{R}_s (I_{X(s)} \otimes \widetilde{R}_t)(I_{X(s)} \otimes I_{X(t)} \otimes a)(I_{X(s)} \otimes \widetilde{R}_t^*) \widetilde{R}_s^* \\
&= \widetilde{R}_{s+t} (U_{s,t} \otimes I_G)(I_{X(s)} \otimes I_{X(t)} \otimes a)(U_{s,t}^* \otimes I_G) \widetilde{R}_{s+t}^* \\
&= \widetilde{R}_{s+t} (I_{X(s\cdot t)} \otimes a) \widetilde{R}_{s+t}^* \\
&= \Theta_{s+t}(a).
\end{aligned}$$

Using the fact that R_0 is unital, we have

$$\begin{aligned}
\Theta_0(a)h &= \widetilde{R_0}(I_{\mathcal{N}} \otimes a)\widetilde{R_0}^* h \\
&= \widetilde{R_0}(I_{\mathcal{N}} \otimes a)(1_{\mathcal{N}} \otimes h) \\
&= R_0(1_{\mathcal{N}})ah \\
&= ah,
\end{aligned}$$

thus $\Theta_0(a) = a$ for all $a \in \mathcal{N}$. □

We will now show that *every* *cp*-semigroup is given by a subproduct representation as in (2.1) above. We recall some constructions from [29] (building on the foundations set in [4]).

Fix a CP map Θ on von Neumann algebra $\mathcal{M} \subseteq B(H)$. We define $\mathcal{M} \otimes_\Theta H$ to be the Hausdorff completion of the algebraic tensor product $\mathcal{M} \otimes H$ with respect to the sesquilinear positive semidefinite form

$$\langle T_1 \otimes h_1, T_2 \otimes h_2 \rangle = \langle h_1, \Theta(T_1^* T_2) h_2 \rangle .$$

We define a representation π_Θ of $\mathcal{M}$ on $\mathcal{M} \otimes_\Theta H$ by

$$\pi_\Theta(S)(T \otimes h) = ST \otimes h,$$

and we define a (contractive) linear map $W_\Theta : H \to \mathcal{M} \otimes H$ by

$$W_\Theta(h) = I \otimes h.$$

If Θ is unital then W_Θ is an isometry, and if Θ is an endomorphism then W_Θ is a coisometry. The adjoint of W_Θ is given by

$$W_\Theta^*(T \otimes h) = \Theta(T)h.$$

For a given CP semigroup Θ on $\mathcal{M}$, Muhly and Solel defined in [29] a W^*-correspondence E_Θ over $\mathcal{M}'$ and a c.c. representation (σ, T_Θ) of E_Θ on H such that for all $a \in \mathcal{M}$

$$\Theta(a) = \widetilde{T}_\Theta \left(I_{E_\Theta} \otimes a\right) \widetilde{T}_\Theta^*. \tag{2.2}$$

The W^*-correspondence E_Θ is defined as the intertwining space

$$E_\Theta = \mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_\Theta H),$$

where

$$\mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_\Theta H) := \{X \in B(H, \mathcal{M} \otimes_\Theta H) \big| \; XT = \pi_\Theta(T)X, \; T \in \mathcal{M}\}.$$

The left and right actions of $\mathcal{M}'$ are given by

$$S \cdot X = (I \otimes S)X \quad , \quad X \cdot S = XS$$

for all $X \in E_\Theta$ and $S \in \mathcal{M}'$. The $\mathcal{M}'$-valued inner product on E_Θ is defined by $\langle X, Y \rangle = X^*Y$. E_Θ is called *the Arveson-Stinespring correspondence* (associated with Θ).
The representation (σ, T_Θ) is defined by letting $\sigma = \mathbf{id}_{\mathcal{M}'}$, the identity representation of $\mathcal{M}'$ on H, and by defining

$$T_\Theta(X) = W_\Theta^* X.$$

$(\mathbf{id}_{\mathcal{M}'}, T_\Theta)$ is called *the identity representation* (associated with Θ). We remark that the paper [29] focused on unital CP maps, but the results we cite are true for nonunital CP maps, with the proofs unchanged.
The case where $\mathcal{M} = B(H)$ in the following theorem appears, in essence at least, in [4].

Theorem 2.2. *Let $\Theta = \{\Theta_s\}_{s \in \mathcal{S}}$ be a cp-semigroup on a von Neumann algebra $\mathcal{M} \subseteq B(H)$, and for all $s \in \mathcal{S}$ let $E(s) := E_{\Theta_s}$ be the Arveson-Stinespring correspondence associated with Θ_s, and let $T_s := T_{\Theta_s}$ denote the identity representation for Θ_s. Then $E = \{E(s)\}_{s \in \mathcal{S}}$ is a subproduct system of $\mathcal{M}'$-correspondences, and $(\mathbf{id}_{\mathcal{M}'}, T)$ is a representation of E on H that satisfies*

$$\Theta_s(a) = \widetilde{T}_s \left(I_{E(s)} \otimes a\right) \widetilde{T}_s^* \tag{2.3}$$

for all $a \in \mathcal{M}$ and all $s \in \mathcal{S}$. T_s is injective for all $s \in \mathcal{S}$. If Θ is an e-semigroup (cp$_0$-semigroup), then $(\mathbf{id}_{\mathcal{M}'}, T)$ is isometric (fully coisometric).

Proof. We begin by defining the subproduct system maps $U_{s,t} : E(s) \otimes E(t) \to E(s+t)$. We use the constructions made in [29, Proposition 2.12] and the surrounding discussion. We define

$$U_{s,t} = V_{s,t}^* \Psi_{s,t} \; ,$$

where the map

$$\Psi_{s,t} : \mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_{\Theta_s} H) \otimes \mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_{\Theta_t} H) \to \mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_{\Theta_t} \mathcal{M} \otimes_{\Theta_s} H)$$

is given by $\Psi_{s,t}(X \otimes Y) = (I \otimes X)Y$, and

$$V_{s,t} : \mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_{\Theta_{s+t}} H) \to \mathcal{L}_\mathcal{M}(H, \mathcal{M} \otimes_{\Theta_t} \mathcal{M} \otimes_{\Theta_s} H)$$

is given by $V_{s,t}(X) = \Gamma_{s,t} \circ X$, where $\Gamma_{s,t} : \mathcal{M} \otimes_{\Theta_{s+t}} H \to \mathcal{M} \otimes_{\Theta_t} \mathcal{M} \otimes_{\Theta_s} H$ is the isometry

$$\Gamma_{s,t} : S \otimes_{\Theta_{s+t}} h \mapsto S \otimes_{\Theta_t} I \otimes_{\Theta_s} h.$$

By [29, Proposition 2.12], $U_{s,t}$ is a coisometric bimodule map for all $s,t \in \mathcal{S}$. To see that the U's compose associatively as in (1.1), take $s,t,u \in \mathcal{S}$, $X \in E(s), Y \in E(t), Z \in E(u)$, and compute:

$$\begin{aligned}
U_{s,t+u}(I_{E(s)} \otimes U_{t,u})(X \otimes Y \otimes Z) &= U_{s,t+u}(X \otimes V_{t,u}^*(I \otimes Y)Z) \\
&= V_{s,t+u}^* \big((I \otimes X)V_{t,u}^*(I \otimes Y)Z\big) \\
&= \Gamma_{s,t+u}^*(I \otimes X)\Gamma_{t,u}^*(I \otimes Y)Z
\end{aligned}$$

and

$$\begin{aligned}
U_{s+t,u}(U_{s,t} \otimes I_{E(u)})(X \otimes Y \otimes Z) &= U_{s+t,u}(V_{s,t}^*(I \otimes X)Y \otimes Z) \\
&= V_{s+t,u}^* \big((I \otimes V_{s,t}^*(I \otimes X)Y)Z\big) \\
&= \Gamma_{s+t,u}^*(I \otimes \Gamma_{s,t}^*)(I \otimes I \otimes X)(I \otimes Y)Z\ .
\end{aligned}$$

So it suffices to show that

$$\Gamma_{s,t+u}^*(I \otimes X)\Gamma_{t,u}^* = \Gamma_{s+t,u}^*(I \otimes \Gamma_{s,t}^*)(I \otimes I \otimes X)$$

It is easier to show that their adjoints are equal. Let $a \otimes h$ be a typical element of $\mathcal{M} \otimes_{\Theta_{s+t+u}} h$.

$$\begin{aligned}
\Gamma_{t,u}(I \otimes X^*)\Gamma_{s,t+u}(a \otimes_{\Theta_{s+t+u}} h) &= \Gamma_{t,u}(I \otimes X^*)(a \otimes_{\Theta_{t+u}} I \otimes_{\Theta_s} h) \\
&= \Gamma_{t,u}(a \otimes_{\Theta_{t+u}} X^*(I \otimes_{\Theta_s} h)) \\
&= a \otimes_{\Theta_u} I \otimes_{\Theta_t} X^*(I \otimes_{\Theta_s} h).
\end{aligned}$$

On the other hand

$$\begin{aligned}
&(I \otimes I \otimes X^*)(I \otimes \Gamma_{s,t})\Gamma_{s+t,u}(a \otimes_{\Theta_{s+t+u}} h) = \\
&\qquad = (I \otimes I \otimes X^*)(I \otimes \Gamma_{s,t})(a \otimes_{\Theta_u} I \otimes_{\Theta_{s+t}} h) \\
&\qquad = (I \otimes I \otimes X^*)(a \otimes_{\Theta_u} I \otimes_{\Theta_t} I \otimes_{\Theta_s} h) \\
&\qquad = a \otimes_{\Theta_u} I \otimes_{\Theta_t} X^*(I \otimes_{\Theta_s} h).
\end{aligned}$$

This shows that the maps $\{U_{s,t}\}$ make E into a subproduct system.
Let us now verify that T is a representation of subproduct systems. That $(\mathbf{id}_{\mathcal{M}'}, T_s)$ is a c.c. representation of $E(s)$ is explained in [29, page 878]. Let $X \in E(s), Y \in E(t)$.

$$T_{s+t}(U_{s,t}(X \otimes Y)) = W_{\Theta_{s+t}}^* \Gamma_{s,t}^*(I \otimes X)Y,$$

while

$$T_s(X)T_t(Y) = W_{\Theta_s}^* X W_{\Theta_t}^* Y.$$

But for all $h \in H$,

$$\begin{aligned}
W_{\Theta_t} X^* W_{\Theta_s} h &= W_{\Theta_t} X^*(I \otimes_{\Theta_s} h) \\
&= I \otimes_{\Theta_t} X^*(I \otimes_{\Theta_s} h) \\
&= (I \otimes X^*)(I \otimes_{\Theta_t} I \otimes_{\Theta_s} h) \\
&= (I \otimes X^*)\Gamma_{s,t}(I \otimes_{\Theta_{s+t}} h) \\
&= (I \otimes X^*)\Gamma_{s,t} W_{\Theta_{s+t}} h,
\end{aligned}$$

which implies $W^*_{\Theta_s} X W^*_{\Theta_t} Y = W^*_{\Theta_{s+t}} \Gamma^*_{s,t}(I \otimes X)Y$, so we have the desired equality

$$T_{s+t}(U_{s,t}(X \otimes Y)) = T_s(X)T_t(Y).$$

Equation (2.3) is a consequence of (2.2). The injectivity of the identity representation has already been noted by Solel in [48] (for all $h, g \in H$ and $a \in M$, $\langle W^*_\Theta X a^* h, g\rangle = \langle X a^* h, I \otimes g\rangle = \langle (I \otimes a^*)Xh, I \otimes g\rangle = \langle Xh, a \otimes g\rangle$). The final assertion of the theorem is trivial (if Θ_s is a $*$-endomorphism, then W_{Θ_s} is a coisometry). □

Definition 2.3. *Given a cp-semigroup Θ on a W^* algebra $\mathcal{M}$, the pair (E, T) constructed in Theorem 2.2 is called* the identity representation *of Θ, and E is called* the Arveson-Stinespring subproduct system *associated with Θ.*

Remark 2.4. If follows from [29, Proposition 2.14], if Θ is an e-semigroup, then the identity representation subproduct system is, in fact, a (full) product system.

Remark 2.5. In [27], Daniel Markiewicz has studied the Arveson-Stinespring subproduct system of a CP_0-semigroup over $\mathbb{R}_+$ acting on $B(H)$, and has also shown that it carries a structure of a *measurable Hilbert bundle*.

2.2. Essentially all injective subproduct system representations come from *cp*-semigroups. The following generalizes and is motivated by [48, Proposition 5.7]. We shall also repeat some arguments from [32, Theorem 2.1].

By Theorem 2.2, with every *cp*-semigroup $\Theta = \{\Theta_s\}_{s \in \mathcal{S}}$ on $\mathcal{M} \subseteq B(H)$ we can associate a pair (E, T) - the identity representation of Θ - consisting of a subproduct system E (of correspondences over $\mathcal{M}'$) and an injective subproduct system c.c. representation T. Let us write $(E, T) = \Xi(\Theta)$. Conversely, given a pair (X, R) consisting of a subproduct system X of correspondences over $\mathcal{M}'$ and a c.c. representation R of X such that $R_0 = id$, one may define by equation (2.1) a *cp*-semigroup Θ acting on $\mathcal{M}$, which we denote as $\Theta = \Sigma(X, R)$. The meaning of equation (2.3) is that $\Sigma \circ \Xi$ is the identity map on the set of *cp*-semigroups of $\mathcal{M}$. We will show below that $\Xi \circ \Sigma$, when restricted to pairs (X, R) such that R is injective, is also, essentially, the identity. When (X, R) is not injective, we will show that $\Xi \circ \Sigma(X, R)$ "sits inside" (X, R).

Theorem 2.6. *Let $\mathcal{N}$ be a W^*-algebra, let $X = \{X(s)\}_{s \in \mathcal{S}}$ be a subproduct system of $\mathcal{N}$-correspondences, and let R be a c.c. representation of X on H, such that $\sigma := R_0$ is faithful and nondegenerate. Let $\mathcal{M} \subseteq B(H)$ be the commutant $\sigma(\mathcal{N})'$ of $\sigma(\mathcal{N})$. Let $\Theta = \Sigma(X, R)$, and let $(E, T) = \Xi(\Theta)$. Then there is a morphism of subproduct systems $W : X \to E$ such that*

$$R_s = T_s \circ W_s \ , \ s \in \mathcal{S}. \tag{2.4}$$

$W_s^* W_s = I_{X(s)} - q_s$, where q_s is the orthogonal projection of $X(s)$ onto $Ker R_s$. In particular, W is an isomorphism if and only if R_s is injective for all $s \in \mathcal{S}$.

Remark 2.7. The construction of the morphism W below basically comes from [48, 32], and it remains only to show that it respects the subproduct system structure. Some of the details in the proof will be left out.

Proof. We may construct a subproduct system X' of $\mathcal{M}'$-correspondences (recall that $\mathcal{M}' = \sigma(\mathcal{N})$), and a representation R' of X' on H such that R'_0 is the identity, in such a way that (X, R) may be naturally identified with (X', R'). Indeed, put

$$X'(0) = \mathcal{M}' \ , \ X'(s) = X(s) \text{ for } s \neq 0,$$

where the inner product is defined by

$$\langle x, y\rangle_{X'} = \sigma(\langle x, y\rangle_X),$$

and the left and right actions are defined by

$$a \cdot x \cdot b := \sigma^{-1}(a)x\sigma^{-1}(b),$$

for $a, b \in \mathcal{M}'$ and $x, y \in X'(s)$, $s \in \mathcal{S} \setminus \{0\}$. Defining $R'_0 = id$ and $W_0 = \sigma$; and $R'_s = R_s$ for and $W_s = id$ for $s \in \mathcal{S} \setminus \{0\}$, we have that W is a morphism $X \to X'$ that sends R to R'.
Assume, therefore, that $\mathcal{N} = \mathcal{M}'$ and that σ is the identity representation.
We begin by defining for every $s \neq 0$

$$v_s : \mathcal{M} \otimes_{\Theta_s} H \to X(s) \otimes H$$

by

$$v_s(a \otimes h) = (I_{X(s)} \otimes a)\widetilde{R}_s^* h.$$

It is straightforward to show that for all $s \in \mathcal{S}$ the map v_s is a well-defined isometry. $[(I_{X(s)} \otimes \mathcal{M})\widetilde{R}_s^* H]$ is invariant under $I_{X(s)} \otimes \mathcal{M}$, thus the projection onto the orthocomplement of this subspace is in $(I_{X(s)} \otimes \mathcal{M})' = \mathcal{L}(X(s)) \otimes I_H$, so it has the form $q_s \otimes I_H$ for some projection $q_s \in \mathcal{L}(X(s))$. In fact, it is easy to check that q_s is the orthogonal projection of $X(s)$ onto $\mathrm{Ker} R_s$.
By the definition of v_s and by the covariance properties of T, we have for all $a \in \mathcal{M}$ and $b \in \mathcal{M}'$,

$$v_s(a \otimes I) = (I \otimes a)v_s \ , \ v_s(I \otimes b) = (b \otimes I)v_s.$$

Fix $s \in \mathcal{S}$ and $x \in E(s)$. For all $\xi \in X(s), h \in H$, write

$$\psi(\xi)h = x^* v_s^*(\xi \otimes h).$$

It is easy to verify that the linear map $\xi \mapsto \psi(\xi)$ maps $X(s)$ into $\mathcal{M}'$ and is a bounded right module map. From the self duality of $X(s)$ it follows that there is a unique element in $X(s)$, which we denote by $V_s(x)$, such that for all $\xi \in X(s), h \in H$,

$$\langle V_s(x), \xi\rangle h = x^* v_s^*(\xi \otimes h). \tag{2.5}$$

The map V_s is then a linear bimodule map. To show that V_s preserves inner products, write L_ξ, $\xi \in X(s)$, for the operator $L_\xi : H \to X(s) \otimes H$ that maps h to $\xi \otimes h$ and note that equation (2.5) becomes

$$L_{V_s(x)}^* L_\xi = x^* v_s^* L_\xi \ , \ \xi \in X(s),$$

or $L_{V_s(x)} = v_s x$, for all $x \in E(s)$. But since v_s preserves inner products, we have for all $x, y \in E(s)$:

$$\langle x, y\rangle = x^* y = x^* v_s^* v_s y = L^*_{V_s(x)} L_{V_s(y)} = \langle V_s(x), V_s(y)\rangle.$$

We now prove that $V_s V_s^* = I_{X(s)} - q_s$. $\xi \in \mathrm{Im} V_s^{\perp}$ if and only if $L_\xi^* v_s E(s) H = 0$. But by [29, Lemma 2.10], $E(s)H = \mathcal{M} \otimes_{\Theta_s} H$, thus $L_\xi^* v_s E(s) H = 0$ if and only if $\langle \xi, \eta\rangle = 0$ for all $\eta \in (I_{X(s)} - q_s)X(s)$, which is the same as $\xi \in q_s X(s)$.
Fix $h, k \in H$. For $x \in E(s)$, we compute:

$$\begin{aligned}
\langle T_s(x)h, k\rangle &= \langle W^*_{\Theta_s} x h, k\rangle \\
&= \langle xh, I \otimes_{\Theta_s} k\rangle \\
&= \langle v_s x h, v_s (I \otimes_{\Theta_s} k)\rangle \\
&= \langle V_s(x) \otimes h, \widetilde{R}_s^* k\rangle \\
&= \langle R_s(V_s(x))h, k\rangle,
\end{aligned}$$

thus $T_s = R_s \circ V_s$ for all $s \in \mathcal{S}$. Define $W_s = V_s^*$. Then $T_s = R_s \circ W_s^*$. Multiplying both sides by W_s we obtain $T_s \circ W_s = R_s \circ W_s^* W_s$. But $W_s^* W_s = I - q_s$ is the orthogonal projection onto $(\mathrm{Ker} R_s)^{\perp}$, thus we obtain (2.4).
Finally, we need to show that $W = \{W_s\}$ respects the subproduct system structure: for all $s, t \in \mathcal{S}$, $x \in X(s)$ and $y \in X(t)$, we must show that

$$W_{s+t}(U^X_{s,t}(x \otimes y)) = U^E_{s,t}(W_s(x) \otimes W_t(y)).$$

Since T_{s+t} is injective, it is enough to show that after applying T_{s+t} to both sides of the above equation we get the same thing. But T_{s+t} applied to the left hand side gives

$$T_{s+t} W_{s+t}(U^X_{s,t}(x \otimes y)) = R_{s+t}(U^X_{s,t}(x \otimes y)) = R_s(x) R_t(y),$$

and T_{s+t} applied to the right hand side gives

$$T_{s+t}(U^E_{s,t}(W_s(x) \otimes W_t(y))) = T_s(W_s(x)) T_t(W_t(y)) = R_s(x) R_t(y).$$

□

Corollary 2.8. *Let X be a subproduct system that has an isometric representation V such that V_0 is faithful and nondegenerate. Then X is a (full) product system.*

Proof. Let $\Theta = \Sigma(X, V)$. Then Θ is an e-semigroup. Thus, if $(E, T) = \Xi(\Theta)$ is the identity representation of Θ, then, by Remark 2.4, E is a (full) product system. But if V_0 is faithful and V is isometric then V is injective. By the above theorem, X is isomorphic to E, so it is a product system. □

2.3. Subproduct systems arise from *cp*-semigroups. The shift representation. A question arises: *does every subproduct system arise as the Arveson-Stinespring subproduct system associated with a cp-semigroup?* By Theorem 2.6, this is equivalent to the question *does every subproduct system*

have an injective representation? We shall answer this question in the affirmative by constructing for every such subproduct system a canonical injective representation.
The following constructs will be of most interest when $\mathcal{S}$ is a countable semigroup, such as $\mathbb{N}^k$.

DEFINITION 2.9. *Let $X = \{X(s)\}_{s\in\mathcal{S}}$ be a subproduct system. The* X-Fock space $\mathfrak{F}_X$ *is defined as*

$$\mathfrak{F}_X = \bigoplus_{s\in\mathcal{S}} X(s).$$

The vector $\Omega := 1 \in \mathcal{N} = X(0)$ is called the vacuum vector of $\mathfrak{F}_X$. *The* X-shift *representation of X on $\mathfrak{F}_X$ is the representation*

$$S^X : X \to B(\mathfrak{F}_X),$$

given by $S^X(x)y = U^X_{s,t}(x \otimes y)$, for all $x \in X(s), y \in X(t)$ and all $s,t \in \mathcal{S}$.

Strictly speaking, S^X as defined above is not a representation because it represents X on a C*-correspondence rather than on a Hilbert space. However, since for any C*-correspondence E, $\mathcal{L}(E)$ is a C*-algebra, one can compose a faithful representation $\pi : \mathcal{L}(E) \to B(H)$ with S^X to obtain a representation on a Hilbert space.
A direct computation shows that $\widetilde{S}^X_s : X(s) \otimes \mathfrak{F}_X \to \mathfrak{F}_X$ is a contraction, and also that $S^X(x)S^X(y) = S^X(U^X_{s,t}(x \otimes y))$ so S^X is a completely contractive representation of X. S^X is also injective because $S^X(x)\Omega = x$ for all $x \in X$. Thus,

COROLLARY 2.10. *Every subproduct system is the Arveson-Stinespring subproduct system of a cp-semigroup.*

3. SUBPRODUCT SYSTEM UNITS AND *cp*-SEMIGROUPS

In this section, following Bhat and Skeide's constructions from [15], we show that subproduct systems and their units may also serve as a tool for studying *cp*-semigroups.

PROPOSITION 3.1. *Let $\mathcal{N}$ be a von Neumann algebra and let X be a subproduct system of $\mathcal{N}$-correspondences over $\mathcal{S}$, and let $\xi = \{\xi_s\}_{s\in\mathcal{S}}$ be a contractive unit of X, such that $\xi_0 = 1_{\mathcal{N}}$. Then the family of maps*

$$\Theta_s : a \mapsto \langle \xi_s, a\xi_s \rangle \,, \tag{3.1}$$

is a semigroup of CP maps on $\mathcal{N}$. Moreover, if ξ is unital, then Θ_s is a unital map for all $s \in \mathcal{S}$.

Proof. It is standard that Θ_s given by (3.1) is a contractive completely positive map on $\mathcal{N}$, which is unital if and only if ξ is unital. The fact that Θ_s is normal goes a little bit deeper, but is also known (one may use [29, Remark 2.4(i)]).

We show that $\{\Theta_s\}_{s\in\mathcal{S}}$ is a semigroup. It is clear that $\Theta_0(a) = a$ for all $a \in \mathcal{N}$. For all $s,t \in \mathcal{S}$,

$$\begin{aligned}\Theta_s(\Theta_t(a)) &= \langle \xi_s, \langle \xi_t, a\xi_t\rangle \xi_s\rangle \\ &= \langle \xi_t \otimes \xi_s, , a\xi_t \otimes \xi_s\rangle \\ &= \left\langle U^*_{t,s}\xi_{s+t}, , aU^*_{t,s}\xi_{s+t}\right\rangle \\ &= \langle \xi_{s+t}, a\xi_{s+t}\rangle \\ &= \Theta_{s+t}(a).\end{aligned}$$

□

We recall a central construction in Bhat and Skeide's approach to dilation of *cp*-semigroup [15], that goes back to Paschke [36]. Let $\mathcal{M}$ be a W^*-algebra, and let Θ be a normal completely positive map on $\mathcal{M}$ [3]. *The GNS representation of* Θ is a pair (F_Θ, ξ_Θ) consisting of a Hilbert W^*-correspondence F_Θ and a vector $\xi_\Theta \in F_\Theta$ such that

$$\Theta(a) = \langle \xi_\Theta, a\xi_\Theta\rangle \quad \text{for all } a \in \mathcal{M}.$$

F_Θ is defined to be the correspondence $\mathcal{M} \otimes_\Theta \mathcal{M}$ - which is the self-dual extension of the Hausdorff completion of the algebraic tensor product $\mathcal{M} \otimes \mathcal{M}$ with respect to the inner product

$$\langle a\otimes b, c\otimes d\rangle = b^*\Theta(a^*c)d.$$

ξ_Θ is defined to be $\xi_\Theta = 1 \otimes 1$. Note that ξ_Θ is a unit vector, that is - $\langle \xi_\Theta, \xi_\Theta\rangle = 1$, if and only if Θ is unital.

Theorem 3.2. *Let $\Theta = \{\Theta_s\}_{s\in\mathcal{S}}$ be a cp-semigroup on a W^*-algebra $\mathcal{M}$. For every $s \in \mathcal{S}$ let $(F(s), \xi_s)$ be the GNS representation of Θ_s. Then $F = \{F(s)\}_{s\in\mathcal{S}}$ is a subproduct system of $\mathcal{M}$-correspondences, and $\xi = \{\xi_s\}_{s\in\mathcal{S}}$ is a generating contractive unit for F that gives back Θ by the formula*

$$\Theta_s(a) = \langle \xi_s, a\xi_s\rangle \quad \textit{for all } a \in \mathcal{M}. \tag{3.2}$$

Θ is a cp$_0$-semigroup if and only if ξ is a unital unit.

Proof. For all $s,t \in \mathcal{S}$ define a map $V_{s,t} : F(s+t) \to F(s) \otimes F(t)$ by sending ξ_{s+t} to $\xi_s \otimes \xi_t$ and extending to a bimodule map. Because

$$\begin{aligned}\langle a\xi_s \otimes \xi_t b, c\xi_s \otimes \xi_t d\rangle &= \langle \xi_t b, \langle a\xi_s, c\xi_s\rangle \xi_t d\rangle \\ &= \langle \xi_t b, \Theta_s(a^*c)\xi_t d\rangle \\ &= b^* \langle \xi_t, \Theta_s(a^*c)\xi_t\rangle d \\ &= b^*\Theta_{t+s}(a^*c)d \\ &= \langle a\xi_{t+s} b, c\xi_{t+s} d\rangle ,\end{aligned}$$

$V_{s,t}$ extends to a well defined isometric bimodule map from $F(s+t)$ into $F(s)\otimes F(t)$. We define the map $U_{s,t}$ to be the adjoint of $V_{s,t}$ (here it is important that

[3] The construction works also for completely positive maps on unital C*-algebras, but in Theorem 3.2 below we will need to work with normal maps on W*-algebras.

we are working with W^* algebras - in general, an isometry from one Hilbert C^*-module into another need not be adjointable, but bounded module maps between *self-dual* Hilbert modules are always adjointable, [36, Proposition 3.4]). The collection $\{U_{s,t}\}_{s,t\in\mathcal{S}}$ makes F into a subproduct system. Indeed, these maps are coisometric by definition, and they compose in an associative manner. To see the latter, we check that $(I_{F(r)}\otimes V_{s,t})V_{r,s+t} = (V_{r,s}\otimes I_{F(t)})V_{r+s,t}$ and take adjoints.

$$\begin{aligned}(I_{F(r)}\otimes V_{s,t})V_{r,s+t}(a\xi_{r+s+t}b) &= (I_{F(r)}\otimes V_{s,t})(a\xi_r\otimes\xi_{s+t}b)\\ &= a\xi_r\otimes\xi_s\otimes\xi_t b.\end{aligned}$$

Similarly, $(V_{r,s}\otimes I_{F(t)})V_{r+s,t}(a\xi_{r+s+t}b) = a\xi_r\otimes\xi_s\otimes\xi_t b$. Since $F(r+s+t)$ is spanned by linear combinations of elements of the form $a\xi_{r+s+t}b$, the U's make F into a subproduct system, and ξ is certainly a unit for F. Equation (3.2) follows by definition of the GNS representation. Now,

$$\langle\xi_s,\xi_s\rangle = \Theta_s(1)\ ,\ s\in\mathcal{S},$$

so ξ is a contractive unit because $\Theta_s(1)\leq 1$, and ξ it is unital if and only if Θ_s is unital for all s. ξ is in fact more then just a generating unit, as $F(s)$ is spanned by elements with the form described in equation (1.4) with $(s_1,\ldots,s_n)$ a *fixed* n-tuple such that $s_1+\cdots+s_n=s$. $\square$

DEFINITION 3.3. *Given a cp-semigroup Θ on a W^* algebra $\mathcal{M}$, the subproduct system F and the unit ξ constructed in Theorem 3.2 are called, respectively,* the GNS subproduct system *and* the GNS unit *of Θ. The pair (F,ξ) is called* the GNS representation *of Θ.*

REMARK 3.4. There is a precise relationship between the identity representation (Definition 2.3) and the GNS representation of a *cp*-semigroup. The GNS representation of a CP map is the *dual* of the identity representation in a sense that is briefly described in [31]. This notion of duality has been used to move from the product-system-and-representation picture to the product-system-with-unit picture, and vice versa. See for example [44] and the references therein. It is more-or-less straightforward to use this duality to get Theorem 3.2 from Theorem 2.2 (or the other way around).

4. $*$-AUTOMORPHIC DILATION OF AN e_0-SEMIGROUP

We now apply some of the tools developed above to dilate an e_0-semigroup to a semigroup of $*$-automorphisms. We shall need the following proposition, which is a modification (suited for *sub*product systems) of the method introduced in [41] for representing a product system representation as a semigroup of contractive operators on a Hilbert space.

PROPOSITION 4.1. *Let $\mathcal{N}$ be a von Neumann algebra and let X be a subproduct system of W^*-correspondences over $\mathcal{S}$. Let (σ,T) be a fully coisometric covariant representation of X on the Hilbert space H, and assume that σ is unital.*

Denote

$$H_s := \big(X(s) \otimes_\sigma H\big) \Big/ \mathrm{Ker}\widetilde{T}_s$$

and

$$\mathcal{H} = \bigoplus_{s \in \mathcal{S}} H_s .$$

Then there exists a semigroup of coisometries $\hat{T} = \{\hat{T}_s\}_{s \in \mathcal{S}}$ *on* $\mathcal{H}$ *such that for all* $s \in \mathcal{S}$, $x \in X(s)$ *and* $h \in H$,

$$\hat{T}_s \left(\delta_s \cdot x \otimes h\right) = T_s(x)h.$$

$\hat{T}$ *also has the property that for all* $s \in \mathcal{S}$ *and all* $t \geq s$

$$\hat{T}_s^* \hat{T}_s \big|_{H_t} = I_{H_t} \quad , \quad (t \geq s). \tag{4.1}$$

Proof. First, we note that the assumptions on σ and on the left action of $\mathcal{N}$ imply that $H_0 \cong H$ via the identification $a \otimes h \leftrightarrow \sigma(a)h$. This identification will be made repeatedly below.
Define $\mathcal{H}_0$ to be the space of all finitely supported functions f on $\mathcal{S}$ such that for all $s \in \mathcal{S}$,

$$f(s) \in H_s .$$

$\mathcal{H}_0$ is generated by the functions $\delta_s \cdot \xi$, where δ_s is the function taking the value 1 at s and 0 elsewhere, and $\xi \in H_s$. We equip $\mathcal{H}_0$ with the inner product

$$\langle \delta_s \cdot \xi, \delta_t \cdot \eta \rangle = \delta_{s,t} \langle \xi, \eta \rangle ,$$

for all $s, t \in \mathcal{S}, \xi \in H_s, \eta \in H_t$ (here $\delta_{s,t}$ is Kronecker's delta function). Let $\mathcal{H}$ be the completion of $\mathcal{H}_0$ with respect to this inner product. We have

$$\mathcal{H} \cong \bigoplus_{s \in \mathcal{S}} H_s .$$

It will sometimes be convenient to identify the subspace $\delta_s \cdot H_s \subseteq \mathcal{H}$ with H_s, and for $s = 0$ this gives us an inclusion $H \subseteq \mathcal{H}$. We define a family $\hat{T} = \{\hat{T}_s\}_{s \in \mathcal{S}}$ of operators on $\mathcal{H}_0$ as follows. First, we define $\hat{T}_0$ to be the identity. Now assume that $s > 0$. If $t \in \mathcal{S}$ and $t \ngeq s$, then we define $\hat{T}_s(\delta_t \cdot \xi) = 0$ for all $\xi \in H_t$. If $t \geq s > 0$ we would like to define (as we did in [41])

$$\hat{T}_s \left(\delta_t \cdot (x_{t-s} \otimes x_s \otimes h)\right) = \delta_{t-s} \cdot \Big(x_{t-s} \otimes \widetilde{T}_s (x_s \otimes h) \Big) , \tag{4.2}$$

but since X is not a true product system, we cannot identify $X(t-s) \otimes X(s)$ with $X(t)$. For a fixed $t > 0$, we define for all $s \leq t$, $\xi \in X(t)$ and $h \in H$

$$\check{T}_s \left(\delta_t \cdot (\xi \otimes h)\right) = \delta_{t-s} \cdot \Big((I_{X(t-s)} \otimes \widetilde{T}_s)(U^*_{t-s,s} \xi \otimes h) \Big) .$$

$\check{T}_s$ can be extended to a well defined contraction from $X(t) \otimes H$ to $X(t-s) \otimes H$, for all $t \geq s$, and has an adjoint given by

$$\check{T}_s^* \delta_{t-s} \cdot \eta \otimes h = \delta_t \cdot \Big((U_{t-s,s} \otimes I_H)(\eta \otimes \widetilde{T}_s^* h) \Big) . \tag{4.3}$$

We are going to obtain $\hat{T}_s$ as the map $H_t \to H_{t-s}$ induced by $\check{T}_s$. Let $Y \in H_t$ satisfy $\widetilde{T}_t(Y) = 0$. We shall show that $\check{T}_s \delta_t \cdot Y = 0$ in $\delta_{t-s} \cdot H_{t-s}$. But

$$\check{T}_s \delta_t \cdot Y = \delta_{t-s} \cdot \Big((I_{X(t-s)} \otimes \widetilde{T}_s)(U^*_{t-s,s} \otimes I_H) Y \Big),$$

and

$$\begin{aligned} \widetilde{T}_{t-s} \Big((I_{X(t-s)} \otimes \widetilde{T}_s)(U^*_{t-s,s} \otimes I_H) Y \Big) (*) &= \widetilde{T}_t (U_{t-s,s} \otimes I_H)(U^*_{t-s,s} \otimes I_H) Y \\ (**) &= \widetilde{T}_t(Y) = 0, \end{aligned}$$

where the equation marked by (*) follows from the fact that T is a representation of subproduct systems, and the one marked by (**) follows from the fact that $U_{t-s,s}$ is a coisometry. Thus, for all $s, t \in \mathcal{S}$,

$$\check{T}_s \left(\delta_t \cdot \mathrm{Ker} \widetilde{T}_t \right) \subseteq \delta_{t-s} \cdot \mathrm{Ker} \widetilde{T}_{t-s},$$

thus $\check{T}_s$ induces a well defined contraction $\hat{T}_s$ on $\mathcal{H}$ given by

$$\hat{T}_s \left(\delta_t \cdot (\xi \otimes h) \right) = \delta_{t-s} \cdot \Big((I_{X(t-s)} \otimes \widetilde{T}_s)(U^*_{t-s,s} \xi \otimes h) \Big), \tag{4.4}$$

where $\xi \otimes h$ and $(I_{X(t-s)} \otimes \widetilde{T}_s)(U^*_{t-s,s} \xi \otimes h)$ stand for these elements' equivalence classes in $\big(X(t) \otimes H \big) / \mathrm{Ker} \widetilde{T}_t$ and $\big(X(t-s) \otimes H \big) / \mathrm{Ker} \widetilde{T}_{t-s}$, respectively. It follows that we have the following, more precise, variant of (4.2):

$$\hat{T}_s \left(\delta_t \cdot (U_{t-s,s}(x_{t-s} \otimes x_s) \otimes h) \right) = \delta_{t-s} \cdot \Big(x_{t-s} \otimes \widetilde{T}_s(x_s \otimes h) \Big).$$

In particular,

$$\hat{T}_s \left(\delta_s \cdot x_s \otimes h \right) = T_s(x_s) h,$$

for all $s \in \mathcal{S}, x_s \in X(s), h \in H$.

It will be very helpful to have a formula for $\hat{T}^*_s$ as well. Assume that $\sum_i \xi_i \otimes h_i \in \mathrm{Ker} \widetilde{T}_t$.

$$\check{T}^*_s \left(\delta_t \cdot \sum_i \xi_i \otimes h_i \right) = \delta_{s+t} \cdot \left((U_{t,s} \otimes I_H)(\sum_i \xi_i \otimes \widetilde{T}^*_s h_i) \right),$$

and applying $\widetilde{T}_{s+t}$ to the right hand side (without the δ) we get

$$\begin{aligned} \widetilde{T}_{s+t} \left((U_{t,s} \otimes I_H)(\sum_i \xi_i \otimes \widetilde{T}^*_s h_i) \right) &= \widetilde{T}_t (I_{X(t)} \otimes \widetilde{T}_s)(\sum_i \xi_i \otimes \widetilde{T}^*_s h_i) \\ &= \widetilde{T}_t (\sum_i \xi_i \otimes \widetilde{T}_s \widetilde{T}^*_s h_i) \\ &= \widetilde{T}_t (\sum_i \xi_i \otimes h_i) = 0, \end{aligned}$$

because T is a fully coisometric representation. So

$$\check{T}^*_s \left(\delta_t \cdot \mathrm{Ker} \widetilde{T}_t \right) \subseteq \delta_{s+t} \cdot \mathrm{Ker} \widetilde{T}_{s+t},$$

and this means that $\check{T}_s^*$ induces on $\mathcal{H}$ a well defined contraction which is equal to $\hat{T}_s^*$, and is given by the formula (4.3).
We now show that $\hat{T}$ is a semigroup. Let $s,t,u \in \mathcal{S}$. If either $s = 0$ or $t = 0$ then it is clear that the semigroup property $\hat{T}_s\hat{T}_t = \hat{T}_{s+t}$ holds. Assume that $s,t > 0$. If $u \ngeq s+t$, then both $\hat{T}_s\hat{T}_t$ and $\hat{T}_{s+t}$ annihilate $\delta_u \cdot \xi$, for all $\xi \in H_u$. Assuming $u \geq s+t$, we shall show that $\hat{T}_s\hat{T}_t$ and $\hat{T}_{s+t}$ agree on elements of the form

$$Z = \delta_u \cdot \big(U_{u-t,t}(U_{u-t-s,s} \otimes I)(x_{u-s-t} \otimes x_s \otimes x_t)\big) \otimes h,$$

and since the set of all such elements is total in H_u, this will establish the semigroup property.

$$\begin{aligned}
\hat{T}_s\hat{T}_t Z &= \hat{T}_s \left(\delta_{u-t}\left(U_{u-t-s,s}(x_{u-s-t} \otimes x_s) \otimes \widetilde{T}_t(x_t \otimes h)\right)\right) \\
&= \delta_{u-s-t}\left(x_{u-s-t} \otimes \widetilde{T}_s(x_s \otimes \widetilde{T}_t(x_t \otimes h))\right) \\
&= \delta_{u-s-t}\left(x_{u-s-t} \otimes \widetilde{T}_s(I \otimes \widetilde{T}_t)(x_s \otimes x_t \otimes h)\right) \\
&= \delta_{u-s-t}\left(x_{u-s-t} \otimes \widetilde{T}_{s+t}\left(U_{s,t}(x_s \otimes x_t) \otimes h\right)\right) \\
&= \hat{T}_{t+s}\delta_u \cdot \left(U_{u-t-s,t+s}\left(x_{u-s-t} \otimes U_{s,t}(x_s \otimes x_t)\right) \otimes h\right) \\
&= \hat{T}_{t+s} Z.
\end{aligned}$$

The final equality follows from the associativity condition (1.1).
To see that $\hat{T}$ is a semigroup of coisometries, we take $\xi \in X(t), h \in H$, and compute

$$\begin{aligned}
&\widetilde{T}_t\left(\hat{T}_s\hat{T}_s^*\delta_t \cdot (\xi \otimes h)\right) = \\
&= \widetilde{T}_t\left((I_{X(t)} \otimes \widetilde{T}_s)(U_{t,s}^* \otimes I_H)(U_{t,s} \otimes I_H)(I_{X(t)} \otimes \widetilde{T}_s^*)(\xi \otimes h)\right) \\
&= \widetilde{T}_{s+t}(U_{t,s} \otimes I_H)(I_{X(t)} \otimes \widetilde{T}_s^*)(\xi \otimes h) \\
&= \widetilde{T}_t(\xi \otimes \widetilde{T}_s\widetilde{T}_s^* h) = \widetilde{T}_t(\xi \otimes h),
\end{aligned}$$

so $\hat{T}_s\hat{T}_s^*$ is the identity on H_t for all $t \in \mathcal{S}$, thus $\hat{T}_s\hat{T}_s^* = I_{\mathcal{H}}$. Equation (4.1) follows by a similar computation, which is omitted. □

We can now obtain a $*$-automorphic dilation for any e_0-semigroup over any subsemigroup of $\mathbb{R}_+^k$. The following result should be compared with similar-looking results of Arveson-Kishimoto [8], Laca [25], Skeide [45], and Arveson-Courtney [7] (none of these cited results is strictly stronger or weaker than the result we obtain for the case of e_0-semigroups).

Theorem 4.2. *Let Θ be a e_0-semigroup acting on a von Neumann algebra $\mathcal{M}$. Then Θ can be dilated to a semigroup of $*$-automorphisms in the following sense: there is a Hilbert space $\mathcal{K}$, an orthogonal projection p of $\mathcal{K}$ onto a subspace $\mathcal{H}$ of $\mathcal{K}$, a normal, faithful representation $\varphi : \mathcal{M} \to B(\mathcal{K})$ such that*

$\varphi(1) = p$, and a semigroup $\alpha = \{\alpha_s\}_{s\in\mathcal{S}}$ of $*$-automorphisms on $B(\mathcal{K})$ such that for all $a \in \mathcal{M}$ and all $s \in \mathcal{S}$

$$\alpha_s(\varphi(a))\big|_{\mathcal{H}} = \varphi(\Theta_s(a)), \tag{4.5}$$

so, in particular,

$$p\alpha_s(\varphi(a))p = \varphi(\Theta_s(a)). \tag{4.6}$$

The projection p is increasing for α, in the sense that for all $s \in \mathcal{S}$,

$$\alpha_s(p) \geq p. \tag{4.7}$$

Remark 4.3. Another way of phrasing the above theorem is by using the terminology of "weak Markov flows", as used in [15]. Denoting φ by j_0, and defining $j_s := \alpha_s \circ j_0$, we have that $(B(\mathcal{K}), j)$ is a weak Markov flow for Θ on $\mathcal{K}$, which just means that for all $t \leq s \in \mathcal{S}$ and all $a \in \mathcal{M}$,

$$j_t(1)j_s(a)j_t(1) = j_t(\Theta_{s-t}(a)). \tag{4.8}$$

Equation (4.8) for $t = 0$ is just (4.6), and the case $t \geq 0$ follows from the case $t = 0$.

Remark 4.4. The assumption that Θ is a *unital* semigroup is essential, since (4.6) and (4.7) imply that $\Theta(1) = 1$.

Remark 4.5. It is impossible, in the generality we are working in, to hope for a semigroup of automorphisms that extends Θ in the sense that

$$\alpha_s(\varphi(a)) = \varphi(\Theta_s(a)), \tag{4.9}$$

because that would imply that Θ is injective.

Proof. Let (E, T) be the identity representation of Θ. Since Θ preserves the unit, T is a fully coisometric representation. Let $\hat{T}$ and $\mathcal{H}$ be the semigroup and Hilbert space representing T as described in Proposition 4.1. $\{\hat{T}_s^*\}_{s\in\mathcal{S}}$ is a commutative semigroup of isometries. By a theorem of Douglas [20], $\{\hat{T}_s^*\}_{s\in\mathcal{S}}$ can be extended to a semigroup $\{\hat{V}_s^*\}_{s\in\mathcal{S}}$ of unitaries acting on a space $\mathcal{K} \supseteq \mathcal{H}$. We obtain a semigroup of unitaries $V = \{\hat{V}_s\}_{s\in\mathcal{S}}$ that is a dilation of $\hat{T}$, that is

$$P_{\mathcal{H}}\hat{V}_s\big|_{\mathcal{H}} = \hat{T}_s \ , \ s \in \mathcal{S}.$$

For any $b \in B(\mathcal{K})$, and any $s \in \mathcal{S}$, we define

$$\alpha_s(b) = \hat{V}_s b \hat{V}_s^*.$$

Clearly, $\alpha = \{\alpha_s\}_{s\in\mathcal{S}}$ is a semigroup of $*$-automorphisms.
Put $p = P_{\mathcal{H}}$, the orthogonal projection of $\mathcal{K}$ onto $\mathcal{H}$. Define $\varphi : \mathcal{M} \to B(\mathcal{K})$ by $\varphi(a) = p(I \otimes a)p$, where $I \otimes a : \mathcal{H} \to \mathcal{H}$ is given by

$$(I \otimes a)\delta_t \cdot x \otimes h = \delta_t \cdot x \otimes ah \quad , \quad x \otimes h \in E(t) \otimes H.$$

φ is well defined because T is an isometric representation (so $\mathrm{Ker}\widetilde{T}_t$ is always zero). We have that φ is a faithful, normal $*$-representation (the fact that T_0 is the identity representation ensures that φ is faithful). It is clear that $\varphi(1) = p$.

To see (4.7), we note that since $\hat{V}_s^*$ is an extension of $\hat{T}_s^*$, we have $\hat{T}_s^* = \hat{V}_s^* p = p\hat{V}_s^* p$, thus

$$\begin{aligned} p\alpha_s(p)p &= p\hat{V}_s p\hat{V}_s^* p \\ &= p\hat{V}_s\hat{V}_s^* p \\ &= p, \end{aligned}$$

that is, $p\alpha_s(p)p = p$, which implies that $\alpha_s(p) \geq p$.

We now prove (4.6). Let $\delta_t \cdot x \otimes h$ be a typical element of $\mathcal{H}$. We compute

$$\begin{aligned} p\alpha_s(\varphi(a))p\delta_t \cdot x \otimes h &= p\hat{V}_s p(I \otimes a)p\hat{V}_s^* p\delta_t \cdot x \otimes h \\ &= \hat{T}_s(I \otimes a)\hat{T}_s^*\delta_t \cdot x \otimes h \\ &= \hat{T}_s(I \otimes a)\delta_{s+t} \cdot (U_{t,s} \otimes I_H)\Big(x \otimes \widetilde{T}_s^* h\Big) \\ &= \hat{T}_s\delta_{s+t} \cdot (U_{t,s} \otimes I_H)\Big(x \otimes (I \otimes a)\widetilde{T}_s^* h\Big) \\ &= \delta_t \cdot x \otimes \Big(\widetilde{T}_s(I \otimes a)\widetilde{T}_s^* h\Big) \\ &= \delta_t \cdot x \otimes (\Theta_s(a)h) \\ &= \varphi(\Theta_s(a))\delta_t \cdot x \otimes h. \end{aligned}$$

Since both $p\alpha_s(\varphi(a))p$ and $\varphi(\Theta_s(a))$ annihilate $\mathcal{K} \ominus \mathcal{H}$, we have (4.6).

To prove (4.5), it just remains to show that

$$p\alpha_s(\varphi(a))\big|_{\mathcal{H}} = \alpha_s(\varphi(a))\big|_{\mathcal{H}},$$

that is, that $\alpha_s(\varphi(a))\mathcal{H} \subseteq \mathcal{H}$. Now, $\hat{V}_s^*$ is an extension of $\hat{T}_s^*$. Moreover (4.1) shows that if $\xi \in H_u$ with $u \geq s$, then $\|\hat{T}_s(\xi)\| = \|\xi\|$. Thus

$$\|\xi\|^2 = \|\hat{V}_s\xi\|^2 = \|P_{\mathcal{H}}\hat{V}_s\xi\|^2 + \|(I_{\mathcal{K}} - P_{\mathcal{H}})\hat{V}_s\xi\|^2 = \|\hat{T}_s\xi\|^2 + \|(I_{\mathcal{K}} - P_{\mathcal{H}})\hat{V}_s\xi\|^2.$$

So $\hat{V}_s\xi = \hat{T}_s\xi$ for $\xi \in H_u$ with $u \geq s$. Now, for a typical element $\delta_t \cdot x \otimes h$ in H_t, $t \in \mathcal{S}$, we have

$$\begin{aligned} \alpha_s(\varphi(a))\delta_t \cdot x \otimes h &= \hat{V}_s(I \otimes a)\hat{V}_s^*\delta_t \cdot x \otimes h \\ &= \hat{V}_s(I \otimes a)\hat{T}_s^*\delta_t \cdot x \otimes h \\ &= \hat{V}_s\delta_{s+t} \cdot (U_{s,t} \otimes I_H)\Big(x \otimes (I \otimes a)\widetilde{T}_s^* h\Big) \\ &= \hat{T}_s\delta_{s+t} \cdot (U_{s,t} \otimes I_H)\Big(x \otimes (I \otimes a)\widetilde{T}_s^* h\Big) \in \mathcal{H}, \end{aligned}$$

because $\delta_{s+t} \cdot x \otimes (I \otimes a)\widetilde{T}_s^* h \in H_{s+t}$, and $s + t \geq s$. $\square$

5. Dilations and pieces of subproduct system representations

5.1. Dilations and pieces of subproduct system representations.

Definition 5.1. *Let X and Y be subproduct systems of $\mathcal{M}$ correspondences ($\mathcal{M}$ a W^*-algebra) over the same semigroup $\mathcal{S}$. Denote by $U_{s,t}^X$ and $U_{s,t}^Y$ the coisometric maps that make X and Y, respectively, into subproduct systems.*

X is said to be a subproduct subsystem of Y *(or simply a* subsystem of Y *for short) if for all $s \in \mathcal{S}$ the space $X(s)$ is a closed subspace of $Y(s)$, and if the orthogonal projections $p_s : Y(s) \to X(s)$ are bimodule maps that satisfy*

$$p_{s+t} \circ U^Y_{s,t} = U^X_{s,t} \circ (p_s \otimes p_t) \ , \ s,t \in \mathcal{S}. \tag{5.1}$$

One checks that if X is a subproduct subsystem of Y then

$$p_{s+t+u} \circ U^Y_{s,t+u}(I \otimes (p_{t+u} \circ U^Y_{t,u})) = p_{s+t+u} \circ U^Y_{s+t,u}((p_{s+t} \circ U^Y_{s,t}) \otimes I), \tag{5.2}$$

for all $s,t,u \in \mathcal{S}$. Conversely, given a subproduct system Y and a family of orthogonal projections $\{p_s\}_{s\in\mathcal{S}}$ that are bimodule maps satisfying (5.2), then by defining $X(s) = p_s Y(s)$ and $U^X_{s,t} = p_{s+t} \circ U^Y_{s,t}$ one obtains a subproduct subsystem X of Y (with (5.1) satisfied).
The following proposition is a consequence of the definitions.

Proposition 5.2. *There exists a morphism $X \to Y$ if and only if Y is isomorphic to a subproduct subsystem of X.*

Remark 5.3. In the notation of Theorem 2.6, we may now say that given a subproduct system X and a representation R of X, then the Arveson-Stinespring subproduct system E of $\Theta = \Sigma(X,R)$ is isomorphic to a subproduct subsystem of X.

The following definitions are inspired by the work of Bhat, Bhattacharyya and Dey [12].

Definition 5.4. *Let X and Y be subproduct systems of W^*-correspondences (over the same W^*-algebra $\mathcal{M}$) over $\mathcal{S}$, and let T be a representation of Y on a Hilbert space K. Let H be some fixed Hilbert space, and let $S = \{S_s\}_{s\in\mathcal{S}}$ be a family of maps $S_s : X(s) \to B(H)$. (Y,T,K) is called a* dilation *of (X,S,H) if*

(1) *X is a subsystem of Y,*
(2) *H is a subspace of K, and*
(3) *for all $s \in \mathcal{S}$, $\widetilde{T}_s^* H \subseteq X(s) \otimes H$ and $\widetilde{T}_s^*\big|_H = \widetilde{S}_s^*$.*

In this case we say that S is an X-piece of T, *or simply a* piece of T. *T is said to be an* isometric *dilation of S if T is an isometric representation.*

The third item can be replaced by the three conditions

1' $T_0(\cdot)P_H = P_H T_0(\cdot) P_H = S_0(\cdot)$,
2' $P_H \widetilde{T}_s\big|_{X(s)\otimes H} = \widetilde{S}_s$ for all $s \in \mathcal{S}$, and
3' $P_H \widetilde{T}_s\big|_{Y(s)\otimes K \ominus X(s)\otimes H} = 0$.

So our definition of dilation is identical to Muhly and Solel's definition of dilation of representations when $X = Y$ is a product system [29, Theorem and Definition 3.7].

Proposition 5.5. *Let T be a representation of Y, let X be a subproduct subsystem of Y, and let S an X-piece of T. Then S is a representation of X.*

Proof. S is a completely contractive linear map as the compression of a completely contractive linear map. Item 1' above together with the coinvariance of T imply that S is coinvariant: if $a, b \in \mathcal{M}$ and $x \in X(s)$, then

$$\begin{aligned} S_s(axb) = P_H T_s(axb) P_H &= P_H T_0(a) T_s(x) T_0(b) P_H \\ &= P_H T_0(a) P_H T_s(x) P_H T_0(b) P_H \\ &= S_0(a) S_s(x) S_0(b). \end{aligned}$$

Finally, (using Item 3' above),

$$\begin{aligned} S_{s+t}(U^X_{s,t}(x \otimes y))h &= S_{s+t}(p_{s+t} U^Y_{s,t}(x \otimes y))h \\ &= \widetilde{S}_{s+t}(p_{s+t} U^Y_{s,t}(x \otimes y) \otimes h) \\ &= P_H \widetilde{T}_{s+t}(U^Y_{s,t}(x \otimes y) \otimes h) \\ &= P_H T_s(x) T_t(y) h \\ &= P_H T_s(x) P_H T_t(y) h \\ &= S_s(x) S_t(y) h. \end{aligned}$$

□

Example 5.6. Let E be a Hilbert space of dimension d, and let X be the symmetric subproduct system constructed in Example 1.3. Fix an orthonormal basis $\{e_1, \ldots, e_n\}$ of E. There is a one-to-one correspondence between c.c. representations S of X (on some H) and commuting row contractions $(S_1, \ldots, S_d)$ (of operators on H), given by

$$S \leftrightarrow \underline{S} = (S(e_1), \ldots, S(e_d)).$$

If Y is the full product system over E, then any dilation (Y, T, K) gives rise to a tuple $\underline{T} = (T(e_1), \ldots, T(e_d))$ that is a dilation of $\underline{S}$ in the sense of [12], and *vice versa*. Moreover, $\underline{S}$ is then a commuting piece of $\underline{T}$ in the sense of [12].

Consider a subproduct system Y and a representation T of Y on K. Let X be some subproduct subsystem of Y. Define the following set of subspaces of K:

$$\mathcal{P}(X, T) = \{H \subseteq K : \widetilde{T}_s^* H \subseteq X(s) \otimes H \quad \text{for all } s \in \mathcal{S}\}. \tag{5.3}$$

As in [12], we observe that $\mathcal{P}(X, T)$ is closed under closed linear spans (and intersections), thus we may define

$$K^X(T) = \bigvee_{H \in \mathcal{P}(X,T)} H.$$

$K^X(T)$ is the maximal element of $\mathcal{P}(X, T)$.

Definition 5.7. *The representation T^X of X on $K^X(T)$ given by*

$$T^X(x)h = P_{K^X(T)} T(x) h,$$

for $x \in X(s)$ and $h \in K^X(T)$, is called the maximal X-piece of T.

By Proposition 5.5, T^X is indeed a representation of X.

5.2. Consequences in dilation theory of *cp*-semigroups.

Proposition 5.8. *Let X and Y be subproduct systems of W^*-correspondences (over the same W^*-algebra $\mathcal{M}$) over $\mathcal{S}$, and let S and T be representations of X on H and of Y on K, respectively. Assume that (Y, T, K) is a dilation of (X, S, H). Then the cp-semigroup Θ acting on $V_0(\mathcal{M})'$, given by*

$$\Theta_s(a) = \widetilde{T}_s(I_{Y(s)} \otimes a)\widetilde{T}_s^* \ , \ a \in V_0(\mathcal{M})',$$

is a dilation of the cp-semigroup Φ acting on $T_0(\mathcal{M})'$ given by

$$\Phi_s(a) = \widetilde{S}_s(I_{X(s)} \otimes a)\widetilde{S}_s^* \ , \ a \in T_0(\mathcal{M})',$$

in the sense that for all $b \in V_0(\mathcal{M})'$ and all $s \in \mathcal{S}$,

$$\Phi_s(P_H b P_H) = P_H \Theta_s(b) P_H .$$

Proof. This follows from the definitions. □

Although the above proposition follows immediately from the definitions, we hope that it will prove to be important in the theory of dilations of *cp*-semigroups, because it points to a conceptually new way of constructing dilations of *cp*-semigroups, as the following proposition and corollary illustrate.

Proposition 5.9. *Let $X = \{X(s)\}_{s\in\mathcal{S}}$ be a subproduct system, and let S be a fully coisometric representation of X on H such that S_0 is unital. If there exists a (full)* product system *$Y = \{Y(s)\}_{s\in\mathcal{S}}$ such that X is a subproduct subsystem of Y, then S has an isometric and fully coisometric dilation.*

Proof. Define a representation T of Y on H by

$$T_s = S_s \circ p_s, \tag{5.4}$$

where, as above, p_s is the orthogonal projection $Y(s) \to X(s)$. A straightforward verification shows that T is indeed a fully coisometric representation of Y on H. By [43, Theorem 5.2], (Y, T, H) has a minimal isometric and fully coisometric dilation (Y, V, K). (Y, V, K) is also clearly a dilation of (X, S, H). □

Corollary 5.10. *Let $\Theta = \{\Theta_s\}_{s\in\mathcal{S}}$ be a cp_0-semigroup and let $(E, T) = \Xi(\Theta)$ be the Arveson-Stinespring representation of Θ. If there is a (full)* product system *Y such that E is a subproduct subsystem of Y, then Θ has an e_0-dilation.*

Proof. Combine Propositions 2.1, 5.8 and 5.9. □

Thus, the problem of constructing e_0-dilations to cp_0-semigroups is reduced to the problem of embedding a subproduct system into a full product system. In the next subsection we give an example of a subproduct system that cannot be embedded into full product system. When this can be done in general is a challenging open question.

Corollary 5.11. *Let $\Theta = \{\Theta_s\}_{s\in\mathbb{N}^k}$ be a cp-semigroup generated by k commuting CP maps $\theta_1,\ldots,\theta_k$, and let $(E,T)=\Xi(\Theta)$ be the Arveson representation of Θ. Assume, in addition, that*

$$\sum_{i=1}^{k}\|\theta_i\|\leq 1.$$

If there is a (full) product system Y *such that E is a subproduct subsystem of Y, then Θ has an e-dilation.*

Proof. As in (5.4), we may extend T to a product system representation of Y on H, which we also denote by T. Denote by $\mathbf{e_i}$ the element of $\mathbb{N}^k$ with 1 in the ith element and zeros elsewhere. Then

$$\sum_{i=1}^{k}\|\widetilde{T}_{\mathbf{e_i}}\widetilde{T}^*_{\mathbf{e_i}}\|=\sum_{i=1}^{k}\|\theta_i\|\leq 1.$$

By the methods of [41], one may show that S has a minimal (regular) isometric dilation. This isometric dilation provides the required e-dilation of Θ. □

Theorem 5.12. *Let $\mathcal{M}\subseteq B(H)$ be a von Neumann algebra, let X be a subproduct system of $\mathcal{M}'$-correspondences, and let R be an injective representation of X on a Hilbert space H. Let $\Theta=\Sigma(X,R)$ be the cp-semigroup acting on $R_0(\mathcal{M}')'$ given by (2.1). Assume that $(\alpha,K,\mathcal{R})$ is an e-dilation of Θ, and let $(Y,V)=\Xi(\alpha)$ be the Arveson-Stinespring subproduct system of α together with the identity representation. Assume, in addition, that the map $\mathcal{R}'\ni b\mapsto P_HbP_H$ is a $*$-isomorphism of $\mathcal{R}'$ onto $R_0(\mathcal{M}')$. Then (Y,V,K) is a dilation of (X,R,H).*

Proof. For every $s\in\mathcal{S}$, define $W_s : Y(s)\to B(H)$ by $W_s(y)=P_HV_s(y)P_H$. We claim that $W=\{W_s\}_{s\in\mathcal{S}}$ is a representation of Y on H. First, note that $P_H\alpha_s(I-P_H)P_H=\Theta_s(P_H(I-P_H)P_H)=0$, thus $P_H\widetilde{V}_s(I\otimes(I-P_H))\widetilde{V}_s^*P_H=0$, and consequently $P_H\widetilde{V}_s(I\otimes P_H)=P_H\widetilde{V}_s$. It follows that $W_s(y)=P_HV_s(y)P_H=P_HV_s(y)$. From this it follows that

$$\begin{aligned}W_s(y_1)W_t(y_2)&=P_HV_s(y_1)P_HV_t(y_2)=P_HV_s(y_1)V_t(y_2)\\&=P_HV_{s+t}(U^Y_{s,t}(y_1\otimes y_2))=W_{s+t}(U^Y_{s,t}(y_1\otimes y_2)).\end{aligned}$$

By Theorem 2.6, we may assume that $(X,R)=(E,T)=\Xi(\Theta)$ is the Arveson-Stinespring representation of Θ. Because α is a dilation of Θ, we have

$$\widetilde{W}_s(I\otimes a)\widetilde{W}_s^*=P_H\widetilde{V}_s(I\otimes a)\widetilde{V}_s^*P_H=\Theta_s(a),$$

That is, $\Theta=\Sigma(Y,W)$. Thus, by Theorem 2.6 and Remark 5.3, we may assume that E is a subproduct subsystem of Y, and that $T_s\circ p_s=W_s$, p_s being the projection of $Y(s)$ onto $E(s)$. In other words, for all $y\in Y$,

$$\widetilde{T}_s(p_s\otimes I_H)=P_H\widetilde{W}_s.$$

Therefore, $\widetilde{W}_s^* H \subseteq E(s) \otimes H$, and $\widetilde{W}_s^*\big|_H = \widetilde{T}_s^*$. That is, (Y, W, H) is a dilation of (E, T, H). But (Y, V, K) is a dilation of (Y, W, H), so it is also a dilation of (E, T, H). □

The assumption that $\mathcal{R}' \ni b \mapsto P_H b P_H \in \mathcal{M}'$ is a $*$-isomorphism is satisfied when $\mathcal{M} = B(H)$ and $\mathcal{R} = B(K)$. More generally, it is satisfied whenever the central projection of P_H in $\mathcal{R}$ is I_K (see Propositions 5.5.5 and 5.5.6 in [22]). Let $(\alpha, K, \mathcal{R})$ be an e-dilation of a semigroup Θ on $\mathcal{M} \subseteq B(H)$. $(\alpha, K, \mathcal{R})$ is called a *minimal dilation* if the central support of P_H in $\mathcal{R}$ is I_K and if

$$\mathcal{R} = W^* \left(\bigcup_{s \in \mathcal{S}} \alpha_s(\mathcal{M}) \right).$$

COROLLARY 5.13. *Let Θ be cp-semigroup on $\mathcal{M} \subseteq B(H)$, and let $(\alpha, K, \mathcal{R})$ be a minimal dilation of Θ. Then $\Xi(\alpha)$ is an isometric dilation of $\Xi(\Theta)$.*

5.3. *cp*-SEMIGROUPS WITH NO e-DILATIONS. OBSTRUCTIONS OF A NEW NATURE. By Parrot's famous example [34], there exist 3 commuting contractions that do not have a commuting isometric dilation. In 1998 Bhat asked whether 3 commuting CP maps necessarily have a commuting $*$-endomorphic dilation [10]. Note that it is not obvious that the non-existence of an isometric dilation for three commuting contractions would imply the non-existence of a $*$-endomorphic dilation for 3 commuting CP maps. However, it turns out that this is the case.

THEOREM 5.14. *There exists a cp-semigroup $\Theta = \{\Theta_n\}_{n \in \mathbb{N}^3}$ acting on a $B(H)$ for which there is no e-dilation $(\alpha, K, B(K))$. In fact, Θ has no* minimal *e-dilation $(\alpha, K, \mathcal{R})$ on any von Neumann algebra $\mathcal{R}$.*

Proof. Let $T_1, T_2, T_3 \in B(H)$ be three commuting contractions that have no isometric dilation and such that $T_1^{n_1} T_2^{n_2} T_3^{n_3} \neq 0$ for all $n = (n_1, n_2, n_3) \in \mathbb{N}^3$ (one may take commuting contractions R_1, R_2, R_3 with no isometric dilation as in Parrot's example [34], and define $T_i = R_i \oplus 1$). For all $n = (n_1, n_2, n_3) \in \mathbb{N}^3$, define

$$\Theta_n(a) = T_1^{n_1} T_2^{n_2} T_3^{n_3} a (T_3^{n_3})^* (T_2^{n_2})^* (T_1^{n_1})^* \ , \ a \in B(H).$$

Note that $\Theta = \Sigma(X, R)$, where $X = \{X(n)\}_{n \in \mathbb{N}^3}$ is the subproduct system given by $X(n) = \mathbb{C}$ for all $n \in \mathbb{N}^3$, and R is the (injective) representation that sends $1 \in X(n)$ to $T_1^{n_1} T_2^{n_2} T_3^{n_3}$ (the product in X is simply multiplication of scalars).

Assume, for the sake of obtaining a contradiction, that $\Theta = \{\Theta_n\}_{n \in \mathbb{N}^3}$ has an e-dilation $(\alpha, K, B(K))$. Let $(Y, V) = \Xi(\alpha)$ be the Arveson-Stinespring subproduct system of α together with the identity representation. By Theorem 5.12, (Y, V, K) is a dilation of (X, R, H). It follows that V_1, V_2, V_3 are a commuting isometric dilation of T_1, T_2, T_3 where $V_1 := V(1)$ with $1 \in X(1, 0, 0)$, $V_2 := V(1)$ with $1 \in X(0, 1, 0)$, and $V_3 := V(1)$ with $1 \in X(0, 0, 1)$. This is a contradiction.

Finally, a standard argument shows that if $(\alpha, K, \mathcal{R})$ is a minimal dilation of Θ, then $\mathcal{R} = B(K)$. □

Until this point, all the results that we have seen in the dilation theory of *cp*-semigroups have been anticipated by the classical theory of isometric dilations. We shall now encounter a phenomena that has no counterpart in the classical theory.

By [49, Proposition 9.2], if $T_1, \ldots, T_k$ is a commuting k-tuple of contractions such that

$$\sum_{i=1}^{k} \|T_i\|^2 \leq 1, \tag{5.5}$$

then $T_1, \ldots, T_k$ has a commuting regular unitary dilation (and, in particular, an isometric dilation). One is tempted to conjecture that if $\theta_1, \ldots, \theta_k$ is a commuting k-tuple of CP maps such that

$$\sum_{i=1}^{k} \|\theta_i\| \leq 1, \tag{5.6}$$

then the tuple $\theta_1, \ldots, \theta_k$ has an e-dilation. Indeed, if $\theta_i(a) = T_i a T_i^*$, where $T_1, \ldots, T_k$ is a commuting k-tuple satisfying (5.5), then it is easy to construct an e-dilation of $\theta_1, \ldots, \theta_k$ from the isometric dilation of $T_1, \ldots, T_k$. However, it turns out that (5.6) is far from being sufficient for an e-dilation to exist. We need some preparations before exhibiting an example.

Proposition 5.15. *There exists a subproduct system that is not a subsystem of any product system.*

Proof. We construct a counter example over $\mathbb{N}^3$. Let e_1, e_2, e_3 be the standard basis of $\mathbb{N}^3$. We let $X(e_1) = X(e_2) = X(e_3) = \mathbb{C}^2$. Let $X(e_i + e_j) = \mathbb{C}^2 \otimes \mathbb{C}^2$ for all $i, j = 1, 2, 3$. Put $X(n) = \{0\}$ for all $n \in \mathbb{N}^k$ such that $|n| > 2$. To complete the construction of X we need to define the product maps $U^X_{m,n}$. Let $U^X_{e_i,e_j}$ be the identity on $\mathbb{C}^2 \otimes \mathbb{C}^2$ for all i, j except for $i = 3, j = 2$, and let $U^X_{e_3,e_2}$ be the flip. Define the rest of the products to be zero maps (except the maps $U^X_{0,n}, U^X_{m,0}$ which are identities). This product is evidently coisometric, and it is also associative, because the product of any three nontrivial elements vanishes.

Let Y be a product system "dilating" X. Then for all $k, m, n \in \mathbb{N}^k$ we have

$$U^Y_{k+m,n}(U^Y_{k,m} \otimes I) = U^Y_{k,m+n}(I \otimes U^Y_{m,n}),$$

or

$$U^Y_{k+m,n} = U^Y_{k,m+n}(I \otimes U^Y_{m,n})(U^Y_{k,m} \otimes I)^*,$$

and

$$U^Y_{k,m+n} = U^Y_{k+m,n}(U^Y_{k,m} \otimes I)(I \otimes U^Y_{m,n})^*.$$

Iterating these identities, we have, on the one hand,

$$\begin{aligned}
U_{e_3,e_1+e_2} &= U^Y_{e_3+e_2,e_1}(U^Y_{e_3,e_2}\otimes I)(I\otimes U^Y_{e_2,e_1})^*\\
&= U^Y_{e_2,e_3+e_1}(I\otimes U^Y_{e_3,e_1})(U^Y_{e_2,e_3}\otimes I)^*(U^Y_{e_3,e_2}\otimes I)(I\otimes U^Y_{e_2,e_1})^*\\
&= U^Y_{e_1+e_2,e_3}(U^Y_{e_2,e_1}\otimes I)(I\otimes U^Y_{e_1,e_3})^*\\
&\qquad (I\otimes U^Y_{e_3,e_1})(U^Y_{e_2,e_3}\otimes I)^*(U^Y_{e_3,e_2}\otimes I)(I\otimes U^Y_{e_2,e_1})^*,
\end{aligned}$$

and on the other hand

$$\begin{aligned}
U_{e_3,e_1+e_2} &= U^Y_{e_3+e_1,e_2}(U^Y_{e_3,e_1}\otimes I)(I\otimes U^Y_{e_1,e_2})^*\\
&= U^Y_{e_1,e_3+e_2}(I\otimes U^Y_{e_3,e_2})(U^Y_{e_1,e_3}\otimes I)^*(U^Y_{e_3,e_1}\otimes I)(I\otimes U^Y_{e_1,e_2})^*\\
&= U^Y_{e_1+e_2,e_3}(U^Y_{e_1,e_2}\otimes I)(I\otimes U^Y_{e_2,e_3})^*\\
&\qquad (I\otimes U^Y_{e_3,e_2})(U^Y_{e_1,e_3}\otimes I)^*(U^Y_{e_3,e_1}\otimes I)(I\otimes U^Y_{e_1,e_2})^*.
\end{aligned}$$

Canceling $U^Y_{e_1+e_2,e_3}$, we must have

$$\begin{aligned}
&(U^Y_{e_1,e_2}\otimes I)(I\otimes U^Y_{e_2,e_3})^*(I\otimes U^Y_{e_3,e_2})(U^Y_{e_1,e_3}\otimes I)^*(U^Y_{e_3,e_1}\otimes I)(I\otimes U^Y_{e_1,e_2})^*\\
&= (U^Y_{e_2,e_1}\otimes I)(I\otimes U^Y_{e_1,e_3})^*(I\otimes U^Y_{e_3,e_1})(U^Y_{e_2,e_3}\otimes I)^*(U^Y_{e_3,e_2}\otimes I)(I\otimes U^Y_{e_2,e_1})^*.
\end{aligned}$$

Now, $U^X_{e_i,e_j}$ were unitary to begin with, so the above identity implies

$$\begin{aligned}
&(U^X_{e_1,e_2}\otimes I)(I\otimes U^X_{e_2,e_3})^*(I\otimes U^X_{e_3,e_2})(U^X_{e_1,e_3}\otimes I)^*(U^X_{e_3,e_1}\otimes I)(I\otimes U^X_{e_1,e_2})^*\\
&= (U^X_{e_2,e_1}\otimes I)(I\otimes U^X_{e_1,e_3})^*(I\otimes U^X_{e_3,e_1})(U^X_{e_2,e_3}\otimes I)^*(U^X_{e_3,e_2}\otimes I)(I\otimes U^X_{e_2,e_1})^*.
\end{aligned}$$

Recalling the definition of the product in X (the product is usually the identity), this reduces to

$$I\otimes U^X_{e_3,e_2} = U^X_{e_3,e_2}\otimes I.$$

This is absurd. Thus, X cannot be dilated to a product system. □

We can now strengthen Theorem 5.14:

Theorem 5.16. *There exists a cp-semigroup $\Theta = \{\Theta_n\}_{n\in\mathbb{N}^3}$ acting on a $B(H)$, such that for all $\lambda > 0$, $\lambda\Theta$ has no e-dilation $(\alpha, K, B(K))$, and no* minimal *e-dilation $(\alpha, K, \mathcal{R})$ on any von Neumann algebra $\mathcal{R}$.*

Proof. Let X be as in Proposition 5.15. Let Θ be the *cp*-semigroup generated by the X-shift, as in Section 2.3 of the paper. Of course, Θ, as a semigroup over $\mathbb{N}^3$, can be generated by three commuting CP maps $\theta_1,\theta_2,\theta_3$. X cannot be embedded into a full product system, so by Theorem 5.12, Θ has no minimal e-dilation, nor does it have an e-dilation acting on a $B(K)$. Note that if Θ is scaled *its product system is left unchanged* (this follows from Theorem 2.6: if you take X and scale the representation S^X you get a scaled version of Θ). So no matter how small you take $\lambda > 0$, $\lambda\theta_1, \lambda\theta_2, \lambda\theta_3$ cannot be dilated to three commuting $*$-endomorphisms on $B(K)$, nor to a minimal three-tuple on any von Neumann algebra. □

Note that the obstruction here seems to be of a completely different nature from the one in the example given in Theorem 5.14. The subproduct system arising there is already a product system, and, indeed, the *cp*-semigroup arising there can be dilated once it is multiplied by a small enough scalar.

Part 2. Subproduct systems over $\mathbb{N}$

6. Subproduct systems of Hilbert spaces over $\mathbb{N}$

We now specialize to subproduct systems of Hilbert W*-correspondences over the semigroup $\mathbb{N}$, so from now on any subproduct system is to be understood as such (soon we will specialize even further to subproduct systems of Hilbert spaces).

6.1. Standard and maximal subproduct systems.

If X is a subproduct system over $\mathbb{N}$, then $X(0) = \mathcal{M}$ (some von Neumann algebra), $X(1)$ equals some W*-correspondence E, and $X(n)$ can be regarded as a subspace of $E^{\otimes n}$. The following lemma allows us to consider $X(m+n)$ as a subspace of $X(m) \otimes X(n)$.

Lemma 6.1. *Let $X = \{X(n)\}_{n\in\mathbb{N}}$ be a subproduct system. X is isomorphic to a subproduct system $Y = \{Y(n)\}_{n\in\mathbb{N}}$ with coisometries $\{U^Y_{m,n}\}_{m,n\in\mathbb{N}}$ that satisfies*

$$Y(1) = X(1)$$

and

$$Y(m+n) \subseteq Y(m) \otimes Y(n). \tag{6.1}$$

Moreover, if p_{m+n} is the orthogonal projection of $Y(1)^{\otimes(m+n)}$ onto $Y(m+n)$, then

$$U^Y_{m,n} = p_{m+n}\Big|_{Y(m)\otimes Y(n)} \tag{6.2}$$

and the projections $\{p_n\}_{n\in\mathbb{N}}$ satisfy

$$p_{k+m+n} = p_{k+m+n}(I_{E^{\otimes k}} \otimes p_{m+n}) = p_{k+m+n}(p_{k+m} \otimes I_{E^{\otimes n}}). \tag{6.3}$$

Proof. Denote by $U^X_{m,n}$ the subproduct system maps $X(s)\otimes X(t) \to X(s+t)$. Denote $E = X(1)$. We first note that for every n there is a well defined coisometry $U_n : E^{\otimes n} \to X(n)$ given by composing in any way a sequence of maps $U^X_{k,m}$ (for example, one can take $U_3 = U^X_{2,1}(U^X_{1,1} \otimes I_E)$ and so on). We define $Y(n) = \mathrm{Ker}(U_n)^\perp$, and we let p_n be the orthogonal projection from $E^{\otimes n}$ onto $Y(n)$. $p_n = U_n^* U_n$, so, in particular, p_n is a bimodule map. For all $m, n \in \mathbb{N}$ we have that

$$E^{\otimes m} \otimes \mathrm{Ker}(U_n) \subseteq \mathrm{Ker}(U_{m+n}).$$

Thus $E^{\otimes m} \otimes \mathrm{Ker}(U_n)^\perp \supseteq \mathrm{Ker}(U_{m+n})^\perp$, so $p_{m+n} \le I_{E^{\otimes m}} \otimes p_n$. This means that (6.3) holds. In addition, defining $U^Y_{m,n}$ to be p_{m+n} restricted to $Y(m)\otimes Y(n) \subseteq E^{\otimes(m+n)}$ gives Y the associative multiplication of a subproduct system.

It remains to show that X is isomorphic to Y. For all n, $X(n)$ is spanned by elements of the form $U_n(x_1 \otimes \cdots \otimes x_n)$, with $x_1, \ldots, x_n \in E$. We define a map $V_n : X(n) \to Y(n)$ by

$$V_n\big(U_n(x_1 \otimes \cdots \otimes x_n)\big) = p_n(x_1 \otimes \cdots \otimes x_n).$$

It is immediate that V_n preserves inner products (thus it is well defined) and that it maps $X(n)$ onto $Y(n)$. Finally, for all $m, n \in \mathbb{N}$ and $x \in E^{\otimes m}, y \in E^{\otimes n}$,

$$\begin{aligned} V_{m+n}\big(U^X_{m,n}(U_m(x) \otimes U_n(y))\big) &= V_{m+n}\big(U_{m+n}(x \otimes y)\big) \\ &= p_{m+n}(x \otimes y) \\ &= p_{m+n}(p_m x \otimes p_n y) \\ &= p_{m+n}\big((V_m U_m(x)) \otimes (V_n U_n(y))\big) \\ &= U^Y_{m+n}\big((V_m U_m(x)) \otimes (V_n U_n(y))\big), \end{aligned}$$

and (1.2) holds. □

Definition 6.2. *A subproduct system Y satisfying (6.1), (6.2) and (6.3) above will be called a* standard *subproduct system.*

Note that a standard subproduct system is a subproduct subsystem of the full product system $\{E^{\otimes n}\}_{n \in \mathbb{N}}$.

Corollary 6.3. *Every cp-semigroup over $\mathbb{N}$ has an e-dilation.*

Proof. The unital case follows from Corollary 5.10 together with the above lemma. The nonunital case follows from a similar construction (where the dilation of a non-fully-coisometric representation is obtained by adapting [41, Theorem 4.2] instead of [43, Theorem 5.2]). □

Let $k \in \mathbb{N}$, and let $E = X(1), X(2), \ldots, X(k)$ be subspaces of $E, E^{\otimes 2}, \ldots, E^{\otimes k}$, respectively, such that the orthogonal projections $p_n : E^{\otimes n} \to X(n)$ satisfy

$$p_n \leq I_{E^{\otimes i}} \otimes p_j$$

and

$$p_n \leq p_i \otimes I_{E^{\otimes j}}$$

for all $i, j, n \in \mathbb{N}_+$ satisfying $i + j = n \leq k$. In this case one can define a maximal standard subproduct system X with the prescribed fibers $X(1), \ldots, X(k)$ by defining inductively for $n > k$

$$X(n) = \left(\bigcap_{i+j=n} E^{\otimes i} \otimes X(j) \right) \bigcap \left(\bigcap_{i+j=n} X(i) \otimes E^{\otimes j} \right).$$

It is easy to see that

$$X(n) = \bigcap_{n_1 + \ldots + n_m = n} X(n_1) \otimes \cdots \otimes X(n_m) = \bigcap_{i+j=n} X(i) \otimes X(j).$$

We then have obvious formulas for the projections $\{p_n\}_{n\in\mathbb{N}}$ as well, for example

$$p_n = \bigwedge_{i+j=n} p_i \otimes p_j \ , \ (n > k).$$

6.2. Examples.

Example 6.4. In the case $k = 1$, the maximal standard subproduct system with prescribed fiber $X(1) = E$, with E a Hilbert space, is the full product system F_E of Example 1.2. If $\dim E = d$, we think of this subproduct system as the product system representing a (row-contractive) d-tuple $(T_1, \ldots, T_d)$ of non commuting operators, that is, d operators that are not assumed to satisfy any relations (the idea behind this last remark must be rather vague at this point, but it shall become clearer as we proceed). In the case $k = 2$, if $X(2)$ is the symmetric tensor product E with itself then the maximal standard subproduct system with prescribed fibers $X(1), X(2)$ is the symmetric subproduct system SSP_E of Example 1.3. We think of SSP as the subproduct system representing a commuting d-tuple.

Example 6.5. Let E be a two dimensional Hilbert space with basis $\{e_1, e_2\}$. Let $X(2)$ be the space spanned by $e_1 \otimes e_1, e_1 \otimes e_2$, and $e_2 \otimes e_1$. In other words, $X(2)$ is what remains of $E^{\otimes 2}$ after we declare that $e_2 \otimes e_2 = 0$. We think of the maximal standard subproduct system X with prescribed fibers $X(1) = E, X(2)$ as the subproduct system representing pairs (T_1, T_2) of operators subject only to the condition $T_2^2 = 0$. $E^{\otimes n}$ has a basis consisting of all vectors of the form $e_\alpha = e_{\alpha_1} \otimes \cdots \otimes e_{\alpha_n}$ where $\alpha = \alpha_1 \cdots \alpha_n$ is a word of length n in "1" and "2". $X(n)$ then has a basis consisting of all vectors e_α where α is a word of length n not containing "22" as a subword. Let us compute $\dim X(n)$, that is, the number of such words.
Let A_n denote the number of words not containing "22" that have leftmost letter "1", and let B_n denote the number of words not containing "22" that have leftmost letter "2". Then we have the recursive relation $A_n = A_{n-1} + B_{n-1}$ and $B_n = A_{n-1}$. The solution of this recursion gives

$$\dim X(n) = A_n + B_n \approx \left(\frac{1+\sqrt{5}}{2}\right)^n .$$

As one might expect, the dimension of $X(n)$ grows exponentially fast.

Example 6.6. Suppose that we want a "subproduct system that will represent a pair of operators (T_1, T_2) such that $T_i T_2 = 0$ for $i = 1, 2$". Although we have not yet made clear what we mean by this, let us proceed heuristically along the lines of the preceding examples. We let E be as above, but now we declare $e_1 \otimes e_2 = e_2 \otimes e_2 = 0$. In other words, we define $X(2) = \{e_1 \otimes e_2, e_2 \otimes e_2\}^\perp$. One checks that the maximal standard subproduct system X with prescribed fibers $X(1) = E, X(2)$ is given by $X(n) = \mathrm{span}\{e_1 \otimes e_1 \otimes \cdots \otimes e_1, e_2 \otimes e_1 \otimes \cdots \otimes e_1\}$. This is an example of a subproduct system with two dimensional fibers.

At this point two natural questions might come to mind. First, *is every standard subproduct system X the maximal subproduct system with prescribed fibers $X(1), \ldots, X(k)$ for some $k \in \mathbb{N}$?* Second, *does* $\dim X(n)$ *grow exponentially fast (or remain a constant) for every subproduct system X?* The next example answers both questions negatively.

Example 6.7. Let E be as in the preceding examples, and let $X(n)$ be a subspace of $E^{\otimes n}$ having basis the set

$$\{e_\alpha : |\alpha| = n, \alpha \textit{ does not contain the words } 22, 212, 2112, 21112, \ldots\}.$$

Then $X = \{X(n)\}_{n\in\mathbb{N}}$ is a standard subproduct system, but it is smaller than the maximal subproduct system defined by any initial k fibers. Also, $X(n)$ is the span of e_α with $\alpha = 11\cdots 11, 21\cdots 11, 121\cdots 11, \ldots, 11\cdots 12$, thus

$$\dim X(n) = n + 1,$$

so this is an example of a subproduct system with fibers that have a linearly growing dimension.

Of course, one did not have to go far to find an example of a subproduct system with linearly growing dimension: indeed, the dimension of the fibers of the symmetric subproduct system $SSP_{\mathbb{C}^d}$ is known to be

$$\dim SSP_{\mathbb{C}^d}(n) = \left(\begin{array}{c} n+d-1 \\ n \end{array} \right).$$

Taking $d = 2$ we get the same dimension as in Example 6.7. However, $SSP := SSP_{\mathbb{C}^2}$ and the subproduct system X of Example 6.7 are not isomorphic: for any nonzero $x \in SSP(1)$, the "square" $U_{1,1}^{SSP}(x \otimes x) \in SSP(2)$ is never zero, while $U_{1,1}^X(e_2 \otimes e_2) = 0$.
Here is an interesting question that we do not know the answer to: *given a solution $f : \mathbb{N} \to \mathbb{N}$ to the functional inequality*

$$f(m+n) \leq f(m)f(n) \ , \ m, n \in \mathbb{N},$$

does there exists a subproduct system X such that $\dim X(n) = f(n)$ *for all $n \in \mathbb{N}$?*

Remark 6.8. One can cook up simple examples of subproduct systems that are not standard. We will not write these examples down, as we already know that such a subproduct system is isomorphic to a standard one.

6.3. Representations of subproduct systems. Fix a W*-correspondence E. Every completely contractive linear map $T_1 : E \to B(H)$ gives rise to a c.c. representation T^n of the full product system $F_E = \{E^{\otimes n}\}_{n\in\mathbb{N}}$ by defining for all $x \in E^{\otimes n}$ and $h \in H$

$$T^n(x)h = \widetilde{T}_1\big(I_E \otimes \widetilde{T}_1\big) \cdots \big(I_{E^{\otimes(n-1)}} \otimes \widetilde{T}_1\big)(x \otimes h), \tag{6.4}$$

where $\widetilde{T}_1 : E \otimes H \to H$ is given by $\widetilde{T}_1(e \otimes h) = T_1(e)h$. We will denote the operator acting on $x \otimes h$ in the right hand side of (6.4) as $\widetilde{T}^n$, so as not to confuse with $\widetilde{T}_n$, which sometimes has a different meaning (namely: if T denotes a c.c. representation of a subproduct system X then

$$\widetilde{T}_n : X(n) \otimes H \to H$$

is given by

$$\widetilde{T}_n(x \otimes h) = T(x)h$$

for all $x \in X(n), h \in H$. Of course, when $X = F_E$, T is a representation of F_E and T_1 is the restriction of T to E, then $\widetilde{T}^n = \widetilde{T}_n$ for all n). If X is a standard subproduct system and $X(1) = E$, we obtain a completely contractive representation of $X(n)$ by restricting T^n to $X(n)$. Let us denote this restriction by T_n, and denote the family $\{T_n\}_{n\in\mathbb{N}}$ by T.

Proposition 6.9. *Let X be a standard subproduct system with projections $\{p_n\}_{n\in\mathbb{N}}$, and let $T_1 : E \to B(H)$ be a completely contractive map. Construct the family of maps $T = \{T_n\}_{n\in\mathbb{N}}$, with $T_n : X(n) \to B(H)$ as in the preceding paragraph. Then the following are equivalent:*

(1) *T is a representation of X.*
(2) *For all $m, n \in \mathbb{N}$,*

$$\widetilde{T}_m(I_{X(m)} \otimes \widetilde{T}_n)(p_m \otimes p_n \otimes I_H)(p_{m+n}^{\perp} \otimes I_H) = 0. \tag{6.5}$$

(3) *For all $n \in \mathbb{N}$,*

$$\widetilde{T}^n(p_n^{\perp} \otimes I_H) = 0. \tag{6.6}$$

Proof. If T is a representation, then

$$\widetilde{T}_m(I_{X(m)}\otimes\widetilde{T}_n)(p_m\otimes p_n\otimes I_H)(p_{m+n}^{\perp}\otimes I_H) = \widetilde{T}_{m+n}(p_{m+n}\otimes I_H)(p_{m+n}^{\perp}\otimes I_H) = 0,$$

so $1 \Rightarrow 2$. To prove $2 \Rightarrow 3$ note first that (6.6) is clear for $n = 1$. Assuming that (6.6) holds for $n = 1, 2, \ldots, k-1$, we will show that it holds for $n = k$.

$$\begin{aligned}
\widetilde{T}^k(p_k^{\perp} \otimes I_H) &= \widetilde{T}^1(I \otimes \widetilde{T}^{k-1})(p_k^{\perp} \otimes I_H) \\
&= \widetilde{T}^1(I \otimes \widetilde{T}^{k-1})(I_E \otimes p_{k-1}^{\perp} \otimes I_H + I_E \otimes p_{k-1} \otimes I_H)(p_k^{\perp} \otimes I_H) \\
(*) &= \widetilde{T}^1(I \otimes \widetilde{T}^{k-1}(p_{k-1} \otimes I_H))(p_k^{\perp} \otimes I_H) \\
&= \widetilde{T}_1(I \otimes \widetilde{T}_{k-1}(p_{k-1} \otimes I_H))(p_k^{\perp} \otimes I_H) \\
(**) &= 0.
\end{aligned}$$

The equality marked by (*) is true by the inductive hypothesis, and the one marked by (**) follows from (6.5).

Finally, $3 \Rightarrow 1$: by (6.6) we have $\widetilde{T}^n(p_n \otimes I_H) = \widetilde{T}^n$. On the other hand, $\widetilde{T}^n(p_n \otimes I_H) = \widetilde{T}_n(p_n \otimes I_H)$. Thus

$$\begin{aligned}
\widetilde{T}_{m+n}(p_{m+n} \otimes I_H) &= \widetilde{T}^{m+n}(p_{m+n} \otimes I_H) \\
&= \widetilde{T}^{m+n} \\
&= \widetilde{T}^m(I_{X(m)} \otimes \widetilde{T}^n) \\
&= \widetilde{T}_m(I_{X(m)} \otimes \widetilde{T}_n)(p_m \otimes p_n \otimes I_H),
\end{aligned}$$

which shows that T is a representation. □

Proposition 6.10. *Let X be the maximal standard subproduct system with prescribed fibers $X(1), \ldots, X(k)$, and let $T_1 : E \to B(H)$ be a completely contractive map. Construct T as in Proposition 6.9. Then T is a representation of X if and only if*

$$\widetilde{T}^n(p_n^\perp \otimes I_H) = 0 \quad \textit{for all} \quad n = 1, 2, \ldots, k. \tag{6.7}$$

Proof. The necessity of (6.7) follows from Proposition 6.9. By the same proposition, to show that the condition is sufficient it is enough to show that (6.7) holds for all $n \in \mathbb{N}$. Given $m \in \mathbb{N}$, we have $p_m = \bigwedge_q q$, where q runs over all projections of the form $q = I_{X(i)} \otimes p_j$ or $q = p_i \otimes I_{X(j)}$, with $i, j \in \mathbb{N}_+$ and $i + j = m$. But then $p_m^\perp = \bigvee_q q^\perp$, thus if (6.7) holds for all $n < m$ then it also holds for $n = m$. □

6.4. Fock spaces and standard shifts.

Definition 6.11. *Let X be a subproduct system of Hilbert spaces. Fix an orthonormal basis $\{e_i\}_{i \in \mathcal{I}}$ of $E = X(1)$. $X(n)$, when considered as a subspace of $\mathfrak{F}_X$, is called* the n particle space. *The* standard X-shift (related to $\{e_i\}_{i \in \mathcal{I}}$) on $\mathfrak{F}_X$ *is the tuple of operators $\underline{S}^X = \left(S_i^X\right)_{i \in \mathcal{I}}$ in $B(\mathfrak{F}_X)$ given by*

$$S_i^X(x) = U_{1,n}(e_i \otimes x),$$

for all $i \in \mathcal{I}$, $n \in \mathbb{N}$ and $x \in X(n)$.

It is clear that $S_i^X = S^X(e_i)$, where S^X is the shift representation given by Definition 2.9.

If F denotes the usual full product system (Example 1.2) then $\mathfrak{F}_F$ is the usual Fock space and the tuple $(S_i^F)_{i \in \mathcal{I}}$ is the standard shift (the $\mathcal{I}$ orthogonal shift of [37]). We shall denote $\mathfrak{F}_F$ as $\mathfrak{F}$ and $(S_i^F)_{i \in \mathcal{I}}$ as $(S_i)_{i \in \mathcal{I}}$. It is then obvious that the tuple $\left(S_i^X\right)_{i \in \mathcal{I}}$ is a row contraction, as it is the compression of the row contraction $(S_i)_{i \in \mathcal{I}}$. Indeed, assuming (as we may, thanks to Lemma 6.1) that $U_{m,n}$ is an orthogonal projection $p_{m+n} : X(m) \otimes X(n) \to X(m+n)$, and denoting $p = \oplus_n p_n$, we have for all i that $S_i^X = pS_i\big|\mathfrak{F}_X$.

Example 6.12. The q-commuting Fock space of [19] also fits into this framework. Indeed, let (as in [19]) $\Gamma(\mathbb{C}^d)$ be the full Fock space, let $\Gamma_q(\mathbb{C}^d)$ denote the q-commuting Fock space, and let $Y(n)$ be the "n particle q-commuting space" with orthogonal projection $p_n : (\mathbb{C}^d)^n \to Y(n)$. Then a straightforward calculation shows that the projections $\{p_n\}_{n \in \mathbb{N}}$ satisfy equation (6.3) of Lemma 6.1, thus $Y = \{Y(n)\}_{n \in \mathbb{N}}$ is a subproduct system (satisfying (6.1) and (6.2)). With our notation from above we have that $\mathfrak{F}_Y = \Gamma_q(\mathbb{C}^d)$ and that the tuple $(S_i^Y, \ldots, S_d^Y)$ is the standard q-commuting shift.

S^F, the standard shift of the full product system on the full Fock space, will be denoted by S, and will be called simply *the standard shift*.

By the notation introduced in Definition 5.7, the symbol S^X is also used to denote the maximal X-piece of the standard shift S. The following proposition

– which is a generalization of [12, Proposition 6], [19, Proposition 11] and [39, Proposition 2.9] – shows that this is consistent.

PROPOSITION 6.13. *Let X subproduct subsystem of a subproduct system Y. Then the maximal X-piece of the standard Y-shift is the standard X-shift.*

Proof. Let $E = Y(1)$, and let $F = F_E$ be the full product system. Viewing $F(n) \otimes \mathfrak{F}_F$ as direct sum of $|\mathcal{I}|^n$ copies of $\mathfrak{F}_F$, $(\widetilde{S})_n$ is just the row isometry $(S_{i_1} \circ \cdots \circ S_{i_n})_{i_1,\ldots,i_n \in \mathcal{I}}$ from the space of columns $\mathfrak{F}_F \oplus \mathfrak{F}_F \oplus \cdots$ into $\mathfrak{F}_F$. In other words, for $h \in \mathfrak{F}_F$ and $i_1, \ldots, i_n \in I$,

$$(\widetilde{S})_n\big((e_{i_1} \otimes \cdots \otimes e_{i_n}) \otimes h\big) = S_{i_1} \circ \cdots \circ S_{i_n} h = (e_{i_1} \otimes \cdots \otimes e_{i_n}) \otimes h.$$

This is an isometry, and the adjoint works by sending $(e_{i_1} \otimes \cdots \otimes e_{i_n}) \otimes h \in \mathfrak{F}_F$ back to $(e_{i_1} \otimes \cdots \otimes e_{i_n}) \otimes h \in F(n) \otimes \mathfrak{F}_F$, and by sending the $0, 1, \ldots, n-1$ particle spaces to 0.

Now, if Z is any standard subproduct subsystem of F, then

$$\left(\widetilde{S^Z}\right)_n = P_{\mathfrak{F}_Z} \left(\widetilde{S}\right)_n \big|_{Z(n) \otimes \mathfrak{F}_Z},$$

thus

$$\left(\widetilde{S^Z}\right)_n^* = P_{Z(n) \otimes \mathfrak{F}_Z} \left(\widetilde{S}\right)_n^* \big|_{\mathfrak{F}_Z}. \tag{6.8}$$

Now if h is in the k particle space of $\mathfrak{F}_F$ with $k < n$, then $(\widetilde{S^Z})_n^* h = 0$. If $k \geq n$, then since $Z(k) \subseteq Z(n) \otimes Z(k-n)$ we may write $h = \sum \xi_i \otimes \eta_i$, where $\xi_i \in Z(n)$ and $\eta_i \in Z(k-n)$. Thus by (6.8) we find that

$$(\widetilde{S^Z})_n^* \left(\sum \xi_i \otimes \eta_i\right) = \sum p_n^Z \xi_i \otimes p_{k-n}^Z \eta_i = \sum \xi_i \otimes \eta_i. \tag{6.9}$$

From these considerations it follows that the standard X-shift is in fact an X-piece of the standard Y shift, as $(\widetilde{S^Y})_n^*\big|_{\mathfrak{F}_X} = (\widetilde{S^X})_n^*$. It remains to show that the X-shift is maximal.

Assume that there is a Hilbert space H, $\mathfrak{F}_X \subseteq H \subseteq \mathfrak{F}_Y$, such that the compression of S^Y to H is an X-piece of Y, that is, $H \in \mathcal{P}(X, S^Y)$ (see equation (5.3)). Let $h \in H \ominus \mathfrak{F}_X$. We shall prove that $h = 0$. Being orthogonal to all of $\mathfrak{F}_X$, $p_n^Y h$ must be orthogonal to $X(n)$ for all n. Thus, we may assume that $h \in Y(n) \ominus X(n)$ for some n. But then by (6.9)

$$(\widetilde{S^Y})_n^* h = h \otimes \Omega.$$

But since $H \in \mathcal{P}(X, S^Y)$, we must have $h \otimes \Omega \in X(n) \otimes H$, and this, together with $h \in Y(n) \ominus X(n)$, forces $h = 0$. $\square$

7. ZEROS OF HOMOGENEOUS POLYNOMIALS IN NONCOMMUTATIVE VARIABLES

In the next section we will describe a model theory for representations of subproduct systems. But before that we dedicate this section to build a precise connection between subproduct systems together with their representations and tuples of operators that are the zeros of homogeneous polynomials in non commuting variables.

Remark 7.1. The notions that we are developing give a framework for studying tuples of operators satisfying relations given by homogeneous polynomials. One can go much further by considering subspaces of Fock spaces and "representations", i.e., maps of the Fock space into $B(H)$, that give a framework for studying tuples of operators satisfying arbitrary (not-necessarily homogeneous) polynomial and even analytic identities. Gelu Popescu [39] has already begun developing such a theory.

We begin by setting up the usual notation. Let $\mathcal{I}$ be a fixed set of indices, and let $\mathbb{C}\langle(x_i)_{i\in\mathcal{I}}\rangle$ be the algebra of complex polynomials in the non commuting variables $(x_i)_{i\in\mathcal{I}}$. We denote $\underline{x} = (x_i)_{i\in\mathcal{I}}$, and we consider $\underline{x}$ as a "tuple variable". We shall sometimes write $\mathbb{C}\langle\underline{x}\rangle$ for $\mathbb{C}\langle(x_i)_{i\in\mathcal{I}}\rangle$. The set of all words in $\mathcal{I}$ is denoted by $\mathbb{F}_{\mathcal{I}}^{+}$. For a word $\alpha \in \mathbb{F}_{\mathcal{I}}^{+}$, let $|\alpha|$ denote the length of α, i.e., the number of letters in α.
For every word $\alpha = \alpha_1\cdots\alpha_k$ in $\mathcal{I}$ denote $\underline{x}^\alpha = x_{\alpha_1}\cdots x_{\alpha_k}$. If $\alpha = 0$ is the empty word, then this is to be understood as 1. k is also referred to in this context as the *degree* of the monomial x^α. $\mathbb{C}\langle\underline{x}\rangle$ is by definition the linear span over $\mathbb{C}$ of all such monomials, and every element in $\mathbb{C}\langle\underline{x}\rangle$ is called a polynomial. A polynomial is called *homogeneous* if it is the sum of monomials of equal degree. A *homogeneous ideal* is a two-sided ideal that is generated by homogeneous polynomials.
If $\underline{T} = (T_i)_{i\in\mathcal{I}}$ is a tuple of operators on a Hilbert space H and $\alpha = \alpha_1\cdots\alpha_k$ is a word with letters in $\mathcal{I}$, we define

$$\underline{T}^\alpha = T_{\alpha_1}T_{\alpha_2}\cdots T_{\alpha_k}.$$

We define $\underline{T}^0 = I_H$. If $p(x) = \sum_\alpha c_\alpha \underline{x}^\alpha \in \mathbb{C}\langle\underline{x}\rangle$, we define $p(\underline{T}) = \sum_\alpha c_\alpha \underline{T}^\alpha$.
If E is a Hilbert space with orthonormal basis $\{e_i\}_{i\in\mathcal{I}}$, An element $e_{\alpha_1}\otimes\cdots\otimes e_{\alpha_k} \in E^{\otimes k}$ will be written in short form as e_α, where $\alpha = \alpha_1\cdots\alpha_k$. If $p(x) = \sum_\alpha c_\alpha \underline{x}^\alpha \in \mathbb{C}\langle\underline{x}\rangle$, we define $p(e) = \sum_\alpha c_\alpha e_\alpha$. Here e_0 (0 the empty word) is understood as the vacuum vector Ω.

Proposition 7.2. *Let E be a Hilbert space with orthonormal basis $\{e_i\}_{i\in\mathcal{I}}$. There is an inclusion reversing correspondence between proper homogeneous ideals $I \lhd \mathbb{C}\langle\underline{x}\rangle$ and standard subproduct systems $X = \{X(n)\}_{n\in\mathbb{N}}$ with $X(1) \subseteq E$. When $|\mathcal{I}| < \infty$ this correspondence is bijective.*

Proof. Let X be such a subproduct system. We define an ideal

$$I^X := \mathrm{span}\{p \in \mathbb{C}\langle\underline{x}\rangle : \exists n > 0, p(e) \in E^{\otimes n} \ominus X(n)\}. \tag{7.1}$$

Once it is established that I^X is a two-sided ideal the fact that it is homogeneous will follow from the definition. Let $p \in \mathbb{C}\langle\underline{x}\rangle$ be such that $p(e) \in E^{\otimes n} \ominus X(n)$ for some $n > 0$. It suffices to show that for every monomial $\underline{x}^\alpha$ we have that $\underline{x}^\alpha p(\underline{x}) \in I^X$, that is,

$$e_\alpha \otimes p(e) \in E^{\otimes|\alpha|+n} \ominus X(|\alpha| + n).$$

But since X is standard, $X(|\alpha|+n) \subseteq X(|\alpha|) \otimes X(n)$, thus

$$E^{\otimes|\alpha|} \otimes (E^{\otimes n} \ominus X(n)) \subseteq E^{\otimes|\alpha|+n} \ominus X(|\alpha|+n).$$

It follows that I^X is a homogeneous ideal.

Conversely, let I be a homogeneous ideal. We construct a subproduct system X_I as follows. Let $I^{(n)}$ be the set of all homogeneous polynomials of degree n in I. Define

$$X_I(n) = E^{\otimes n} \ominus \{p(e) : p \in I^{(n)}\}. \tag{7.2}$$

Denote by p_n the orthogonal projection of $E^{\otimes n}$ onto $X_I(n)$. To show that X_I is a subproduct system it is enough (by symmetry) to prove that for all $m, n \in \mathbb{N}$

$$p_{m+n} \leq I_{E^{\otimes m}} \otimes p_n,$$

or, in other words, that

$$X_I(m+n) \subseteq E^{\otimes m} \otimes X_I(n). \tag{7.3}$$

Let $x \in X_I(m+n)$, let $\alpha \in \mathcal{I}^m$, and let $q \in I^{(n)}$. Since I is an ideal, $\underline{x}^\alpha q(\underline{x})$ is in $I^{(m+n)}$, thus $\langle x, e_\alpha \otimes q(e)\rangle = 0$. This proves (7.3).

Assume now that $|\mathcal{I}| < \infty$. We will show that the maps $X \mapsto I^X$ and $I \mapsto X_I$ are inverses of each other. Let J be a homogeneous ideal in $\mathbb{C}\langle \underline{x}\rangle$. Then

$$\begin{aligned} I^{X_J} &= \mathrm{span}\{p \in \mathbb{C}\langle \underline{x}\rangle : \exists n > 0, p(e) \in E^{\otimes n} \ominus X_J(n)\} \\ (*) &= \mathrm{span}\{p \in \mathbb{C}\langle \underline{x}\rangle : \exists n > 0, p(e) \in \{q(e) : q \in J^{(n)}\}\} \\ &= \mathrm{span}\{p : \exists n > 0, p \in J^{(n)}\} \\ (**) &= J, \end{aligned}$$

where (*) follows from the definition of X_J, and (**) from the fact that J is a homogeneous ideal.

For the other direction, let Y be a standard subproduct subsystem of $F_E = \{E^{\otimes n}\}_{n\in\mathbb{N}}$. Clearly, $(I^Y)^{(n)} = \{p \in \mathbb{C}\langle \underline{x}\rangle : p(e) \in E^{\otimes n} \ominus Y(n)\}$. Thus

$$\begin{aligned} X_{I^Y}(n) &= E^{\otimes n} \ominus \{p(e) : p \in (I^Y)^{(n)}\} \\ &= E^{\otimes n} \ominus \{p(e) : p \in \{q \in \mathbb{C}\langle \underline{x}\rangle : q(e) \in E^{\otimes n} \ominus Y(n)\} \\ &= E^{\otimes n} \ominus (E^{\otimes n} \ominus Y(n)) \\ &= Y(n). \end{aligned}$$

□

We record the definitions of I^X and X_I from the above theorem for later use:

Definition 7.3. *Let E be a Hilbert space with orthonormal basis $\{e_i\}_{i\in\mathcal{I}}$ ($|\mathcal{I}|$ is not assumed finite). Given a homogeneous ideal $I \triangleleft \mathbb{C}\langle \underline{x}\rangle$, the subproduct system X_I defined by (7.2) will be called the* subproduct system associated with I. *If X is a given subproduct subsystem of F_E, then the ideal I^X of $\mathbb{C}\langle \underline{x}\rangle$ defined by (7.1) will be called the* ideal associated with X.

We note that X_I depends on the choice of the space E and basis $\{e_i\}_{i\in\mathcal{I}}$, but different choices will give rise to isomorphic subproduct systems.

Proposition 7.4. *Let X and Y be standard subproduct systems with $\dim X(1) = \dim Y(1) = d < \infty$. Then X is isomorphic to Y if and only if there is a unitary linear change of variables in $\mathbb{C}\langle x_1, \ldots, x_d\rangle$ that sends I^X onto I^Y.*

Fix some infinite dimensional separable Hilbert space H. As in classical algebraic geometry, given a homogeneous ideal $I \triangleleft \mathbb{C}\langle \underline{x}\rangle$, it is natural to introduce and to study the *zero set of I*

$$Z(I) := \{\underline{T} = (T_i)_{i\in\mathcal{I}} \in B(H)^{\mathcal{I}} : \forall p \in I, p(\underline{T}) = 0\}.$$

Also, given a set $Z \subseteq B(H)^{\mathcal{I}}$, one may form the following two-sided ideal in $\mathbb{C}\langle \underline{x}\rangle$

$$I(Z) := \{p \in \mathbb{C}\langle \underline{x}\rangle : \forall \underline{T} \in Z,\ p(\underline{T}) = 0\}.$$

In the following theorem we shall use the notation of 6.3. This simple result is the justification for viewing subproduct systems as a framework for studying tuples of operators satisfying certain homogeneous polynomial relations.

Theorem 7.5. *Let E be a Hilbert space with orthonormal basis $\{e_i\}_{i\in\mathcal{I}}$ (not necessarily with $|\mathcal{I}| < \infty$), and let I be a proper homogeneous ideal in $\mathbb{C}\langle (x_i)_{i\in\mathcal{I}}\rangle$. Let X_I be the associated subproduct system. Let $T_1 : E \to B(H)$ be a given representation of E. Define a tuple $\underline{T} = (T(e_i))_{i\in\mathcal{I}}$. Construct the family of maps $T = \{T_n\}_{n\in\mathbb{N}}$, with $T_n : X(n) \to B(H)$ as in the paragraphs before Proposition 6.9. Then T is a representation of X if and only if $\underline{T} \in Z(I)$.*

Proof. On the one hand, $E^{\otimes n} \ominus X_I(n) = \overline{\mathrm{span}}\{p(e) : p \in I^{(n)}\}$. On the other hand, for every $p \in I^{(n)}$ and every $h \in H$,

$$\widetilde{T}^n(p(e)\otimes h) = p(\underline{T})h.$$

Hence, the Theorem follows from Proposition 6.9. □

Lemma 7.6. *Let $J \triangleleft \mathbb{C}\langle (x_i)_{i\in\mathcal{I}}\rangle$, $|\mathcal{I}| < \infty$, be a proper homogeneous ideal. Let S^{X_J} be the X_J-shift representation, and define $\underline{T} = (T_i)_{i\in\mathcal{I}}$ by $T_i = S^{X_J}(e_i)$, $i \in \mathcal{I}$. If $p \in \mathbb{C}\langle \underline{x}\rangle$ is a homogeneous polynomial, then $p(\underline{T}) = 0$ if and only if $p \in J$.*

Proof. The "if" part follows from Theorem 7.5. For the "only if" part, let $p \notin J$ be a homogeneous polynomial of degree n. Applying $p(\underline{T})$ to the vacuum vector Ω, we have

$$p(\underline{T})\Omega = Pp(e),$$

where P is the orthogonal projection of $E^{\otimes n}$ onto $X_J(n)$. But as $p \notin J$, $p(e)$ is not in $E^{\otimes n} \ominus X_J(n) = \ker P$, thus $Pp(e) \neq 0$. In particular, $p(\underline{T}) \neq 0$. □

We have the following noncommutative projective Nullstellensatz.

Theorem 7.7. *Let H be a fixed infinite dimensional separable Hilbert space. Let J be a homogeneous ideal in $\mathbb{C}\langle (x_i)_{i\in\mathcal{I}}\rangle$, with $|\mathcal{I}| < \infty$. Then*

$$I(Z(J)) = J.$$

In particular, $Z(J) = \{\underline{0} = (0,0,\ldots)\}$ *if and only if* J *is the ideal generated by all the* $x_i, i \in \mathcal{I}$.

Proof. $I(Z(J)) \supseteq J$ is immediate. To see the converse, first note that equality is obvious when $J = \mathbb{C}\langle \underline{x} \rangle$, so we may assume that J is proper. Also note that since J is homogeneous $Z(J)$ is scale invariant. From this it follows that $I(Z(J))$ is also a homogeneous ideal. Indeed, if $h, g \in H$, and $p(\underline{x}) = \sum_\alpha c_\alpha \underline{x}^\alpha \in I(Z(J))$, then for all $\lambda \in \mathbb{C}$ one has for every tuple $\underline{T} = (T_i)_{i\in\mathcal{I}} \in Z(I)$,

$$0 = \langle p(\lambda \underline{T})h, g\rangle = \sum_k \left(\sum_{|\alpha|=k} c_\alpha \langle \underline{T}^\alpha h, g\rangle \right) \lambda^k,$$

and since a nonzero univariate polynomial has only finitely many zeros, it follows the homogeneous components of p are all in $I(Z(J))$.
Assume now that p is a homogeneous polynomial not in J. Let S^{X_J} be the X_J-shift representation, and define $\underline{T} = (T_i)_{i\in\mathcal{I}}$ by $T_i = S^{X_J}(e_i)$, $i \in \mathcal{I}$. It is clear that $B(H)^{\mathcal{I}}$ contains some unitarily equivalent copy of $\underline{T}$, which we also denote by $\underline{T}$. By Theorem 7.5, $\underline{T} \in Z(J)$. But by Lemma 7.6, $p(\underline{T}) \neq 0$, so $p \notin I(Z(J))$. This completes the proof. □

8. Universality of the shift: universal algebras and models

In [5], Arveson established a model for commuting, row-contractive tuples. Using an idea from that paper that appeared also in [12] and [19] – an idea that rests upon Popescu's "Poisson Transform" introduced in [38] (and pushed forward in [33] and [39]) – we construct below a model for representations of subproduct systems. Roughly speaking, we will show that every representation of a subproduct system X is a piece of a scaled inflation of the shift. Our model should be compared with a similar model obtained by Popescu in [39]. We will also see below that the operator algebra generated by the shift S^X is the universal operator algebra generated by a representation of X.

8.1. Notation for this section. We continue to use the notation set in the previous section. Let X be a standard subproduct system of Hilbert spaces over $\mathbb{N}$, to be fixed throughout this section. Let $p_n : E^{\otimes n} \to X(n)$ be the projections. Denote $E = X(1)$. Let $\{e_i\}_{i\in\mathcal{I}}$ be an orthonormal basis for E, fixed once and for all.
We denote the standard X-shift tuple by $\underline{S^X} = (S^X_i)_{i\in\mathcal{I}}$, and we denote the standard X-shift representation of X on $\mathfrak{F}_X$ by S^X. We consider $\mathfrak{F}_X$ to be a subspace of the full Fock space $\mathfrak{F}$, we denote the full shift by $\underline{S} = (S_i)_{i\in\mathcal{I}}$, and we denote the full shift representation of F on $\mathfrak{F} := \mathfrak{F}_F$ by S.
Given a representation $T : X \to B(H)$, we will write $\underline{T} = (T_i)_{i\in\mathcal{I}}$ for the tuple $(T(e_i))_{i\in\mathcal{I}}$.
We denote by $\mathcal{A}_X$ the unital algebra

$$\mathcal{A}_X := \overline{\mathrm{span}}\{\underline{S^X}^\alpha : \alpha \in \mathbb{F}^+_{\mathcal{I}}\}.$$

We denote by $\mathcal{E}_X$ the operator system

$$\mathcal{E}_X := \overline{\text{span}} \mathcal{A}_X \mathcal{A}_X^*,$$

and by $\mathcal{T}_X = C^*(\underline{S^X})$ the C*-algebra generated by S_i^X, $i \in \mathcal{I}$ and $I_{\mathfrak{F}_X}$. We denote by $\mathcal{K}(\mathfrak{F}_X)$ the algebra of compact operators on $\mathfrak{F}_X$
If T and U are two representations of X on Hilbert spaces H and K, respectively, then we define

$$T \oplus U$$

to be the representation of X on $H \oplus K$ given by $(T \oplus U)(x) = T(x) \oplus U(x)$.
We also define

$$T \otimes I_K$$

to be the representation of X on $H \otimes K$ given by $(T \otimes I_K)(x) = T(x) \otimes I_K$.

8.2. POPESCU'S "POISSON TRANSFORM". After obtaining the results of this section, we discovered that they were obtained earlier by Popescu [39]. We are presenting them here since they are important for the rest of this paper but we leave out some of the arguments.

PROPOSITION 8.1. $\mathcal{K}(\mathfrak{F}_X) \subseteq \mathcal{E}_X$.

Proof. The result follows from the equations

$$I - \sum_{i \in \mathcal{I}} S_i^X \left(S^X\right)_i^* = P_{\mathbb{C}}. \tag{8.1}$$

and

$$(\underline{S^X})^{\beta} \left(I - \sum_{i \in \mathcal{I}} S_i^X \left(S^X\right)_i^* \right) \underline{S^X}^{\alpha *} x = p_{|\beta|} \langle e_\alpha, x \rangle e_\beta.$$

As the elements $p_{|\beta|} e_\beta$ span $\mathfrak{F}_X$, it follows that $\mathcal{K}(\mathfrak{F}_X) \subseteq \mathcal{E}_X$.
Full details can be found in [39, Theorem 1.3] □

Given a representation T of X on a Hilbert space H and given an integer $m \in \mathbb{N}$, we denote by $m \cdot T$ the representation

$$m \cdot T : X \to B(\underbrace{H \oplus H \oplus \cdots \oplus H}_{m \text{ times}})$$

given by $m \cdot T(x) = \underbrace{T(x) \oplus T(x) \oplus \cdots \oplus T(x)}_{m \text{ times}}$. $\underline{T}$ is a row contraction (i.e., $\sum_{i \in \mathcal{I}} T_i T_i^* \leq I_H$) if and only if T is completely contractive. When $\underline{T}$ is a row contraction the *defect operator* $\Delta(\underline{T})$ is defined as

$$\Delta(\underline{T}) = I - \sum_{i \in \mathcal{I}} T_i T_i^*,$$

and the *Poisson Kernel* [38] associated with $\underline{T}$ is the family of isometries $\{K_r(\underline{T})\}_{0 \leq r < 1}$

$$K_r(\underline{T}) : H \to \mathfrak{F} \otimes H,$$

given by

$$K_r\left(\underline{T}\right)h = \sum_{\alpha\in\mathbb{F}_{\mathcal{I}}^+} e_\alpha \otimes \left(r^{|\alpha|}\Delta(r\underline{T})^{1/2}\underline{T}^{\alpha*}h\right).$$

(See the beginning of [38, Section 8] for the remark that $\underline{T}$ has "property (P)", and [38, Lemma 3.2] for the fact that these are isometries). When it makes sense, we also define $K_1\left(\underline{T}\right)$ by the same formula with $r = 1$. The *Poisson transform* is then defined as a map

$$\Phi = \Phi_{\underline{T}} : C^*(\underline{S}) \to B(H)$$

$$\Phi(a) = \Phi_{\underline{T}}(a) = \lim_{r\nearrow 1} K_r\left(\underline{T}\right)^* (a\otimes I)K_r\left(\underline{T}\right).$$

By [38, Theorem 3.8], Φ is a unital, completely positive, completely contractive, satisfies

$$\Phi(\underline{S}^\alpha \underline{S}^{\beta*}) = \underline{T}^\alpha \underline{T}^{\beta*},$$

and is multiplicative on $Alg(\underline{S}, I_{\mathfrak{F}})$, the algebra generated by $\underline{S}$ and $I_{\mathfrak{F}}$ (Φ is in fact an $Alg(\underline{S}, I_{\mathfrak{F}})$-morphism).

Theorem 8.2. *Let T be a c.c. representation of X on H. There exists a unital, completely positive, completely contractive map*

$$\Psi : \mathcal{E}_X \to B(H)$$

that satisfies

$$\Psi\left((\underline{S^X})^\alpha(\underline{S^X})^{\beta*}\right) = \underline{T}^\alpha\underline{T}^{\beta*}\ ,\ \alpha,\beta\in\mathbb{F}_{\mathcal{I}}^+$$

and

$$\Psi(ab) = \Psi(a)\Psi(b)\ ,\ a\in\mathcal{A}_X, b\in\mathcal{E}_X. \tag{8.2}$$

Proof. By the lemma below, the range of $K_r\left(\underline{T}\right)$ is contained in $\mathfrak{F}_X\otimes H$ for all $0\le r<1$, thus

$$(P_{\mathfrak{F}_X}\otimes I_H)K_r\left(\underline{T}\right) = K_r\left(\underline{T}\right).$$

We may then define

$$\begin{aligned}\Psi(\underline{T})\left(\left((\underline{S^X})^\alpha(\underline{S^X})^{\beta*}\right)\right) &= \lim_{r\nearrow 1} K_r\left(\underline{T}\right)^*\left(\left((\underline{S^X})^\alpha(\underline{S^X})^{\beta*}\right)\otimes I\right)K_r\left(\underline{T}\right)\\ (*) &= \lim_{r\nearrow 1} K_r\left(\underline{T}\right)^*\left(\left(\underline{S}^\alpha\underline{S}^{\beta*}\right)\otimes I\right)K_r\left(\underline{T}\right)\\ &= \underline{T}^\alpha\underline{T}^{\beta*},\end{aligned}$$

where in (*) we have made use of the coinvariance of $\mathfrak{F}_X$ under $\underline{S}$. This obviously extends to the desired map on $\mathcal{E}_X$. □

Lemma 8.3. $K_r\left(\underline{T}\right)H\subseteq\mathfrak{F}_X\otimes H$.

Proof. This was proved in [39, Equation (2.5)] for $r = 1$. The same argument (using the fact that p, there, can be chosen homogeneous) works also for $r < 1$. □

8.3. THE UNIVERSAL ALGEBRA GENERATED BY A TUPLE SUBJECT TO HOMOGENEOUS POLYNOMIAL IDENTITIES.

THEOREM 8.4. *$J \lhd \mathbb{C}\langle (x_i)_{i\in\mathcal{I}}\rangle$, be a homogeneous ideal. Then $\mathcal{A}_{X_J}$ is the universal unital operator algebra generated by a row contraction in $Z(J)$, that is: $\mathcal{A}_{X_J}$ is a norm closed unital operator algebra generated by a tuple in $Z(J)$, (namely, $(S_i^{X_J})_{i\in\mathcal{I}}$), and if $\mathcal{B} \subseteq B(H)$ is another norm closed unital operator algebra generated by a row contraction $(T_i)_{i\in\mathcal{I}} \in Z(J)$, then there is a unique unital and completely contractive homomorphism φ of $\mathcal{A}_{X_J}$ onto $\mathcal{B}$, such that $\varphi(S_i^{X_J}) = T_i$ for all $i \in \mathcal{I}$.*

Proof. This follows from Theorems 7.5 and 8.2. □

8.4. A MODEL FOR REPRESENTATIONS: EVERY COMPLETELY BOUNDED REPRESENTATION OF X IS A PIECE OF AN INFLATION OF S^X. We will now construct a model for representations of subproduct systems. In [39, Section 2], a similar but different model – that includes also a fully coisometric part and not only the shift – has been obtained.

THEOREM 8.5. *Let $\underline{T}$ be a completely bounded representation of the subproduct system X on a separable Hilbert space H, and let K be an infinite dimensional, separable Hilbert space. Then for all $r > \|T\|_{cb}$, T is unitarily equivalent to a piece of*

$$S^X \otimes rI_K. \tag{8.3}$$

Moreover, $\|T\|_{cb}$ is equal the infimum of r such that T is a piece of an operator as in (8.3).

Proof. It is known that $\|T\|_{cb} = \|(T_i)_{i\in\mathcal{I}}\|_{row}$, where $T_i = T(e_i)$. Thus if $r > r_0 = \|T\|_{cb}$, then $\sum_{i\in\mathcal{I}} T_iT_i^* \leq r_0^2 I < r^2 I$. Put $W_i = r^{-1}T_i$, so $\sum_{i\in\mathcal{I}} W_iW_i^* \leq r_0^2/r^2 I$. Then $K_1(\underline{W})$ is an isometry (it is equal to $K_{r_0/r}(r/r_0\underline{W})$, and $r/r_0\underline{W}$ is a row contraction). Thus we may define a map (as in the proof of Theorem 8.2)

$$\Psi : B(\mathfrak{F}_X) \to B(H)$$

by

$$\Psi(a) = K_1(\underline{W})^* (a \otimes I) K_1(\underline{W}).$$

Ψ is a normal completely positive unital map that satisfies

$$\Psi\left((\underline{S^X})^\alpha (\underline{S^X})^{\beta *}\right) = \underline{W}^\alpha \underline{W}^{\beta *}, \ \alpha, \beta \in \mathbb{F}_{\mathcal{I}}^+.$$

Since Ψ is normal it has a *normal* minimal Stinespring dilation $\Psi(a) = V^*\pi(a)V$, with $\pi : B(\mathfrak{F}_X) \to B(L)$ a normal $*$-homomorphism and $V : H \to L$ an isometry. It is well known that π is equivalent to a multiple of the identity representation. Thus we obtain, up to unitary equivalence and after identifying H with VH, that $r^{-1}T_i = P_H\pi(S_i^X)P_H = P_H(S_i^X \otimes I_G)P_H$, for some Hilbert space G. To see that T is a piece of $S^X \otimes I_G$ we need to show that

$(S_i^X \otimes I_G)^*\big|_H = T_i^*$ for all $i \in \mathcal{I}$. In other words, we need to show that $P_H\pi(S_i^X) = P_H\pi(S_i^X)P_H$. But, for all $b \in \mathcal{E}_X$,

$$\begin{aligned} P_H\pi(S_i^X)\pi(b)P_H &= P_H\pi(S_i^X b)P_H \\ &= \Psi(S_i^X b) \\ (*) &= \Psi(S_i^X)\Psi(b) \\ &= P_H\pi(S_i^X)P_H\pi(b)P_H, \end{aligned}$$

where (*) follows from (8.2). By Proposition 8.1, the strong operator closure of $\mathcal{E}_X$ is $B(\mathfrak{F}_X)$. $P_H\pi(S_i^X) = P_H\pi(S_i^X)P_H$ now follows from the minimality and normality of the dilation.
It is clear that $r^{-1}T$ is a also piece of $S^X \otimes I_K$ for every K with $\dim K \geq \dim G$, so we may choose K to be infinite dimensional.
We want to show that necessarily $\dim K \geq \dim H$. Since $S^X \otimes I_K$ is a dilation of $r^{-1}T$, $I_L - \sum_{i\in\mathcal{I}} S_i^X(S_i^X)^* \otimes I_K$ is a dilation of $I_H - \sum_{i\in\mathcal{I}} r^{-2}T_iT_i^*$. But the latter operator is invertible so it has rank $\dim H$. Thus the rank of $P_{\mathbb{C}} \otimes I_K = I_L - \sum_{i\in\mathcal{I}} S_i^X(S_i^X)^* \otimes I_K$, which is $\dim K$, must be greater.
Now the final assertion is clear. $\square$

We can now obtain a general von Neumann inequality.

Theorem 8.6. *Let X be a subproduct system, and let T be a c.c. representation of X on a Hilbert space H. Let $\{e_1, \dots, e_d\}$ be an orthonormal set in $X(1)$, and define $T_i = T(e_i)$ and $S_i^X = S^X(e_i)$ for $i = 1, \dots, d$. Then for every polynomials p and q in d non commuting variables,*

$$\|p(T_1, \dots, T_d)q(T_1, \dots, T_d)^*\| \leq \|p(S_1^X, \dots, S_d^X)q(S_1^X, \dots, S_d^X)^*\|.$$

Proof. Since T is a piece of $S^X \otimes rI_K$ for all $r > 1$, we have

$$p(T_1, \dots, T_d)q(T_1, \dots, T_d)^* = P\Big(p(rS_1, \dots, rS_d)q(rS_1, \dots, rS_d)^* \otimes I_K\Big)P$$

for some projection P, and the result follows by taking $r \searrow 1$. $\square$

9. The operator algebra associated to a subproduct system

9.1. Let X be a subproduct system. Recall the definitions of $\mathcal{A}_X$ and $\mathcal{E}_X$ from 8.1. If $\{e_i\}_{i\in\mathcal{I}}$ is an orthonormal basis for $X(1)$, then $\mathcal{A}_X$ is the unital operator algebra generated by $(S_i^X)_{i\in\mathcal{I}}$ with $S_i^X = S^X(e_i)$. If $\{f_i\}_{i\in\mathcal{I}}$ is another orthonormal basis then the tuple $(S^X(f_i))_{i\in\mathcal{I}}$ is not necessarily unitarily equivalent to $(S_i^X)_{i\in\mathcal{I}}$. For instance (with the above notation), if X and $\{e_1, e_2\}$ are as in Example 6.7, and

$$f_1 = \frac{1}{\sqrt{2}}(e_1 + e_2) \quad , \quad f_2 = \frac{1}{\sqrt{2}}(e_1 - e_2),$$

then S_1^X, S_2^X are partial isometries, whereas $T_1 = S^X(f_1)$ and $T_2 = S^X(f_2)$ are not. Thus, the unitary equivalence of the row $(\underline{S}_i^X)$ does not determine the isomorphism class of the subproduct system X.

Proposition 9.1. *Let X and Y be two subproduct systems with $X(1) = E$ and $Y(1) = F$. Assume that $\{e_i\}_{i\in\mathcal{I}}$ is an orthonormal basis for E and that $\{f_i\}_{i\in\mathcal{I}}$ is an orthonormal basis for F. Then the shifts $(S_i^X)_{i\in\mathcal{I}}$ and $(S_i^Y)_{i\in\mathcal{I}}$ are unitarily equivalent as rows (i.e., there exists a unitary $V : \mathfrak{F}_X \to \mathfrak{F}_Y$ such that $VS_i^X = S_i^Y V$ for all $i \in \mathcal{I}$), if and only if there is an isomorphism of subproduct systems $W : X \to Y$ such that $We_i = f_i$ for all $i \in \mathcal{I}$.*

Proof. If X and Y are isomorphic with the isomorphism W sending e_i to f_i, then define a unitary $V : \mathfrak{F}_X \to \mathfrak{F}_Y$ by

$$V = \bigoplus_{n\in\mathbb{N}} W\big|_{X(n)}.$$

$VS_i^X = S_i^Y V$ follows from the properties of W. Conversely, a unitary V intertwining $\underline{S}^X$ and $\underline{S}^Y$ must send Ω_X to Ω_Y. Indeed, such a unitary must send $\{\Omega_X\}^\perp$ (which is equal to $\vee_i \mathrm{Im} S_i^X$) onto a subspace of $\{\Omega_Y\}^\perp$ that has codimension 1 in $\mathfrak{F}_Y$, thus it must send $\{\Omega_X\}^\perp$ onto $\{\Omega_Y\}^\perp$. It follows that $V\Omega_X = \Omega_Y$. Thus, given a unitary V intertwining $\underline{S}^X$ and $\underline{S}^Y$, we may define $W\big|X(n) : X(n) \to Y(n)$ by

$$WS_\alpha^X\Omega = VS_\alpha^X\Omega = S_\alpha^Y\Omega,$$

for all $|\alpha| = n$, and it is easy to see that the maps $W\big|_{X(n)}$ assemble to form an isomorphism of subproduct systems. □

In the example preceding the proposition, we saw how the shift "tuple" (S_1^X, S_2^X) depends essentially on the choice of basis in E. However, the closed unital algebra generated by (S_1^X, S_2^X) is isomorphic to the one generated by (T_1, T_2). Similar remarks hold for $\mathcal{E}_X$ and $\mathcal{T}_X$.

Example 9.2. Let X be the subproduct system given by $X(0) = \mathbb{C}, X(1) = \mathbb{C}^2$ and $X(n) = 0$ for all $n \geq 2$. Let Y be the subproduct system given by $Y(0) = Y(1) = Y(2) = \mathbb{C}$ and $Y(n) = 0$ for all $n \geq 3$. Then since $\mathcal{E}_X$ and $\mathcal{E}_Y$ contain the compact operators on $\mathcal{F}_X$ and $\mathcal{F}_Y$ (the Fock spaces), we have $\mathcal{E}_X = \mathcal{T}_X \cong M_3(\mathbb{C}) \cong \mathcal{T}_Y = \mathcal{E}_Y$.
On the other hand, let $\{e_1, e_2\}$ be an orthonormal basis for $X(1)$. Then if Ω is the vacuum vector, then $\mathcal{A}_X$ is generated by $S^X(\Omega) = I, S^X(e_1), S^X(e_2)$. In the base $\{\Omega, e_1, e_2\}$ for $\mathcal{F}_X$, these operators have the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Thus,

$$\mathcal{A}_X \cong \left\{ \begin{pmatrix} a & 0 & 0 \\ b & a & 0 \\ c & 0 & a \end{pmatrix} \Big| a, b, c \in \mathbb{C} \right\}.$$

On the other hand, $\mathcal{A}_Y$ is generated by

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, S^Y(f_1) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \left(S^Y(f_1)\right)^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

where $\{f_1\}$ is an orthonormal basis for $Y(1)$. Thus

$$\mathcal{A}_Y \cong \left\{ \begin{pmatrix} a & 0 & 0 \\ b & a & 0 \\ c & b & a \end{pmatrix} \Big| a, b, c \in \mathbb{C} \right\}.$$

So $\mathcal{A}_X \ncong \mathcal{A}_Y$ (in $\mathcal{A}_X$ the solutions of $T^2 = 0$ form a two dimensional subspace, and in $\mathcal{A}_Y$ they form a one dimensional subspace).

9.2. $\mathcal{A}_X$ as a graded algebra. For every subproduct system X there exists a unique completely contractive multiplicative linear functional $\rho_0 : \mathcal{E}_X \to \mathbb{C}$ that sends λI to λ and S^X_α to 0 when $|\alpha| > 0$. The existence of ρ_0 follows from Theorem 8.2 (using the Poisson Transform), but it is also clear that ρ_0 is just the vector state associated with the vacuum vector Ω_X:

$$\rho_0(T) = \langle T\Omega_X, \Omega_X \rangle \ , \ T \in \mathcal{A}_X.$$

ρ_0 can be considered also as a conditional expectation $\rho_0 : \mathcal{A}_X \to \mathbb{C} \cdot \Omega_X$, inducing a direct sum

$$\mathcal{A}_X = \rho_0 \mathcal{A}_X \oplus \ker \rho_0 = \mathbb{C} \cdot I \oplus \sum_i S^X_i \mathcal{A}_X. \tag{9.1}$$

$\mathcal{A}_X$ contains a dense graded subalgebra, with the homogeneous elements of degree n being $S^X(\xi)$, where $\xi \in X(n)$. To be precise, we have the following proposition.

Proposition 9.3. *Every $T \in \mathcal{A}_X$ can be written in a unique way as*

$$T = \sum_{n=0}^{\infty} T_n,$$

where $T_n \in \overline{span}\{S^X(\xi) : \xi \in X(n)\}$ and the sum is Cesaro convergent in the norm topology.

Proof. The proof uses a familiar gadget in operator algebra theory, the *gauge action of the torus*. For every $t \in [-\pi, \pi]$, let $W_t : X \to X$ be the subproduct system automorphism given by

$$X(n) \ni \xi \mapsto e^{int}\xi \in X(n).$$

The *gauge action* on $\mathcal{A}_X$ is given by

$$\gamma_t(T) = W_t T W_t^* \ , \ T \in \mathcal{A}_X.$$

Note that if $\alpha \in \mathcal{I}^n$, then

$$\gamma_t(S^X_\alpha) = e^{int} S^X_\alpha,$$

and it follows also that for all $T \in \overline{\text{span}}\{S^X(\xi) : \xi \in X(n)\}$,

$$\gamma_t(T) = e^{int}T.$$

Moreover, for all $T \in \mathcal{A}_X$, the path $t \mapsto \gamma_t(T)$ is strong operator continuous. Given $T \in \mathcal{A}_X$, we define

$$\Phi_n(T) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \gamma_t(T)e^{-int}dt,$$

where this is interpreted in the strong operator sense.
It is easy to show that Φ_n is an idempotent whose range is $\overline{\text{span}}\{S^X(\xi) : \xi \in X(n)\}$.
Define linear maps on $\mathcal{A}_X$ by

$$\Psi_N(T) = \sum_{n=0}^{N}\left(1 - \frac{n}{N}\right)\Phi_n(T).$$

It is then a standard argument (using the Fejer kernel) to prove that $\sum_n \Phi_n(T)$ is Cesaro convergent to T in the norm topology, that is, to show that for all $T \in \mathcal{A}_X$,

$$\|\Psi_N(T) - T\| \stackrel{N\to\infty}{\longrightarrow} 0.$$

It remains to prove the uniqueness assertion. Assume that $T = \sum_n T_n$, where the sum is Cesaro convergent to T, and $T_n \in \overline{\text{span}}\{S^X(\xi) : \xi \in X(n)\}$. Then for all $N > n$,

$$\Phi_n\left(\sum_{m=0}^{N}\left(1 - \frac{m}{N}\right)T_m\right) = \left(1 - \frac{n}{N}\right)T_n \stackrel{N\to\infty}{\longrightarrow} T_n.$$

On the other hand,

$$\Phi_n\left(\sum_{m=0}^{N}\left(1 - \frac{m}{N}\right)T_m\right) \stackrel{N\to\infty}{\longrightarrow} \Phi_n(T),$$

whence $T_n = \Phi_n(T)$. □

9.3. Vacuum state preserving isometric isomorphisms of $\mathcal{A}_X$.

Lemma 9.4. *Let $\varphi : \mathcal{A}_X \to \mathcal{A}_Y$ be an isometric isomorphism. Then φ is unital.*

Proof. A theorem of Arazy and Solel [1] implies that an isometric map between $\mathcal{A}_X$ and $\mathcal{A}_Y$ must send $I \in \mathcal{A}_X$ to an isometry in $\mathcal{A}_X \cap \mathcal{A}_X^*$. It follows that $\varphi(I) = cI$, $|c| = 1$. But since φ is a homomorphism, then $c = 1$. □

Lemma 9.5. *For all $n \in \mathbb{N}$, $\xi \in X(n)$*

$$\|S^X(\xi)\| = \|S^X(\xi)\Omega_X\| = \|\xi\|.$$

Proof. Because $S^X(\xi)$ maps the orthogonal summands $X(k)$ of $\mathfrak{F}_X$ into the orthogonal summands $X(k+n)$, it suffices to show that for all $\eta \in X(k)$, $\|S^X(\xi)\eta\| \leq \|\xi\|\|\eta\|$ (because $S^X(\xi)\Omega_X = \xi$). Now, $S^X(\xi)\eta = p^X_{n+k}(\xi \otimes \eta)$, thus

$$\|S^X(\xi)\eta\|^2 \leq \|\xi \otimes \eta\|^2 = \|\xi\|^2\|\eta\|^2.$$

□

Lemma 9.6. *Let $\varphi : \mathcal{A}_X \to \mathcal{A}_Y$ be an isometric isomorphism that preserves the direct sum decomposition (9.1). Then φ preserves the grading: if $\xi \in X(n)$ then $\varphi(S^X(\xi))$ is in the norm closure of span$\{S^Y(\eta) : \eta \in Y(n)\}$.*

Proof. Since φ is a homomorphism, it suffices to show, say, that $\varphi(S_1^X)$ has "degree one", that is, it is in the norm closure of span$\{S^Y(\eta) : \eta \in Y(1)\}$. By assumption, we may write $\varphi(S_1^X) = \sum_i a_i S_i^Y + T$, with T in the closure of span$\{S^Y(\eta) : \eta \in Y(n), n \geq 2\}$. But $\varphi^{-1}(\sum_i a_i S_i^Y + T) = S_1^X$, and $\varphi^{-1}(T)$ is in the norm closure of span$\{S^X(\xi) : \eta \in X(n), n \geq 2\}$, so $\varphi^{-1}(\sum_i a_i S_i^Y) = S_1^X + B$, with $B = -\varphi^{-1}(T)$ (note that φ^{-1} also preserves the direct sum decomposition (9.1)).
If $T = 0$ then we are done, so assume $T \neq 0$. Then $B \neq 0$, also. But

$$1 = \|S_1^X\| = \|S_1^X \Omega_X\| < \|(S_1^X + B)\Omega_X\| \leq \|S_1^X + B\| = \|\sum_i a_i S_i^Y\|,$$

and at the same time

$$\|\sum_i a_i S_i^Y\| = \|\sum_i a_i S_i^Y \Omega_Y\| < \|(\sum_i a_i S_i^Y + T)\Omega_Y\| \leq$$
$$\leq \|\sum_i a_i S_i^Y + T\| = \|S_1^X\| = 1.$$

From $T \neq 0$ we arrived at $1 < 1$, thus $T = 0$. □

Theorem 9.7. *$X \cong Y$ if and only if $\mathcal{A}_X$ and $\mathcal{A}_Y$ are isometrically isomorphic with an isomorphism that preserves the direct sum decomposition (9.1), and this happens if and only if $\mathcal{A}_X$ and $\mathcal{A}_Y$ are isometrically isomorphic with a grading preserving isomorphism. In fact, if $\varphi : \mathcal{A}_X \to \mathcal{A}_Y$ is a grading preserving isometric isomorphism then there is an isomorphism $V : X \to Y$ such that for all $T \in \mathcal{A}_X$, $\varphi(T) = VTV^*$.*

Proof. $X \cong Y$ implies $\mathcal{A}_X \cong \mathcal{A}_Y$ because these algebras are then generated by unitarily equivalent tuples.
For the converse, we will assume that X and Y are standard subproduct systems. The isomorphism $V : X \to Y$ is defined on the fiber $X(n)$ by

$$V(\xi) = V(S^X(\xi)\Omega_X) = \varphi(S^X(\xi))\Omega_Y \ , \ \xi \in X(n).$$

If it is well defined, then it is onto. Lemma 9.5 shows that V is an isometry on the fibers:

$$\|S^X(\xi)\Omega_X\| = \|S^X(\xi)\| = \|\varphi(S^X(\xi))\| = \|\varphi(S^X(\xi))\Omega_Y\|.$$

Lemma 9.6 implies that $V(\xi)$ sits in $Y(n)$. V respects the subproduct structure: if $m,n\in\mathbb{N}$, $\xi\in X(n),\eta\in X(m)$, then

$$\begin{aligned}
Vp^X_{m,n}(\xi\otimes\eta) &= VS^X(p^X_{m,n}(\xi\otimes\eta))\Omega_X\\
&= \varphi(S^X(p^X_{m,n}(\xi\otimes\eta)))\Omega_Y\\
&= \varphi(S^X(\xi)S^X(\eta))\Omega_Y\\
&= \varphi(S^X(\xi))\varphi(S^X(\eta))\Omega_Y\\
(*) &= p^Y_{m,n}\left(\varphi(S^X(\xi))\Omega_Y\otimes\varphi(S^X(\eta))\Omega_Y\right)\\
&= p^Y_{m,n}(V(\xi)\otimes V(\eta)).
\end{aligned}$$

(*) follows from the facts $S^Y(y)\Omega_Y=y$ and $S^Y(y_1)S^Y(y_2)\Omega_Y=S^Y(p^Y_{m,n}(y_1\otimes y_2))\Omega_Y=p^Y_{m,n}(y_1\otimes y_2)=p^Y_{m,n}(S^Y(y_1)\Omega_Y\otimes S^Y(y_2)\Omega_Y)$.
Finally, let us show that for all $T\in\mathcal{A}_X$, $\varphi(T)=VTV^*$. What we mean by this is that for all $\xi\in X$, $\varphi(S^X(\xi))=VS^X(\xi)V^*$. Let $\varphi(S^X(\eta))\Omega_Y=V(\eta)$ be a typical element in $\mathfrak{F}_Y$.

$$\begin{aligned}
VS^X(\xi)V^*\varphi(S^X(\eta))\Omega_Y &= VS^X(\xi)\eta\\
&= Vp^X(\xi\otimes\eta)\\
&= \varphi(S^X(p^X(\xi\otimes\eta)))\Omega_Y\\
&= \varphi(S^X(\xi)S^X(\eta))\Omega_Y\\
&= \varphi(S^X(\xi))\varphi(S^X(\eta))\Omega_Y,
\end{aligned}$$

This completes the proof. □

10. Classification of the universal algebras of q-commuting tuples

Definition 10.1. *A matrix q is called* admissible *if $q_{ii}=0$ and $0\neq q_{ij}=q_{ji}^{-1}$ for all $i\neq j$.*

10.1. The q-commuting algebras $\mathcal{A}_q$ and their universality. Let $\{e_1,\dots,e_d\}$ be an orthonormal basis for $E:=\mathbb{C}^d$, to be fixed (together with d) throughout this section. Let $q\in M_d(\mathbb{C})$ be an admissible matrix, and let X_q be the maximal standard subproduct system with fibers

$$X_q(1)=E\ ,\ X_q(2)=E\otimes E\ominus\operatorname{span}\{e_i\otimes e_j-q_{ij}e_j\otimes e_i:1\leq i,j\leq d,i\neq j\}.$$

When $q_{ij}=1$ for all $i<j$, then X_q is the symmetric subproduct system SSP. The Fock spaces $\mathfrak{F}_{X_q}$ have been studied in [19].
For brevity, we shall write S^q_i instead of $S^{X_q}_i$. We denote by $\mathcal{A}_q$ the algebra $\mathcal{A}_{X_q}$. By Theorem 8.4, the algebra $\mathcal{A}_q$ is the universal norm closed unital operator algebra generated by a row contraction $(T_1,\dots,T_d)$ satisfying the relations

$$T_iT_j=q_{ij}T_jT_i\ ,\ 1\leq i<j\leq d.$$

10.2. The character space of $\mathcal{A}_q$. Let $\mathcal{M}_q$ be the space of all (contractive) multiplicative and unital linear functionals on $\mathcal{A}_q$, endowed with the weak-$*$ topology. We shall call $\mathcal{M}_q$ *the character space of* $\mathcal{A}_q$. Every $\rho \in \mathcal{M}_q$ is uniquely determined by the d-tuple of complex numbers $(x_1, \ldots, x_d)$, where $x_i = \rho(S_i^q)$ for $i = 1, \ldots, d$. Since a contractive linear functional is completely contractive, $(x_1, \ldots, x_d)$ must be a row contraction, that is, $|x_1|^2 + \ldots + |x_d|^2 \leq 1$. In other words, $(x_1, \ldots, x_d)$ is in the unit ball B_d of $\mathbb{C}^d$. The multiplicativity of ρ implies that $(x_1, \ldots, x_d)$ must lie inside the set

$$Z_q := \{(z_1, \ldots, z_d) \in B_d : (1 - q_{ij}) z_i z_j = 0, 1 \leq i < j \leq d\}.$$

Conversely, Theorem 8.4 implies that every $(x_1, \ldots, x_d) \in Z_q$ gives rise to a character $\rho \in \mathcal{M}_q$ that sends S_i^q to x_i. Thus the map

$$\mathcal{M}_q \ni \rho \mapsto (\rho(S_1^q), \ldots, \rho(S_d^q)) \in Z_q$$

is injective and surjective. It is also obviously continuous (with respect to the weak-$*$ and standard topologies). Since $\mathcal{M}_q$ is compact, we have the homeomorphism

$$\mathcal{M}_q \cong Z_q. \tag{10.1}$$

Note that the vacuum state ρ_0 corresponds to the point $0 \in Z_q \subset \mathbb{C}^d$.
When $q_{ij} = 1$, the condition $(1 - q_{ij}) z_i z_j = 0$ is trivially satisfied, so when $q_{i,j} = 1$ for all i, j, then Z_q is the unit ball B_d. When $q_{ij} \neq 1$, the condition is that either $z_i = 0$ or $z_j = 0$. Thus, if for all i, j, $q_{ij} \neq 1$, then Z_q is the union of d discs glued together at their origins.

10.3. Classification of the $\mathcal{A}_q$, $q_{ij} \neq 1$. Given a permutation σ (on a set with d elements), let U_σ be the matrix that induces the same permutation on the standard basis of $\mathbb{C}^d$.

Proposition 10.2. *Let q and r be two admissible $d \times d$ matrices. Assume that there is a permutation $\sigma \in S_d$ such that $r = U_\sigma q U_\sigma^{-1}$, and let $\lambda_1, \ldots, \lambda_d$ be any complex numbers on the unit circle. Then the map*

$$e_i \mapsto \lambda_i e_{\sigma(i)} \tag{10.2}$$

extends to an isomorphism of X_q onto X_r, and thus the map

$$S_i^q \mapsto \lambda_i S_{\sigma(i)}^r$$

extends to a completely isometric isomorphism between $\mathcal{A}_q$ and $\mathcal{A}_r$.

Proof. For all n, the map (10.2) extends to a unitary V_n of $E^{\otimes n}$. For $n = 2$, this unitary sends $e_i \otimes e_j - q_{ij} e_j \otimes e_i$ to $\lambda_i \lambda_j e_{\sigma(i)} \otimes e_{\sigma(j)} - \lambda_i \lambda_j q_{ij} e_{\sigma(j)} \otimes e_{\sigma(i)}$. But $r = U_\sigma q U_\sigma^{-1}$ implies $r_{\sigma(i)\sigma(j)} = q_{ij}$, thus

$$V_2 : e_i \otimes e_j - q_{ij} e_j \otimes e_i \mapsto \lambda_i \lambda_j e_{\sigma(i)} \otimes e_{\sigma(j)} - \lambda_i \lambda_j r_{\sigma(i)\sigma(j)} e_{\sigma(j)} \otimes e_{\sigma(i)},$$

so V_2 is a unitary between $X_q(2)$ and $X_r(2)$ that respects the product. By induction, it follows that $V = \{V_n\big|_{X_q(n)}\}_n$ is an isomorphism of subproduct systems. The final assertion follows from Proposition 9.1. $\square$

Theorem 10.3. *Let q and r be two admissible $d\times d$ matrices such that $q_{ij}, r_{ij} \neq 1$ for all i,j. Then X_q is isomorphic to X_r if and only if there is a permutation $\sigma \in S_d$ such that $r = U_\sigma q U_\sigma^{-1}$. In this case the isomorphisms are precisely those of the form*

$$e_i \mapsto \lambda_i e_{\sigma(i)},$$

where $\lambda_1, \dots, \lambda_d$ are any complex numbers on the unit circle, and σ is such that $r = U_\sigma q U_\sigma^{-1}$.

Proof. One direction is Proposition 10.2, so assume that there is an isomorphism of subproduct systems $V : X_q \to X_r$. Let $f_i := V^{-1}e_i$. There is a $d \times d$ unitary matrix $U = (u_{ij})$ such that $f_i = \sum_j u_{ij}e_j$. As V is an isomorphism of subproduct systems, we have for all $i \neq j$

$$Vp_2^{X_q}(f_i \otimes f_j - r_{ij} f_j \otimes f_i) = p_2^{X_r}(e_i \otimes e_j - r_{ij} e_j \otimes e_i) = 0,$$

thus

$$(\sum_k u_{ik}e_k) \otimes (\sum_l u_{jl}e_l) - r_{ij}(\sum_k u_{jk}e_k) \otimes (\sum_l u_{il}e_l) \in$$
$$\in \operatorname{span}\{e_m \otimes e_n - q_{mn}e_n \otimes e_m : m \neq n\},$$

or

$$\sum_{k,l}(u_{ik}u_{jl} - r_{ij}u_{jk}u_{il})e_k \otimes e_l \in \operatorname{span}\{e_m \otimes e_n - q_{mn}e_n \otimes e_m : m \neq n\}. \tag{10.3}$$

The coefficients of the vectors $e_k \otimes e_k$ in the sum above must vanish, thus $u_{ik}u_{jk} - r_{ij}u_{jk}u_{ik} = 0$ for all $i \neq j$. Since $r_{ij} \neq 1$, we must have $u_{jk}u_{ik} = 0$ for all k and all $i \neq j$. Thus the unitary matrix U has precisely one nonzero element in each column, and it therefore must be of the form $U_\sigma^{-1}D$, where D is a diagonal unitary matrix.
Equation (10.3) becomes

$$u_{i\sigma(i)}u_{j\sigma(j)}e_{\sigma(i)} \otimes e_{\sigma(j)} - r_{ij}u_{j\sigma(j)}u_{i\sigma(i)}e_{\sigma(j)} \otimes e_{\sigma(i)} \in$$
$$\in \operatorname{span}\{e_m \otimes e_n - q_{mn}e_n \otimes e_m : m \neq n\},$$

but this can only happen if

$$u_{i\sigma(i)}u_{j\sigma(j)}e_{\sigma(i)} \otimes e_{\sigma(j)} - r_{ij}u_{j\sigma(j)}u_{i\sigma(i)}e_{\sigma(j)} \otimes e_{\sigma(i)}$$

is proportional to

$$e_{\sigma(i)} \otimes e_{\sigma(j)} - q_{\sigma(i)\sigma(j)}e_{\sigma(j)} \otimes e_{\sigma(i)},$$

that is $u_{i\sigma(i)}u_{j\sigma(j)}q_{\sigma(i)\sigma(j)} = u_{j\sigma(j)}u_{i\sigma(i)}r_{ij}$, or $r_{ij} = q_{\sigma(i)\sigma(j)}$. Replacing σ with σ^{-1}, the proof is complete. □

Corollary 10.4. *Let q be an admissible $d \times d$ matrix such that there is no permutation $\sigma \in S_d$ such that $q = U_\sigma q U_\sigma^{-1}$. Assume that $q_{ij} \neq 1$ for all i,j. Then the only automorphisms of X_q are unitary scalings of the basis $\{e_1, \dots, e_d\}$.*

Theorem 10.5. *Let q and r be two admissible $d \times d$ matrices such that $q_{ij}, r_{ij} \neq 1$ for all i, j. Then $\mathcal{A}_q$ is isometrically isomorphic to $\mathcal{A}_r$ if and only if there is a permutation $\sigma \in S_d$ such that $r = U_\sigma q U_\sigma^{-1}$. In this case the isometric isomorphisms between $\mathcal{A}_q$ and $\mathcal{A}_r$ are precisely those of the form*

$$S_i^q \mapsto \lambda_i S_{\sigma(i)}^r ,$$

where $\lambda_1, \ldots, \lambda_d$ are any complex numbers on the unit circle.

Proof. If $r = U_\sigma q U_\sigma^{-1}$, then by Proposition 10.2 and Theorem 9.7 $\mathcal{A}_q$ and $\mathcal{A}_r$ are isomorphic (with an isomorphism that preserves the direct sum decomposition (9.1)).
Conversely, assume that $\varphi : \mathcal{A}_q \rightarrow \mathcal{A}_r$ is a completely isometric isomorphism. Then φ induces a homeomorphism between $\mathcal{M}_r$ and $\mathcal{M}_q$ by $\rho \mapsto \rho \circ \varphi$. Recall that $\mathcal{M}_q$ and M_r are both homeomorphic to d discs glued together at the origin. Thus the homeomorphism $\rho \mapsto \rho \circ \varphi$ must take ρ_0 of X_r to ρ_0 of X_q, because these are the unique points in $\mathcal{M}_r$ and $\mathcal{M}_q$, respectively, that when removed from M_r and $\mathcal{M}_q$ leave d disconnected punctured discs. Thus φ sends the vacuum state of $\mathcal{A}_r$ to the vacuum state of $\mathcal{A}_q$, and must therefore preserve the direct sum decomposition (9.1). By Theorem 9.7, there is an isomorphism of subproduct systems $V : X_q \rightarrow X_r$ such that $\varphi(\bullet) = V \bullet V^*$. By Theorem 10.3 we conclude that there is a permutation $\sigma \in S_d$ such that $r = U_\sigma q U_\sigma^{-1}$. It also follows that $\varphi(S_i^q) = \lambda_i S_{\sigma(i)}^r$. □

Corollary 10.6. *Let q be an admissible $d \times d$ matrix such that there is no permutation $\sigma \in S_d$ such that $q = U_\sigma q U_\sigma^{-1}$. Then the only isometric automorphisms of $\mathcal{A}_q$ are unitary scalings of the shift $\{S_1^q, \ldots, S_d^q\}$.*

As a corollary of the above discussion we have:

Corollary 10.7. *Let q and r be two admissible $d \times d$ matrices such that $q_{ij}, r_{ij} \neq 1$ for all i, j. Then $\mathcal{A}_q$ is isometrically isomorphic to $\mathcal{A}_r$ if and only if $X_q \cong X_r$.*

10.4. X_q and $\mathcal{A}_q$, $d = 2$. In the particular case $d = 2$, we let a complex number q parameterize the spaces X_q (we may allow also $q = 0$) defined to be the maximal standard subproduct system with fibers

$$X_q(1) = \mathbb{C}^2 \ , \ X_q(2) = \mathbb{C}^2 \otimes \mathbb{C}^2 \ominus \mathrm{span}\{e_1 \otimes e_2 - q e_2 \otimes e_1\}.$$

Since $\mathcal{M}_1 \cong B_2$, $\mathcal{A}_1$ is not isomorphic to any $\mathcal{A}_q$ with $q \neq 1$ (recall that when $q \neq 1$, $\mathcal{M}_q$ is homeomorphic to two discs glued together at the origin). Thus Theorem 10.5 gives:

Corollary 10.8. *Assume that $d = 2$. Then $X_q \cong X_r$ if and only if $\mathcal{A}_q$ is isometrically isomorphic to $\mathcal{A}_r$, and either one of these happens if and only if either $r = q$ or $r = q^{-1}$.*

Elias Katsoulis has pointed out to us that the above corollary also follows from the techniques of [18].

The above result is reminiscent to the fact that two rotation algebras A_θ and $A_{\theta'}$ are isomorphic if and only if either $e^{2\pi i\theta} = e^{2\pi i\theta'}$ or $(e^{2\pi i\theta})^{-1} = e^{2\pi i\theta'}$. One cannot help but wonder whether one can draw a deeper connection between these results then the superficial one, in particular, can the classification of rotation algebras be deduced from the classification of the algebras $\mathcal{A}_q$?
By Corollaries 10.4 and 10.6 we have the following.

Corollary 10.9. *Let $d = 2$ and let $q \neq 1$. Then subproduct system X_q has no automorphisms aside form the unitary scalings of the basis. The algebra $\mathcal{A}_q$ has no isometric automorphisms other than unitary scalings of the generators.*

On the other hand, a direct calculation shows that every unitary on $\mathbb{C}^2$ extends to an automorphism of X_1, and thus induces a non-obvious automorphism of $\mathcal{A}_1$.

11. Standard maximal subproduct systems with $\dim X(1) = 2$ and $\dim X(2) = 3$

Again, let $\{e_1, \ldots, e_d\}$ be an orthonormal basis for $E := \mathbb{C}^d$. We will soon turn attention to the case $d = 2$. For a matrix $A \in M_d(\mathbb{C})$, we define the *symmetric part* of A to be $A^s := (A + A^t)/2$ and the *antisymmetric part* of A to be $A^a := (A - A^t)/2$. Denote by X_A the maximal standard subproduct system with fibers

$$X_A(1) = E \ , \ X_A(2) = E \otimes E \ominus \operatorname{span}\left\{ \sum_{i,j=1}^{d} a_{ij} e_i \otimes e_j \right\}.$$

We will write S^A for the shift S^{X_A}. We will also write $\mathcal{A}_A$ for $\mathcal{A}_{X_A}$.

Proposition 11.1. *Let $A, B \in M_d(\mathbb{C})$. Then there is an isomorphism $V : X_A \to X_B$ if and only if there exists $\lambda \in \mathbb{C}$ and a unitary $d \times d$ matrix U such that $B = \lambda U^t A U$. In this case, U extends to the isomorphism V between X_A and X_B by $V_1 = U$.*

Proof. Let $V : X_A \to X_B$ be an isomorphism of subproduct systems. There is a $d \times d$ unitary matrix $U = (u_{ij})$ such that

$$f_i := V_1(e_i) = \sum_{j=1}^{d} u_{ij} e_j.$$

Then

$$\begin{aligned} 0 &= V_1(p_2^X(\sum_{i,j} a_{ij} e_i \otimes e_j)) \\ &= p_2^Y(\sum_{i,j} a_{ij} f_i \otimes f_j), \end{aligned}$$

so $\sum_{i,j} a_{ij} f_i \otimes f_j$ must be a spanning vector of span $\left\{\sum_{i,j} b_{ij} e_i \otimes e_j\right\}$. Writing out fully what this means,

$$\lambda \sum_{i,j} a_{ij} \sum_{k,l} u_{ik} u_{jl} e_k \otimes e_l = \sum_{k,l} b_{kl} e_k \otimes e_l$$

for some $\lambda \in \mathbb{C}$, so

$$b_{kl} = \lambda \sum_{i,j} a_{ij} u_{ik} u_{jl}.$$

But the right hand side is precisely the kl-th element of $\lambda U^t AU$.
Conversely, assuming $B = \lambda U^t AU$, one can read the above argument from finish to start to obtain an isomorphism $V : X_A \to X_B$. □

We see that for X_A and X_B to be isomorphic the ranks of A and B must be the same, as well as the ranks of their symmetric and anti-symmetric parts. For example, if A is symmetric and B is not then $X_A \ncong X_B$, a result which may not seem obvious at first glance.

Theorem 11.2. *Assume that $d = 2$. Let $A, B \in M_2(\mathbb{C})$ be any two matrices. Then $\mathcal{A}_A$ is isometrically isomorphic to $\mathcal{A}_B$ if and only if $X_A \cong X_B$, and this happens if and only if there exists $\lambda \in \mathbb{C}$ and a unitary 2×2 matrix U such that $B = \lambda U^t AU$.*

The proof of Theorem 11.2 will occupy the rest of this section. Denote by $\mathcal{M}_A$ the character space of $\mathcal{A}_A$, that is, the topological space of contractive multiplicative and unital linear functionals on $\mathcal{A}_A$, endowed with the weak-$*$ topology.

Lemma 11.3. *The topology of $\mathcal{M}_A$ depends on the rank $r(A^s)$ of the symmetric part A^s of A:*

(1) *If $r(A^s) = 0$ then $\mathcal{M}_A \cong B_2$, the unit ball in $\mathbb{C}^2$.*
(2) *If $r(A^s) = 1$ then $\mathcal{M}_A \cong D$, the unit disc in $\mathbb{C}$.*
(3) *If $r(A^s) = 2$ then $\mathcal{M}_A$ is homeomorphic to two discs pasted together at the origin.*

Proof. We proceed similarly to the lines of 10.2. Every character $\rho \in \mathcal{M}_A$ is uniquely determined by $\lambda_1 = \rho(S_1^A)$ and $\lambda_2 = \rho(S_2^A)$, which lie in B_2. Conversely, every $(\lambda_1, \lambda_2) \in B_2$ that satisfies

$$\sum_{i,j} a_{ij} \lambda_i \lambda_j = 0$$

gives rise to a character ρ by defining $\lambda_1 = \rho(S_1^A)$ and $\lambda_2 = \rho(S_2^A)$. Thus,

$$\mathcal{M}_A \cong V_A := \left\{ (\lambda_i, \lambda_j) \in B_2 : \sum_{i,j} a_{ij} \lambda_i \lambda_j = 0 \right\}.$$

Clearly, $V_A = V_{A^s}$. However, every symmetric 2×2 matrix is complex congruent to one of the following:

$$D_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, D_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ or } D_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

i.e., there exists a nonsingular matrix T such that $A^s = T^t D_i T$, for $i = r(A^s)$. But then $V_{A^s} = T^{-1} V_{D_i} \cong V_{D_i}$, so it remains to verify that V_{D_i} is homeomorphic to the spaces listed in the statement of the lemma. $\square$

Corollary 11.4. *If $r(A^s) \neq r(B^s)$ then $\mathcal{A}_A \ncong \mathcal{A}_B$.*

We can use this corollary to break down the classification of the algebras $\mathcal{A}_A$ to the classification of the algebras $\mathcal{A}_A$ with fixed $r(A^s)$. The easiest case is $r(A^s) = 0$, because then A is either the zero matrix or a multiple of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and these two matrices give rise to non isomorphic algebras (these are the algebras generated by the full and symmetric shifts, respectively).
The next easiest case is $r(A^s) = 2$.

Lemma 11.5. *If $A, B \in M_2(\mathbb{C})$ and $r(A^s) = r(B^s) = 2$, then $\mathcal{A}_A$ is isometrically isomorphic to $\mathcal{A}_B$ if and only if $X_A \cong X_B$, and this happens if and only if there exists $\lambda \in \mathbb{C}$ and a unitary 2×2 matrix U such that $B = \lambda U^t A U$. Any isometric isomorphism between $\mathcal{A}_A$ and $\mathcal{A}_B$ arises as conjugation by the subproduct system isomorphism arising from U.*

Proof. In light of Theorem 9.7 and Proposition 11.1, it suffices to show that any isometric isomorphism $\varphi : \mathcal{A}_A \to \mathcal{A}_B$ sends the vacuum state to the vacuum state. But the vacuum state in $\mathcal{M}_A$ and in $\mathcal{M}_B$ corresponds to the point where the two discs are glued together. Since φ induces a homeomorphism between $\mathcal{M}_B$ and $\mathcal{M}_A$, it must send the vacuum state to the vacuum state. $\square$

Remark 11.6. In the previous section we have seen already that there is a continuum of non-(completely isometrically)-isomorphic algebras $\mathcal{A}_A$ and subproduct systems X_A with $r(A^s) = 2$, namely the algebras $\mathcal{A}_q$. One can see that these algebras $\mathcal{A}_A$ are not exhausted by the algebras $\mathcal{A}_q$ of the previous section. For example, all the algebras $\mathcal{A}_A$ with $A = \begin{pmatrix} 1 & 0 \\ 0 & q \end{pmatrix}$, with $q > 0$, are non-isomorphic, and only for $q = 1$ is this algebra isomorphic to an $\mathcal{A}_q$ (in this case $q = -1$).

We now come to the trickiest case, $r(A^s) = 1$.

Lemma 11.7. *If $A, B \in M_2(\mathbb{C})$ are two symmetric matrices of rank 1, then there exists $\lambda \in \mathbb{C}$ and a unitary 2×2 matrix U such that $B = \lambda U^t A U$, and consequently $X_A \cong X_B$ and $\mathcal{A}_A$ is isometrically isomorphic to $\mathcal{A}_B$.*

Proof. We only have to prove the first assertion, and we may assume that $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. We may also assume that there is a unit vector $v = (v_1, v_2)^t$

such that $A = vv^t$. Now let

$$U = \begin{pmatrix} \overline{v_1} & \overline{v_2} \\ \overline{v_2} & -\overline{v_1} \end{pmatrix}.$$

Then

$$U^t A U = \begin{pmatrix} \overline{v_1} & \overline{v_2} \\ \overline{v_2} & -\overline{v_1} \end{pmatrix}^t vv^t \begin{pmatrix} \overline{v_1} & \overline{v_2} \\ \overline{v_2} & -\overline{v_1} \end{pmatrix} = \begin{pmatrix} \overline{v_1} & \overline{v_2} \\ \overline{v_2} & -\overline{v_1} \end{pmatrix}^t \begin{pmatrix} v_1 & 0 \\ v_2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

□

Below we will also need the following lemma.

Lemma 11.8. *Let A be a 2×2 matrix for which $r(A^s) = 1$. Then there exists one and only one $q \geq 0$ for which there is a $\lambda \in \mathbb{C}$ and a unitary U such that*

$$\begin{pmatrix} 1 & q \\ -q & 0 \end{pmatrix} = \lambda U^t A U.$$

Furthermore, if A is non-symmetric then A is congruent to the matrix

$$\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}.$$

Proof. Direct verification, using Lemma 11.7 and the fact that congruations preserves, up to a scalar, the anti-symmetric part. □

Let us write A_q for the matrix

$$A_q = \begin{pmatrix} 1 & q \\ -q & 0 \end{pmatrix}.$$

By the above lemma, we may restrict attention only to the algebras $\mathcal{A}_{A_q}$ with $q \geq 0$.
Recall that the character space $\mathcal{M}_{A_q}$ of $\mathcal{A}_{A_q}$ is identified with the closed unit disc $\overline{\mathbb{D}}$ by

$$\mathcal{M}_{A_q} \ni \rho \longleftrightarrow \rho(S_2^{A_q}) \in \overline{\mathbb{D}}.$$

We write ρ_z for the character that sends $S_2^{A_q}$ to $z \in \overline{\mathbb{D}}$. This identifies the vacuum vector ρ_0 with the point 0. Recall also that if $\varphi : \mathcal{A}_{A_q} \to \mathcal{A}_{A_r}$ is an isometric isomorphism, then it induces a homeomorphism $\varphi_* : \mathcal{M}_{A_r} \to \mathcal{M}_{A_q}$ given by $\varphi_* \rho = \rho \circ \varphi$. We write F_φ for the homeomorphism $\overline{\mathbb{D}} \to \overline{\mathbb{D}}$ induced by φ, that is, F_φ is the unique self map of $\overline{\mathbb{D}}$ that satisfies

$$\varphi_* \rho_z = \rho_{F_\varphi(z)} \ , \ z \in \overline{\mathbb{D}}.$$

Let us introduce the notation

$$\mathcal{O}(0; q, r) = \{F_\varphi(0) \big| \varphi : \mathcal{A}_{A_q} \to \mathcal{A}_{A_r} \text{ is an isometric isomorphism}\},$$

and

$$\mathcal{O}(0; q) = \mathcal{O}(0; q, q).$$

Lemma 11.9. *Let $q, r \geq 0$. If $q \neq r$ then 0 does not lie in $\mathcal{O}(0; q, r)$.*

Proof. Assume that $0 \in \mathcal{O}(0; q, r)$. Then there is some isometric isomorphism $\varphi : \mathcal{A}_{A_q} \to \mathcal{A}_{A_q}$ that preserves the character ρ_0. It follows from Theorem 9.7 and Proposition 11.1 that, for some unitary 2×2 matrix U and some $\lambda \in \mathbb{C}$, $A_q = \lambda U^t A_r U$. But, as noted in Lemma 11.8, this is impossible if $r \neq q$. $\square$

Lemma 11.10. *The sets $\mathcal{O}(0; q, r)$ are invariant under rotations around* 0.

Proof. For λ with $|\lambda| = 1$, write φ_λ for the isometric isomorphism mapping $S_i^{A_q}$ to $\lambda S_i^{A_q}$ $(i = 1, 2)$. For $b = F_\varphi(0) \in \mathcal{O}(0; q, r)$, consider $\varphi \circ \varphi_\lambda$. We have $\rho_0((\varphi \circ \varphi_\lambda)(S_2^{A_q})) = \rho_0(\varphi(\lambda S_2^{A_q})) = \lambda \rho_0(\varphi(S_2^{A_q})) = \lambda b$. Thus $\lambda b \in \mathcal{O}(0; q, r)$. $\square$

Lemma 11.11. *Let $q, r \geq 0$. If $q \neq r$ then $\mathcal{A}_{A_q}$ is not isometrically isomorphic to $\mathcal{A}_{A_r}$.*

Proof. Assume that $\varphi : \mathcal{A}_{A_q} \to \mathcal{A}_{A_r}$ is an isometric isomorphism. We have $\rho_0 \circ \varphi = \rho_b$, with $b = F_\varphi(0)$, and F_φ is a homeomorphism of $\overline{\mathbb{D}}$ onto itself.
By definition, $b \in \mathcal{O}(0; q, r)$. By Lemma 11.9, $b \neq 0$. Denote $C := \{z : |z| = |b|\}$. By Lemma 11.10, $C \subseteq \mathcal{O}(0; q, r)$. Consider $C' := F_\varphi^{-1}(C)$. We have that $C' \subseteq \mathcal{O}(0; r)$. C' is a simply connected closed path in $\mathbb{D}$ that goes through the origin. By Lemma 11.10, the interior of C', $\mathrm{int}(C')$, is in $\mathcal{O}(0; r)$. But then $F_\varphi(\mathrm{int}(C'))$ is the interior of C, and it is in $\mathcal{O}(0; q, r)$. But then $0 \in \mathcal{O}(0; q, r)$, contradicting Lemma 11.9. $\square$

That concludes the proof of Theorem 11.2.

12. The representation theory of Matsumoto's subshift C*-algebras

In [28] Kengo Matsumoto introduced a class of C*-algebras that arise from symbolic dynamical systems called "subshifts" (we note that in the later paper [17] Carlsen and Matsumoto suggest another way of associating a C*-algebra with a subshift. Here we are discussing only the algebras originally introduced in [28]). These *subshift algebras*, as we shall call them, are strict generalizations of Cuntz-Krieger algebras and have been extensively studied by Matsumoto, T. M. Carlsen and others. For example, the following have been studied: criteria for simplicity and pure-infiniteness; conditions on the underlying dynamical systems for subshift algebras to be isomorphic; the automorphisms of the subshift algebras; K-theory of the subshift algebras; and much more. In this section we will use the framework constructed in the previous sections to give a complete description of all representations of a subshift algebra when the subshift is of finite type.

12.1. Subshifts and the corresponding subproduct systems and C*-algebras. Our references for subshifts are [28] and [16, Chapter 3].
Let $\mathcal{I} = \{1, 2, \ldots, d\}$ be a fixed finite set. $\mathcal{I}^{\mathbb{Z}}$ is the space of all two-sided infinite sequences, endowed with the product topology. The *left shift* (or simply *the shift*) on $\mathcal{I}^{\mathbb{Z}}$ is the homeomorphism $\sigma : \mathcal{I}^{\mathbb{Z}} \to \mathcal{I}^{\mathbb{Z}}$ given by $(\sigma(x))_k = x_{k+1}$. Let

Λ be a shift invariant closed subset of $\mathcal{I}^{\mathbb{Z}}$. By this we mean $\sigma(\Lambda) = \Lambda$. The topological dynamical system $(\Lambda, \sigma\big|_\Lambda)$ is called a *subshift*. Sometimes Λ is also referred to as the subshift.
If W is a set of words in $1, 2, \dots, d$, one can define a subshift by forbidding the words in W as follows:

$$\Lambda_W = \{x \in \mathcal{I}^{\mathbb{Z}} : \text{no word in } W \text{ occurs as a block in } x\}.$$

Conversely, every subshift arises this way: i.e., for every subshift Λ there exists a collection of words W, called *the set of forbidden words*, such that $\Lambda = \Lambda_W$. In this context, if W can be chosen finite then $\Lambda = \Lambda_W$ is called *a subshift of finite type*, or *SFT* for short. By replacing $\mathcal{I}$ if needed, we may always assume that W has no words of length one. If W can be chosen such that the longest word in W has length $k+1$ then Λ is called a *k-step SFT*. A 1-step SFT is also called a *topological Markov chain*. A basic result is that every SFT is isomorphic to a topological Markov chain ([16, Proposition 3.2.1]).
For a fixed subshift $(\Lambda, \sigma\big|_\Lambda)$, we set

$$\Lambda^k = \{\alpha : \alpha \text{ is a word with length } k \text{ occurring in some } x \in \Lambda\},$$

and $\Lambda_l = \cup_{k=0}^{l} \Lambda^k$, $\Lambda^* = \cup_{k=0}^{\infty} \Lambda^k$. With the subshift $(\Lambda, \sigma\big|_\Lambda)$ we associate a subproduct system X_Λ as follows. Let $\{e_i\}_{i \in \mathcal{I}}$ be an orthonormal basis of a Hilbert space E. We define

$$X_\Lambda(0) = \mathbb{C},$$

and for $n \geq 1$ we define

$$X_\Lambda(n) = \operatorname{span}\{e_\alpha : \alpha \in \Lambda^n\}.$$

We define a product $U_{m,n} : X_\Lambda(m) \otimes X_\Lambda(n) \to X_\Lambda(m+n)$ by

$$U_{m,n}(e_\alpha \otimes e_\beta) = \begin{cases} e_{\alpha\beta}, & \text{if } \alpha\beta \in \Lambda^{m+n} \\ 0, & \text{else.} \end{cases}$$

Since $\Lambda^{m+n} \subseteq \Lambda^m \cdot \Lambda^n$, X_Λ is a standard subproduct system.

Definition 12.1. *The C^*-algebra associated with a subshift $(\Lambda, \sigma\big|_\Lambda)$ is defined as the quotient algebra*

$$\mathcal{O}_\Lambda := \mathcal{O}_{X_\Lambda} = \mathcal{T}_{X_\Lambda} / \mathcal{K}(\mathfrak{F}_{X_\Lambda}).$$

Remark 12.2. Just to prevent confusion: In [28], $\mathcal{O}_\Lambda$ was defined as the quotient by the compacts of the C*-algebra generated by the "creation operators" (that is, the X-shift) on $\mathfrak{F}_X$, without using the language of subproduct systems.

12.2. Subproduct systems that come from subshifts.

Proposition 12.3. *Let X be a standard subproduct system such that there is an orthonormal basis $\{e_i\}_{i \in \mathcal{I}}$ of $X(1)$, with $\mathcal{I}$ finite, such that*

1. *Every $X(n)$, $n \geq 1$, is spanned by vectors of the form e_α with $|\alpha| = n$.*
2. *For all $m, n \in \mathbb{N}$, $|\alpha| = n$ and $e_\alpha \in X(n)$, implies that there is some $\beta, \gamma \in \mathcal{I}^m$ such that $e_\beta \otimes e_\alpha$ and $e_\alpha \otimes e_\gamma$ are in $X(m+n)$.*

Then there is a shift invariant closed subset Λ of $\mathcal{I}^{\mathbb{Z}}$ such that $X = X_\Lambda$. X is the maximal standard subproduct system with prescribed fibers $X(1), X(2), \dots, X(k+1)$ if and only if Λ is k-step SFT.

Proof. For all $k \in \mathbb{N}$, define

$$\Lambda^{(k)} = \{\alpha \in \mathcal{I}^k : e_\alpha \in X(k)\}.$$

For all $m \in \mathbb{Z}, k \in \mathbb{N}$, define the closed sets

$$A_{m,k} = \{x \in \mathcal{I}^{\mathbb{Z}} : (x_m, x_{m+1}, \dots, x_{m+k-1}) \in \Lambda^{(k)}\}.$$

Condition (2) implies that $X(k)$ always contains a nonzero vector of the form e_α, $|\alpha| = k$. That implies that the family $\{A_{m,k}\}_{m,k}$ has the finite intersection property. Indeed,

$$A_{m_1,k_1} \cap A_{m_2,k_2} \supseteq A_{M,K} \neq \emptyset,$$

where $M = \min\{m_1, m_2\}$, $K = \max\{m_2 + k_2, m_1 + k_1\} - M$. By compactness of $\mathcal{I}^{\mathbb{Z}}$ we conclude that the closed set

$$\Lambda := \bigcap_{m,k} A_{m,k}$$

is non-empty. Λ is invariant under the left and the right shifts, so $\sigma(\Lambda) = \Lambda$, so $(\Lambda, \sigma\big|_\Lambda)$ is a subshift. By condition (2), $\Lambda^k = \Lambda^{(k)}$. Condition (1) together with the definition of X_Λ now imply that $X = X_\Lambda$.
The final assertion follows from the following facts, together with $X = X_\Lambda$. Fact number one:

$$E^{\otimes n} \ominus X_\Lambda(n) = \operatorname{span}\{e_\alpha : \alpha \text{ is a forbidden word of length } n\}.$$

Fact number two: X is the maximal standard subproduct system with prescribed fibers $X(1), \dots, X(k+1)$ if and only if for every $n > k+1$,

$$X(n) = \bigcap_{i+j=n} X(i) \otimes X(j),$$

or in other words, if and only if

$$\begin{aligned} E^{\otimes n} \ominus X(n) &= \bigvee_{i+j=n} \left(E^{\otimes n} \ominus (X(i) \otimes X(j))\right) \\ &= \bigvee_{i+j=n} \left(E^{\otimes i} \otimes (E^{\otimes j} \ominus X(j)) + (E^{\otimes i} \ominus X(i)) \otimes E^{\otimes j}\right). \end{aligned}$$

Fact number three: Λ is a k-step SFT if and only if for every $n > k+1$,

$$\{\text{forbidden words of length } n\} =$$

$$\bigcup_{i+j=n} \left(\mathcal{I}^i \cdot \{\text{forbidden words of length } j\} \cup \{\text{forbidden words of length } i\} \cdot \mathcal{I}^j\right).$$

These facts assemble together to complete the proof. □

Not every subproduct system is isomorphic to one that comes from a subshift. Indeed, in the symmetric subproduct system SSP (see Example 1.3) for any basis $\{e_i\}_{i\in\mathcal{I}}$ of $X(1)$, the product $e_i \otimes e_j$ for $i \neq j$ is never in $X(2)$, and thus the images f_i and f_j of e_i and e_j in any isomorphic subproduct system X can never be such that $f_i \otimes f_j$ is mapped isometrically to $U^X_{1,1}(f_i \otimes f_j)$. Thus if SSP is isomorphic to X_Λ for some subshift Λ, then Λ must be the subshift containing only constant sequences. But such X_Λ is clearly not isomorphic to SSP.
As another example, the subproduct system $X(0) = \mathbb{C}$, $X(1) = \mathbb{C}^2$, and $X(n) = 0$ for $n > 1$, cannot be of the form X_Λ for any $\Lambda \subseteq \mathcal{I}^{\mathbb{Z}}$.

12.3. The representation theory of the C*-algebra associated with a subshift of finite type. Let Λ be a fixed subshift in $\mathcal{I}^{\mathbb{Z}}$ (with $\mathcal{I} = \{1, 2, \ldots, d\}$), and let $X = X_\Lambda$ be the associated subproduct system. We will denote the X-shift by S (instead of S^X) to make some formulas more readable. Let Z_i be the image of S_i in the quotient $\mathcal{O}_\Lambda$. We define for $i \in \mathcal{I}$, $k \in \mathbb{N}$ the sets

$$E^k_i = \{\alpha \in \Lambda^k : i\alpha \in \Lambda^*\}.$$

Lemma 12.4. *If Λ is a k-step SFT, then for all $i \in \mathcal{I}$,*

$$\{\gamma \in \Lambda^* : |\gamma| \geq k, i\gamma \in \Lambda^*\} = \{\alpha\beta \in \Lambda^* : \alpha \in E^k_i, \beta \in \Lambda^*\}.$$

Proof. Assume that $\gamma \in \Lambda^*$ is such that $|\gamma| \geq k$ and $i\gamma \in \Lambda^*$. Defining $\alpha = \gamma_1 \cdots \gamma_k$ and $\beta = \gamma_{k+1} \cdots \gamma_{k+l}$, we have that $\gamma = \alpha\beta$ where $\alpha \in E^k_i$ and $\beta \in \Lambda^*$.
Conversely, if $\gamma = \alpha\beta \in \Lambda^*$ where $\alpha \in E^k_i$ and $\beta \in \Lambda^*$, then $i\gamma$ must be in Λ^*. Indeed, if not, then $i\gamma$ must contain a forbidden word. But $\gamma \in \Lambda^*$, thus the forbidden word must be in $i\alpha$ (since Λ is a k-step SFT). But that is impossible because $\alpha \in E^k_i$. □

Lemma 12.5. *If Λ is a k-step SFT then for all $i, j \in \mathcal{I}$, $i \neq j$,*

$$S^*_i S_j = 0,$$

and

$$S^*_i S_i = \sum_{\alpha \in E^k_i} \underline{S}^\alpha \underline{S}^{\alpha *} \mod \mathcal{K}_X. \tag{12.1}$$

Consequently, $\mathcal{E}_X = \mathcal{T}_X$.

Proof. Since the S_i are partial isometries with orthogonal ranges, we have $S^*_i S_j = 0$ for all $i \neq j$. Since $\mathcal{K}_X \subseteq \mathcal{E}_X \subseteq \mathcal{T}_X$ (Proposition 8.1), $\mathcal{E}_X = \mathcal{T}_X$ will be established once we prove (12.1).
$S^*_i S_i$ is the projection onto the initial space of S_i. Call this space G. We have

$$G = \mathrm{span}\{e_\alpha : \alpha \in \Lambda^* \text{ such that } i\alpha \in \Lambda^*\}.$$

The space

$$G' = \mathrm{span}\{e_\alpha : \alpha \in \Lambda^* \text{ such that } i\alpha \in \Lambda^* \text{ and } |\alpha| \geq k\}$$

has finite codimension in G. But by Lemma 12.4,

$$G' = \{e_{\alpha\beta} : \alpha\beta \in \Lambda^*, \alpha \in E_i^k\},$$

that is, G' is spanned by e_γ where γ runs through all legal words beginning with some $\alpha \in E_i^k$. Thus, G' is the range of the projection $\sum_{\alpha\in E_i^k} \underline{S}^\alpha \underline{S}^{\alpha^*}$. Since G' has finite codimension in G, we have (12.1). □

PROPOSITION 12.6. *For every subshift Λ, the d-tuple $\underline{Z} = (Z_1, \dots, Z_d)$ satisfies the following relations:*

$$p(\underline{Z}) = 0 \text{ , } \textit{for all } p \in I^X, \tag{12.2}$$

$$Z_i^* Z_j = 0 \text{ , } \textit{for all } i, j \in \mathcal{I}, i \neq j, \tag{12.3}$$

and

$$\sum_{i=1}^d Z_i Z_i^* = 1. \tag{12.4}$$

In particular, Z_i is a partial isometry for all $i \in \mathcal{I}$. If Λ is a k-step SFT, the $\underline{Z}$ also satisfies

$$Z_i^* Z_i = \sum_{\alpha\in E_i^k} \underline{Z}^\alpha \underline{Z}^{\alpha *} \text{ , } \textit{for all } i \in \mathcal{I}. \tag{12.5}$$

Proof. The quotient map $\mathcal{T}_X \to \mathcal{O}_\Lambda$ is a $*$-homomorphism, so (12.2) follows from Theorem 7.5. (12.3) and (12.5) follow from the previous lemma, and (12.4) follows ¿¿from equation (8.1). □

THEOREM 12.7. *Let Λ be a k-step SFT. Every unital representation $\pi : \mathcal{O}_\Lambda \to B(H)$ is determined by a row-contraction $\underline{T} = (T_1, \dots, T_d)$ satisfying relations (12.2)-(12.5) such that $\pi(Z_i) = T_i$ for all $i \in \mathcal{I}$. Conversely, every row contraction in $B(H)^d$ satisfying the relations (12.2)-(12.5) gives rise to a unital representation $\pi : \mathcal{O}_\Lambda \to B(H)$ such $\pi(Z_i) = T_i$ for all $i \in \mathcal{I}$.*

Proof. It is the second assertion that is non-trivial, and we will try to convince that it is true. By Theorem 8.2, there is unital completely positive map

$$\Psi : \mathcal{E}_X \to B(H)$$

sending $\underline{S}^\alpha \underline{S}^{\beta *}$ to $\underline{T}^\alpha \underline{T}^{\beta *}$. Since enough of the rank one operators on $\mathfrak{F}_X$ arise as $\underline{S}^\alpha (I - \sum_{i=1}^d S_i S_i^*) \underline{S}^{\beta *}$ (see equation (8.1)), and because $\underline{T}$ satisfies (12.4), we must have that $\Psi(K) = 0$ for every $K \in \mathcal{K}(\mathfrak{F}_X)$. By Lemma 12.5, $\mathcal{E}_X = \mathcal{T}_X$, and it follows that Ψ induces a positive and unital (hence contractive) mapping

$$\pi : \mathcal{O}_\Lambda \to B(H)$$

that sends $\underline{Z}^\alpha \underline{Z}^{\beta *}$ to $\underline{T}^\alpha \underline{T}^{\beta *}$. Roughly speaking: π must be multiplicative because $\underline{Z}$ and $\underline{T}$ satisfy the same relations. In more detail: every product $(\underline{Z}^\alpha \underline{Z}^{\beta *})(\underline{Z}^{\alpha'} \underline{Z}^{\beta' *})$ may be written, using the relations (12.2)-(12.5) as some sum $\sum_{\gamma,\delta} \underline{Z}^{\gamma}\underline{Z}^{\delta *}$. The mapping π then takes this sum to $\sum_{\gamma,\delta} \underline{T}^{\gamma}\underline{T}^{\delta *}$, and this can be rewritten (using the same relations) as

$$(\underline{T}^{\alpha}\underline{T}^{\beta *})(\underline{T}^{\alpha'}\underline{T}^{\beta' *}) = \pi(\underline{Z}^{\alpha}\underline{Z}^{\beta *})\pi(\underline{Z}^{\alpha'}\underline{Z}^{\beta' *}).$$

This shows that

$$\pi\left((\underline{Z}^{\alpha}\underline{Z}^{\beta *})(\underline{Z}^{\alpha'}\underline{Z}^{\beta' *})\right) = \pi(\underline{Z}^{\alpha}\underline{Z}^{\beta *})\pi(\underline{Z}^{\alpha'}\underline{Z}^{\beta' *}),$$

and since the elements of the form $\underline{Z}^{\alpha}\underline{Z}^{\beta *}$ span $\mathcal{O}_{\Lambda}$, and since π is a positive linear map, it follows that π is in fact a $*$-representation. $\square$

References

1. J. Arazy and B. Solel, *Isometries of non-self-adjoint operator algebras*, J. Funct. Anal., Vol. 90 (1990), 284–305.
2. Wm. B. Arveson, *Subalgebras of C^*-algebras*, Acta Math., Vol. 123 (1969), 141–224.
3. Wm. B. Arveson, *Continuous analogues of Fock space*, Mem. Amer. Math. Soc., No. 409, American Mathematical Society, 1989.
4. Wm. B. Arveson, *The index of a quantum dynamical semigroup*, J. Funct. Anal., Vol. 146, No. 2 (1997), 557588.
5. Wm. B. Arveson, *Subalgebras of C^*-algebras III: Multivariable operator theory*, Acta Math., Vol. 181 (1998), 159–228.
6. Wm.B. Arveson, *Non commutative dynamics and E-semigroups*, Springer Monographs in Math., Springer-Verlag, 2003.
7. Wm. B. Arveson and D. Courtney, *Lifting endomorphisms to automorphisms*, preprint, arXiv:math/0703115v2 [math.OA], 2007.
8. Wm. B. Arveson and A. Kishimoto, *A note on extensions of semigroups of $*$-endomorphisms*, Proc. Amer. Math. Soc. Vol. 116, No. 3 (1992), 769–774.
9. B. V. R. Bhat, *An index theory for quantum dynamical semigroups*, Trans. Amer. Math. Soc. Vol. 348, No. 2 (1996), 561–583.
10. B. V. R. Bhat *A generalized intertwining lifting theorem*, in: Operator Algebras and Applications, II, Waterloo, ON, 1994-1995, in: Fields Inst. Commun., vol. 20, Amer. Math. Soc., Providence, RI, 1998, pp. 1–10.
11. B. V. R. Bhat and T. Bhattacharyya, *A model theory for q-commuting contractive tuples*, J. Operator Theory, Vol. 47 (2002), 97–116.
12. B. V. R. Bhat, T. Bhattacharyya and S. Dey, *Standard non commuting and commuting dilations of commuting tuples*, Trans. Amer. Math. Soc., Vol. 356, No. 4 (2003), 1551–1568.
13. B. V. R. Bhat, S. Dey and J. Zacharias, *Minimal Cuntz–Krieger dilations and representations of Cuntz–Krieger algebras*, Proc. Indian Acad. Sci., Math. Sci. 116, No. 2 (2006), 193–220.
14. B. V. R. Bhat and M. Mukherjee, *Inclusion systems and amalgamated products of product systems.* arXiv:0907.0095, 2009.

15. B. V. R. Bhat and M. Skeide, *Tensor product systems of Hilbert modules and dilations of completely positive semigroups*, Infinite Dimensional Analysis, Quantum Probability and Related Topics, Vol. 3, (2000), 519–575.
16. M. Brin and G. Stuck, *Introduction to Dynamical Systems*, Cambridge University Press, 2002.
17. T. M. Carlsen and K. Matsumoto, *Some remarks on the C^*-algebras associated with subshifts*, Math. Scand., Vol. 95, No. 1 (2004), 145–160.
18. K. Davidson and E. Katsoulis, *Isomorphisms between topological conjugacy algebras*, J. Reine Angew. Math., Vol. 621 (2008), 29–51.
19. S. Dey, *Standard dilations of q-commuting tuples*, Colloquium Math. Vol. 107 (2007), 141–165.
20. R. Douglas *On Extending Commutative Semigroups of Isometries*, Bull. London Math. Soc., Vol. 1 (1969), 157–159.
21. N. J. Fowler, *Discrete product systems of Hilbert bimodules*, Pac. J. Math., Vol. 204, No. 2 (2002), 335–375.
22. R. V. Kadison and J. Ringrose, *Fundamentals of the Theory of Operator Algebras, vol. I*, Academic Press, New-York, 1982.
23. R. V. Kadison and J. Ringrose, *Fundamentals of the Theory of Operator Algebras, vol. II*, Academic Press, New-York, 1983.
24. Y. Katznelson, *An Introduction to Harmonic analysis*, 2nd. ed., Dover Publications, New York, 1976.
25. Marcelo Laca, *From endomorphisms to automorphisms and back: dilations and full corners* J. London Math. Soc., II Ser. 61, No. 3 (2000), 893-904.
26. E. C. Lance, *Hilbert C^*-modules: A toolkit for operator algebraists*, 124 of *London Mathematical Society Lecture Note Series,* Cambridge Univ. Press, Cambridge, 1994.
27. D. Markiewicz, *On the product system of a completely positive semigroup*, J. Funct. Anal.. Vol. 200, No. 1 (2003), 237–280.
28. K. Matsumoto, *On C^*-algebras associated with subshifts*, Internat. J. Math., Vol. 8, No. 3 (1997), 357–374.
29. P. Muhly and B. Solel, *Quantum Markov Processes (Correspondences and Dilations)*, Internat. J. Math. Vol. 13, No. 8 (2002), 863–906.
30. P. Muhly and B. Solel, *The curvature and index of completely positive maps*, Proc. Lond. Math. Soc., III. Ser. 87, No. 3 (2003), 748–778.
31. P. Muhly and B. Solel, *Duality of W^*-correspondences and applications*, Quantum Probability and Infinite Dimensional Analysis – From Foundations to Applications (M. Schürman and U. Franz, eds.), Quantum Probability and White Noise Analysis, No. XVIII, World Scientific, 2005, 396–414.
32. P. Muhly and B. Solel, *Quantum Markov Semigroups (Product systems and subordination)*, Internat. J. Math. Vol. 18, No. 6 (2007),633–669.
33. P. Muhly and B. Solel, *The Poisson Kernel for Hardy Algebras*, Complex Anal. Oper. Theory 3 (2009), 221-242.
34. S. Parrot, *Unitary dilations for commuting contractions*, Pacific J. Math. Vol. 34 (1970), 481–490.

35. K. R. Parthasarathy, *An Introduction to Quantum Stochastic Calculus*, Monographs in Mathematics, Vol. 85, Birkhauser, 1992.
36. W. Paschke, *Inner product modules over B^*-algebras*, Trans. Amer. Math. Soc. 182 (1973), 443-468.
37. G. Popescu, *Isometric Dilations for Infinite Sequences of Noncommuting Operators*, Transactions of the American Mathematical Society, Vol 316, No. 2 (1989), 523–536.
38. G. Popescu, *Poisson Transforms on Some C^*-Algebras Generated by Isometries*, J. Funct. Anal., Vol. 161, No. 1 (1999), 27–61.
39. G. Popescu, *Operator Theory on Noncommutative Varieties*, Indiana Univ. Math. J., Vol. 55, No. 2 (2006), 389–441.
40. D. SeLegue, *Minimal Dilations of CP maps and C^*-Extension of the Szegö Limit Theorem,* Ph.D Dissertation, University of California, Berkeley, 1997.
41. O. M. Shalit, *Representing a product system representation as a contractive semigroup and applications to regular isometric dilations*, to appear in Canad. Math. Bull.
42. O. M. Shalit, *E-dilation of strongly commuting CP-semigroups (the nonunital case)*, preprint, arXiv:0711.2885v3 [math.OA], 2007.
43. O. M. Shalit, *E_0-dilation of strongly commuting CP_0-semigroups*, J. Funct. Anal., Vol. 255, No. 1 (2008), 46–89.
44. M. Skeide, *Isometric Dilations of Representations of Product Systems via Commutant*, Internat. J. Math., Vol. 19, No. 5 (2008), 521–539.
45. M. Skeide, *Spatial E_0-semigroups are restrictions of inner automorphism groups* , In L. Accardi, W. Freudenberg, and M. Schürmann, editors, *Quantum Probability and Infinite Dimensional Analysis — Proceedings of the 26th Conference*, number XX in Quantum Probability and White Noise Analysis, pages 348–355. World Scientific, 2007.
46. M. Skeide, *Product Systems; a Survey with Commutants in View*, to appear in the proceedings of the 2006 Quantum Probability Conference in Nottingham.
47. M. Słociński, *Unitary dilation of two-parameter semi-groups of contractions*, Bull. Acad. Polon. Sci. Se'r. Sci. Math. Astronom. Phys., Vol. 22 (1974), 1011–1014.
48. B. Solel, *Representations of product systems over semigroups and dilations of commuting CP maps*, J. Funct. Anal., Vol. 235, No. 2 (2006) 593–618.
49. B. Sz.-Nagy and C. Foiaş, *Harmonic Analysis of Operators in Hilbert Space*, North-Holland, Amsterdam, 1970.

Orr Moshe Shalit
Department of Pure Mathematics
University of Waterloo
200 University Avenue West
Waterloo, ON, Canada N2L 3G1.
oshalit@math.uwaterloo.ca

Baruch Solel
Department of Mathematics,
Technion,
32000, Haifa, Israel.
mabaruch@tx.technion.ac.il