Gordon Moore and His Law: Numerical Methods to the Rescue

Raúl Rojas

Abstract. In this chapter we review the protracted history of "Moore's Law", that is, the expected doubling of the number of transistors in semiconductor chips every 18 months. Such an exponential increase has been possible due to steady improvements in optical imaging methods. The wavelength of light used for photolithography has been reduced every decade, but it is reaching tough limits. Mathematical methods capable of simulating optical systems and their interference properties play now a significant role in semiconductor design and have kept Moore's Law alive for at least the last ten years. As we show, advances in semiconductor integration and numerical optimization methods act synergistically.

2010 Mathematics Subject Classification: 00A69, 01A61
Keywords and Phrases: Fourier optics, photolithography, Moore's law, numerical simulation

1 Introduction

The number of transistors in a modern chip doubles every 18 months: this is the most common mentioned variation of Moore's Law. Actually, what Gordon Moore postulated in 1965 was an annual doubling of electronic components in semiconductor chips. He was talking about resistances, capacitors, and, of course, logic elements such as transistors [10]. In his now famous paper he compared different manufacturing technologies at their respective life-cycle peaks, that is, when they reached minimal production cost. Fig. 1 is the famous graph from Moore's paper. Notice that he extrapolated future growth based on just a few empirical points.

Moore corrected his prediction ten years later, when, looking back to the previous decade, he modified his prediction to a doubling of electronic components every 24 months: "The complexity for minimum component costs has increased at a rate of roughly a factor of two per year" [11]. Finally, the community of semiconductor experts settled somehow on a doubling period of 18

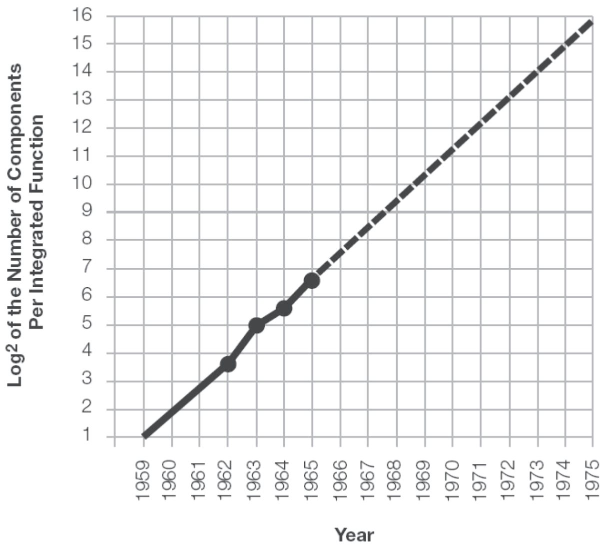

Figure 1: The extrapolated growth curve from Moore's paper of 1965 [10]. Originally Gordon Moore proposed a doubling of components on a chip every 12 months.

months (referring now just to transistors on a chip), which is the modern version of Moore's Law [4]. This prediction has proved very resilient and has been applied to memory chips, microprocessors, and other components, so that we are really faced with a "family" of Laws, all postulating an exponential increase in the number of components per chip (see Fig. 2).

Although more and more transistors can be integrated on a chip every year, and a specific mix of technologies has been responsible for this achievement (for example by designing three-dimensional semiconductor structures [12]), the width of the smallest structures that can be "engraved" on a chip is still the most important parameter in the semiconductor industry. We then talk about chips built with 200 nm, or 100 nm, or even 22 nm technologies. What we mean by this is that photolithographic methods can project small details of that width on layer after layer of semiconductors. The desired two-dimensional logical components are projected on the silicon wafer using a mask and light. Chemicals are used to dissolve, or preserve, the portions of the wafer exposed to light. This so-called photolithography allows engineers to build a chip step by step, like a sandwich of materials and interconnections. The whole process resembles the old photographic methods where an image was produced by exposing the substrate to light, and then chemicals were applied in order to obtain the finished picture. Such projection-processing steps are repeated for different layout masks until a memory chip or microprocessor is packaged.

The problem with optical lithography is that it requires high-quality and

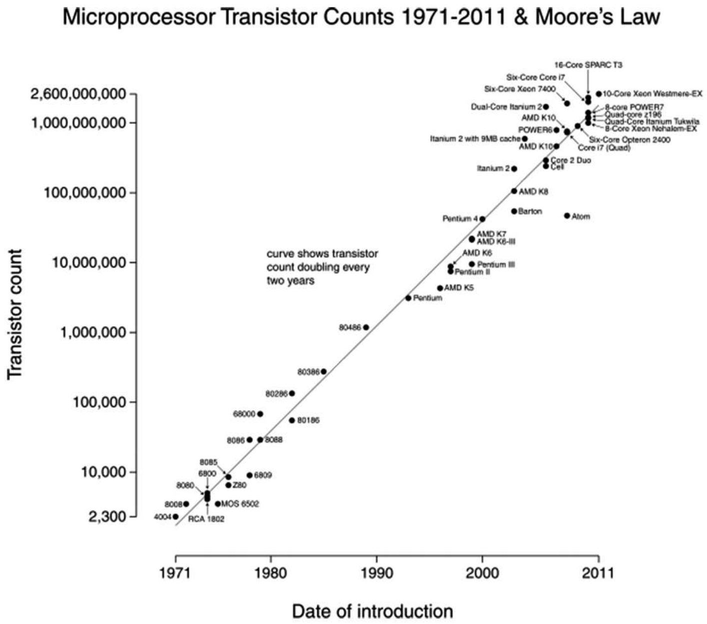

Figure 2: The modern Moore's law interpolated from the transistor count of popular microprocessors (illustration from Wikipedia)

extremely accurate lenses. It is also hampered by the wavelength of the light used for projecting the masks. The width of the current smallest structures projected on commercial chips (22 nm) is already much smaller than the wavelength of the exposure light. For example, for structures of 22nm width a laser of 193nm wavelength can be used. That is almost a factor eight larger than the details size! It is like writing thin lines using a pencil with a tip eight times wider than the lines. It is no wonder that the demise of Moore's Law has been postulated again and again, in view of the physical limits that optical lithography seems to be reaching. However, the death of optical lithography has been greatly exaggerated, as Mark Twain would say, and mathematical methods play an important role in the longevity and endurance of the law. In fact, physicists and engineers have found new techniques for exploiting the interference and wave properties of light in order to produce sharp image details. Now, before a chip is manufactured, extensive optical simulations of the complete imaging process are run on powerful computers. Moore's Law would have stopped being valid a long time ago, were it not for the numerical methods being used today. Thousands and thousands of CPU hours go into the design and optimization of the lithography masks. The whole process is now called "computer lithography".

2 Interference properties of light

The optical imaging difficulties stem from the wave properties of light. In Newton's time there was an intensive discussion about the nature of light. Newton thought that light consists of corpuscles which are so small that they do not make contact. They behaved otherwise as bodies possessing a certain small mass and even a form. Curiously, it was Einstein who in 1905 vindicated Newton, to a certain extent, when he explained the photoelectric effect as interaction of materials with photons behaving as particles.

But it was the wave theory of light which gained prominence due mostly to the work of the Dutch scientist Christiaan Huygens. He could explain phenomena such as reflection, diffraction and refraction of light in a unified way, making use of what we now call "Huygens principle". Huygens worked out this rule in 1690 in his "Traité de la lumière'", postulating that every point in a wave front can be conceived, and can be treated, computationally, as the source of a new secondary wave. The interference of the phases of the many point sources produces the observed expansion of the wave front. Fig. 3 shows an illustration from Huygens' book, where we can see points along a spherical wave acting as the source of new secondary spherical waves.

Light is electromagnetic radiation and each wave can interfere with another. Each wave has a phase (like in a sine curve) and two waves can interfere constructively or destructively. Two waves from a coherent source displaced by half a wavelength can "erase" each other. Adding up secondary waves corresponds to computing every possible interference. Mathematically, all this summing up of secondary waves is equivalent to computing the expected tra-

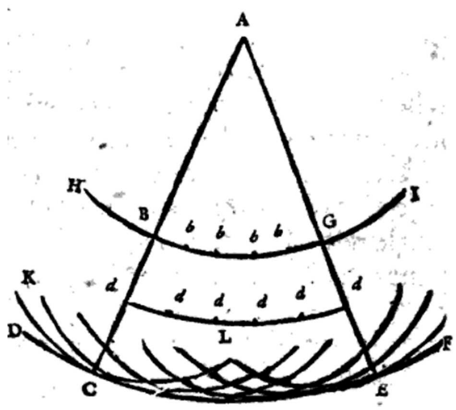

Figure 3: Huygens principle as illustrated in *Traité de la Lumière* (1690). Each point on a spherical wave is a source for secondary waves. Their interference produces the further progress of the wave front.

jectory of photons going in all possible directions, with changing phases along their trajectory.

Diffraction produced by small slits is especially important in photolithography. Light "bends" around obstacles and the smaller the slit, the larger the effect. Photolithographic masks with millions of details can be thought of as millions of small slits and the diffracted light has to be captured by lenses in order to reconstruct the image through controlled refraction. No image frequencies should get lost in the process.

3 The Rayleigh limit and the "Moore gap"

The layout of modern chips looks like a picture of a city, with millions of "streets" connecting millions of components. The chip components must be projected as tiny as possible on the wafer substrate. Smaller elements mean smaller connections and smaller details. The question then is whether optical lithography can still provide the sharp resolution needed (at some point the industry could shift to electron lithography and use electrons as imaging source, for example). Photolithography is the inverse problem to microscopy: in the latter we want to see the smallest details, in the first we want to recreate them by projection. In both cases expensive and accurate systems of lenses are needed. Fig. 4 shows an example of the tower of lenses needed in today's optical lithography. Projection errors, such as chromatic or spherical aberrations, are corrected by the stack of lenses, each of them contributing one small modification to the final light trajectory. Such lens systems are heavy and very expensive.

Two factors are relevant when considering the optical resolution of lenses: the size of the smallest details which can be seen through the system and the depth of focus of the projection (since the chips are planar and the details have to be focused precisely on the surface of the chip). In optics there is an expression for the resolution limit called the Rayleigh limit. This is expressed as

$$d = k\frac{\lambda}{\mathrm{NA}}$$

where λ is the wavelength of the exposure light, NA the so called numerical aperture, and k a constant related to the production process. For lithography, d is the width of the smallest structures that can be brought into focus. If we want to reduce d, we must increase NA or use a smaller wavelength. In the previous decades it was cheaper to move to progressively smaller wavelengths. Now, economics dictates that wavelength reductions are coupled to much higher costs, so that instead of moving to 157 nm exposure wavelength, for example, the industry is still working with the 193 nm alternative. Therefore, NA and k must be optimized. In both cases we have been stretching the limits of the technology for several years now.

Rayleigh's optical resolution limit arises from the interplay of the refracted light waves. Interference effects conspire to wash out the resolution of the

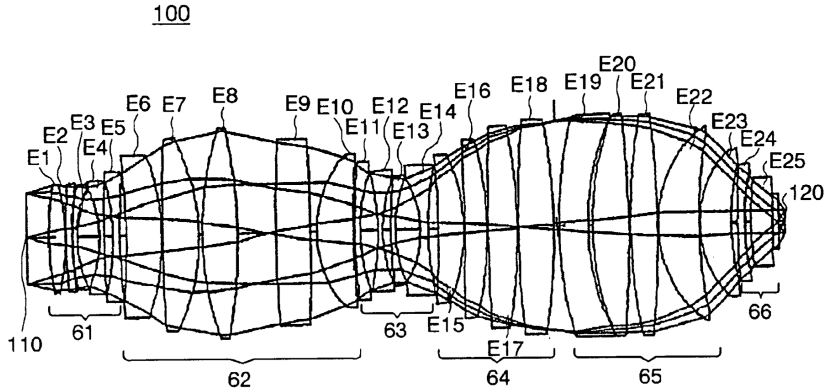

Figure 4: Diagram from a patent application for lithography lenses. The light traverses the system of lenses from left to right. The stack of lenses is positioned vertically in the lithography machine [5].

image when the details are of the same order of magnitude as the wavelength of the light being used. In the past, lithographic equipment had just progressed from one wavelength to the next. The industry moved from light from mercury lamps and 436 nm wavelength, to 365 nm (the *i*-line of mercury lamps), then further to 248 nm (KrF laser), and down to today's 193 nm wavelength (Argon-Fluoride). Also, now lasers, not just lamps, are being used, that is, coherent light sources, allowing a more precise control of the projected shapes. The next step would be moving to Extreme Ultraviolet lithography (EUV) with 13.5 nm wavelength, or still further to X-rays of smaller wavelength. However EUV light is absorbed in the air and the optics, so that the whole process would have to take place in vacuum and employ special lenses combined with mirrors. Glass, for example, is opaque to X-rays, so that no affordable projection systems exist for both kinds of electromagnetic radiation.

Fig. 5 is very interesting in this respect because it shows the gap between the growth trend of Moore's law and the integration effect of smaller wavelengths [9]. The vertical scale is logarithmic, so that Moore's law appears as a linear increase. The effects of improvements in wavelength have not kept pace with Moore's law, so that something different has to be made: instead of just reducing the laser wavelength, the production process must be modified, so that smaller structures can be imaged by means of the same exposure wavelength. Here is where improvements in the optics and tools require numerical methods. Moore's gap is mathematics' opportunity.

4 Immersion-lithography increases the numerical aperture

One production improvement which gave 193 nm lasers an edge over 157 nm lasers is immersion lithography, now almost universally used. Light is focused

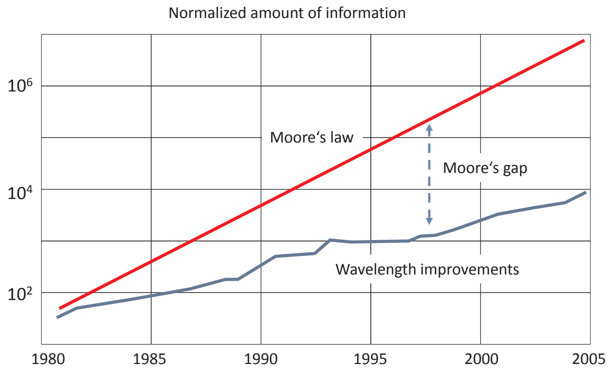

Figure 5: The "Moore gap". The growth in the number of components (proportional to the so-called amount of information) surpasses the growth of wavelength lithographic improvements alone [9]. The gap must be closed using novel techniques.

using lenses but some image frequencies are lost at the interface air-glass-air. Remember that the image produced by a spherical lens at the focal plane can be interpreted as a Fourier decomposition of the image. Lower image frequencies are collected near the optical axis, higher frequencies toward the periphery of the lenses. Some of the frequencies, lost due to the finite size of the lenses, can be kept in the system by moving from a glass-air interface to a glass-water interface. Water has almost the same refraction index as glass (1.44 against 1.5–1.6 for light of 193 nm wavelength). That limits the reflections on the lens surface (internal and external). Fig. 6 shows the trajectory of exposure light in both cases, with a glass-air or a glass-water interface at the wafer. The semiconductor is immersed in water; the water layer between the glass and silicon serves the purpose of capturing the high image frequencies so that the projection is sharper. Immersion lithography can be done with light of 193 nm wavelength but at 157 nm water becomes opaque and cannot be used as shown in Fig. 6. Obviously, introducing water between the lenses and the wafer leads to all kinds of manufacturing problems, but they were quickly sorted out so that the semiconductor industry moved to the new technology in just two years (between 2002 and 2003). Water is also not the last word: better liquids are being sought and could lead to further improvements of the optical process [14].

As Fig. 6 shows, immersion lithography improves mainly the so-called numerical aperture (NA) in Rayleigh's limit expression. The numerical aperture is directly proportional to the refraction index between the lenses and the wafer. NA is also directly proportional to the sine of the maximum projection angle (the angle between the vertical and the rightmost ray in Fig. 6). Since the projection angle cannot be larger than 90 degrees (whose sine is 1), further improvements of NA are limited by the geometrical constraints. This parame-

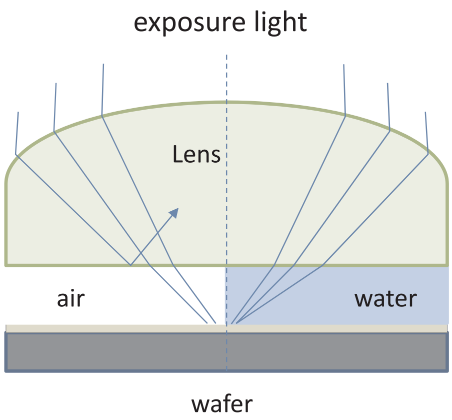

Figure 6: Immersion lithography is used on the right side, there is a glass-air interface on the left side. Undesired reflections at the glass-air interface (left) lead to poor resolution due to the loss of high image frequencies. Adapted from [13].

ter has already given most of what it can provide – alternative optimizations become indispensable.

5 Enter computer lithography

We are left with the constant k in the Rayleigh expression. Numerical methods and computers can contribute now. It is ironic that Moore's Law has led to the fast processors we have now on every desktop, but that the law itself is now dependent on these very same computers in order to continue being valid. Here we have a truly positive feedback system, where synergy between two seemingly separate fields can lead to exponential improvements in each one.

The idea of computer lithography is easy to explain using an example. Since light is diffracted by the structures on the projections masks for chips, what we can do is calculate in advance the effect of interference and modify the shape etched on the mask, so that we obtain the desired sharp image projection. That is, the mask is morphed in such a way that the diffraction, especially at corners, is taken into account from the beginning. Instead of trying to avoid interference, apply it, and make sure that constructive interference happens where you need it, while destructive interference erases undesired "shadows".

An embodiment of this idea is "optical proximity correction" (OPC). Connections with sharp corners can be obtained by adding "serifs" to the mask pattern. Fig. 7 shows an example. We want to obtain a structure shaped like an inverted L. The mask used has the wiggled form shown (in green) which looks like an L with some embellishments at the corners (the serifs). The imaging result is the somewhat rounded L, which is not perfect, but comes very near

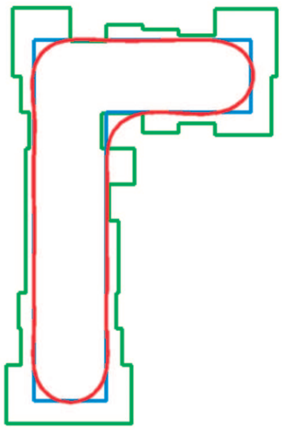

Figure 7: An example of *Optical Proximity Correction*. The green mask produces the red structure after photolithographic imaging (illustration from Wikipedia).

to the desired inverted L shape. The effect of the serifs is to produce the required interference. In order to produce such effects some rules of thumb or heuristics can be followed, but a really good result can only be obtained by simulating the outcome of Huygen's principle in advance.

6 Phase-shift masks and double patterning

It is also possible to manipulate directly the phase of the projected light. In order to do this, the mask has to be manufactured with materials that produce the phase-shift, or it can be manufactured with varying material thickness. A small step protuberance can be embedded in the mask with the only purpose of shifting the phase of the light going through each side of the step. Light waves coming from both step sides interfere then in controllable way. Fig. 8 shows an example. On the right, a mask with a small phase-shifting step has been exposed to a laser. Light going through the mask emerges with different phases on each side of the small step. The final illumination intensity produced by interference is such that total destructive interference can be obtained in the middle of the detail. On the left you can see what happens when no phase-shifting is used and the mask detail is smaller than the wavelength of the light used: the light bends around the obstacle and the detail almost disappears in the resulting low-contrast exposure: The wafer is being illuminated with almost the same intensity everywhere. On the right, on the contrary, a small detail of almost any width can be produced by adjusting the threshold of the photochemical reaction (that is, exposure to how many photons dissolves the material or not). The optical problem becomes manageable and the problem

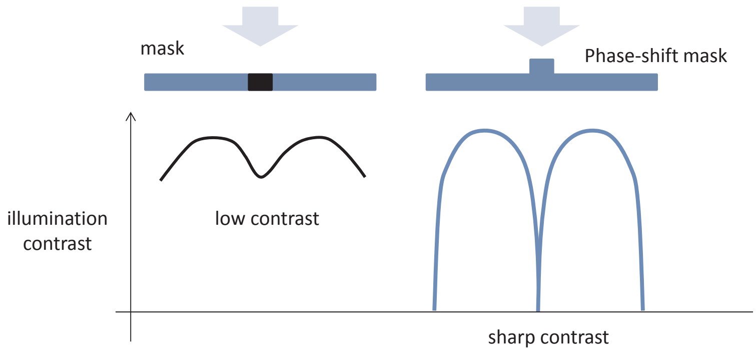

Figure 8: Without phase-shift, a mask produces the illumination shape shown on the left. The small detail in the middle is not projected with enough contrast on the wafer. A phase-shift mask (right side) uses a small step which shifts the phase of the incoming light. The interference effect is such that a sharp edge with high contrast is produced. Adjusting the illumination threshold a bar with any possible small width can thus be imaged on the wafer, theoretically.

is now to find the materials with the right photochemical properties for the obtained imaging contrast [3].

The design problem for the photolithography masks becomes now complicated. Phase-shifted masks represent the state of the art in the semiconductor industry. However, if phase-shifting is used everywhere in the mask, we are left with a combinatorial problem. The phase-shifting steps have to be distributed across the mask, using just two different mask levels. Special software must keep track of the areas where phase-shifting has occurred. Therefore, the layout of the mask must be planned very carefully. Usually, multiple masks are designed and the exposure steps are combined, leading to multiple exposures. Especially thin details can be produced by so-called double patterning [8], in which thin parallel connections are handled by exposing first the even numbered lines, and then the odd numbered ones (if you think of such parallel connections as having been numbered sequentially). The number of lithographic steps increases, and sometimes auxiliary structures become necessary, which have to be dissolved later (think of scaffolding during construction work). There are two main methods for integrating and dissolving the auxiliary structures, called respectively LELE und LFLE (for Lithography-Etch and Lithography-Freeze, and their combinations).

7 Structured light and quantum lithography

There is still another technique used to increase the captured high frequency components in the projected image. The idea is to use "structured light" when illuminating the photomask. This is an old proposal that was first applied to

Figure 9: Iris shapes for modern photolithography

microscopy, and which consists in illuminating not along the optical axis of the lenses but from the side. The same effect can be achieved if the light is first passed through an "iris", that is, an opening with a certain shape. The idea is to diffract the exposure light so that customized wavefronts reach the optics, that is, wavefronts capable of preserving more detail from the mask. Fig. 9 shows four examples of the type of irises used in photolithography for projecting light "structured" in such a way as to preserve more high-frequency details of the mask.

Quantum lithography is also a novel idea that would allow having access to smaller effective wavelengths without having to change the optical system. It consists of producing entangled photons so that they behave like a single quantum mechanical system. It is then possible to produce virtual particles with twice or thrice the energy of the original single photons. The virtual wavelength is reduced by a factor of two or three, as if we were using light of smaller wavelength. However, each particle can still be focused with the same kind of lenses as we have now, so that the problem of glass opacity at higher energies does not arise. The materials on the chip must be exposed in such a way that two or three photons are needed to produce the necessary photochemical reaction. It sounds like a good idea for the future, but low temperatures and very accurate equipment are needed, so that more research is still needed if quantum photolithography is ever to become reality.

8 Koomey's law and the power problem

A negative effect of Moore's law is the increase in heat released pro square millimeter in every chip. Microprocessors can become so hot, that enormous heat exchangers or water cooling becomes necessary. In 2009, Jonathan Koomey studied the historical development of the energy efficiency of computers and came to the conclusion that another power law is here at work. It is interesting that Koomey included in his analysis not just modern microprocessors but also very old machines, trying to find out how much energy has been used per computation in every historical period.

What Koomey found is that the number of operations per kWh follows the following rule: *The number of logical operations that one can obtain for a watt-hour doubles every 18 months* [6]. This rule of thumb is now called "Koomey's Law". If we would consume the same number of operations per second every year, the battery in new laptops would last twice as long as before. We know, however, that new software executes more operations per second so that the annual battery life gains are certainly lower. However, without Kommey's law many mobile applications would not be possible today.

Koomeys law, as first postulated, refers to the number of operations per second. That is not a good metric for comparing microprocessors since some processors can work with simpler instructions as others. Mobile processors, for example, are usually simpler than desktop computers. A better metric is to use the benchmarks produced by the *Standard Performance Evaluation Corporation* (SPEC), an organization whose mission is to provide a set of executable programs which represents real workloads for computer systems. The SPEC benchmarks compare execution times of realistic workloads and allow users to determine whether a processor is really faster than another.

In 2008, the SPEC organization released a new set of benchmarks for measuring the energy consumed by computer systems while executing typical workloads (graphic operations, data bank accesses, and so on). The SPEC Power Benchmarks are a basket of executable programs tested under three different conditions (10 %, 20 % and 100 % processor load). The idea is to test whether a processor which is working only at 10 % capacity is maybe consuming 50 % of the peak energy, for example. At the end, the SPEC Power benchmark shows how much processing the processor can deliver and at what energy cost (energy is measured by plugging the computer to appropriate measuring instruments).

There were 280 reports in the database of the SPEC organization in 2011. Fig. 10 shows the result of plotting this data. The vertical axis shows the SPEC-index (operations for kWh) for every processor and the horizontal axis the introduction year for the processors tested. The line represents the trend of all these measurements.

The graph shows that the operations per Watt have increased continually since 2007 (with a large spread). There are some very efficient processors, i.e., those near the 4500 SPEC power index, and some others which are certainly rather power hungry. The trend in the graph corresponds very closely to Koomey's law though. The SPEC power data shows a doubling of energetic efficiency every 18.8 months, very close to the expected doubling postulated by Koomey. In a certain sense, this law is a complement to Moore's law since not only more transistors per chip are important, but less energy for every logical computation makes many new applications possible.

9 The limits of photolithography

This short review of photolithographic "tricks of the trade" shows that the semiconductor industry has been extremely innovative every time it seems as if the physical limits of the production methods are about to be reached. Modern lithography must be described now using many adjectives: what we have is phase-shifted-double-patterning immersion lithography, based on resolution enhanced technologies (RET), such as Optical proximity correction and structured light. The whole process has to be extensively optimized and tested using computer simulations [12].

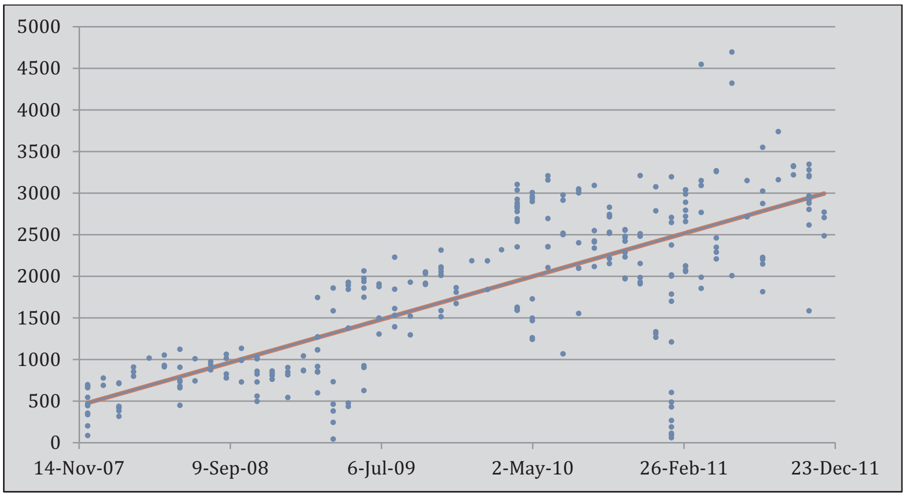

Figure 10: SPEC-Power results (December 2007 to December 2011). Each point corresponds to a processor and the date of the SPEC test. Some processors were tested after their introduction date, producing thus a significant spread of the data.

Photolithography will be further enhanced by using new materials whose photochemical properties can be tuned to the number of photons captured by the material. Low optical contrast can be enhanced using longer imaging periods, so as to be able to produce smaller and smaller structures. Some physicists are now of the opinion that there are no physical limits for optical lithography [1].

Moore's law could however hit a wall of a different nature: heat production in modern chips is already a problem, as Moore predicted in 1965 (notwithstanding Koomey's law), but more important than that is the fact that 22nm structures contain just around 220 atoms. If we reduce the number of atoms in transistors and connections, it could be that we start seeing uncontrollable non-linear effects. Fortunately, the physical limit seems to be still far away, having been reported recently that nanoconnectors with just four atoms still obey Ohm's law [2].

Therefore, the most important obstacle in the horizon seems to be of economic nature. EUV lithography has been postponed due to the enormous costs of the equipment. All new semiconductor factories are ultramodern buildings where hundreds or thousands o production steps must be planned and performed exactly. Intel's newest semiconductor fab is totally robotized and cost billions of dollars.

Physicists are already looking for alternatives, for a new age in which two-dimensional structures will not be enough. Moore's Law could get more oxygen – the production methods and materials used for semiconductors will then change radically within the next twenty years. But one thing is sure: numerical methods and simulation will be even more important in that future. Moore's Law has made numerical methods faster and more powerful, but numerical methods keep now Moore's law alive.

References

[1] S.R.J. Brueck, There are NO Fundamental Limits to Optical Nanolithography, in: A. Guenther (ed.), International Trends in Applied Optics, SPIE, 2002, 85–109.

[2] S. Das, Ohm's Law Survives at the Atomic Scale, IEEE Spectrum, January 2012.

[3] M. Fritze, B. Tyrrell, D. Astolfi, R. Lambert, D. Yost, A. Forte S. Cann, B. Wheeler, Subwavelength Optical Lithography with Phase-Shift Photomasks, Lincoln Laboratory Journal, V. 14, N. 2, 2003, 237–250.

[4] Tom R. Halfhill, The Mythology of Moore's Law, IEEE SSCS Newsletter, Sept. 2006, 21–25.

[5] R. Hudyma, W. Ulrich, H-J. Rostalski, Compact 1½-waist system for sub 100 nm ArF lithography, United States Patent 6906866, Carl Zeiss SMT AG, 2005.

[6] Jonathan G. Koomey, Stephen Berard, Marla Sanchez, Henry Wong, Implications of Historical Trends in the Electrical Efficiency of Computing, *Annals of the History of Computing*, July–Sept. 2011, 46–54.

[7] X. Ma, G. Arce, Computational Lithography, John Wiley & Sons, 6th edition, August 2010.

[8] C. Mack, Seeing Double, IEEE Spectrum, September 2008.

[9] T. Matsuyama, Y. Ohmura, D. Williamson, The Lithographic Lens: its history and evolution, Optical Microlithography XIX, Donis G. Flagello (ed.), Proc. of SPIE, V. 6154, 2006.

[10] Gordon E. Moore, Cramming more components onto integrated circuits, Electronics, V. 38, N. 8, April 19, 1965, 114 ff.

[11] Gordon E. Moore, Progress In Digital Integrated Electronics, Technical Digest, IEEE International Electron Devices Meeting, 1975, 11–13.

[12] Thomas Lee, Leap for Microchips, Scientific American, January 2002, 52–59

[13] G. Stix, Shrinking Circuits with Water, Scientific American, July 2005, 64–67.

[14] M. Yang, S. Kaplan, R. French, J. Burnett, Index of refraction of high-index lithographic immersion fluids and its variability, J. Micro/Nanolith. MEMS MOEMS 8(2), 023005, Apr–June 2009.

Raúl Rojas
Dept. of Mathematics
and Computer Science
Freie Universität Berlin
Arnimallee 7
14195 Berlin
Germany
raul.rojas@fu-berlin.de