Available online at www.sciencedirect.com

ScienceDirect

C. R. Physique 7 (2006) 887–895

Comptes Rendus Physique

http://france.elsevier.com/direct/COMREN/

Ultimate lithography/Lithographie ultime

From 120 to 32 nm CMOS technology: development of OPC and RET to rescue optical lithography

Yorick Trouiller

CEA/LETI, Minatec, 17, rue des Martyrs, 38054 Grenoble cedex 09, France

Available online 21 November 2006

Abstract

Starting from the 120 nm CMOS technology node down to the 32 nm node, we have entered into a new lithographic regime. The wavelength has not changed (only 193 nm), and we move closer and closer to the theoretical optical resolution limit. Therefore, Resolution Enhancement Techniques (RET) have been developed in order to print all shapes properly and close the resolution gap. The primary RET developed are off-axis illumination, sub-resolution assist features and a phase shift mask. Moreover, working closer to the resolution limit implies bigger image distortion between the mask and the silicon. For this purpose OPC (Optical Proximity Correction) has been widely used by making mask pre-compensation of all non linear effects, optical diffraction and interference effects, resist and etch. RET and OPC are also fundamentally linked. RET such as off-axis illumination generates more distortion, and therefore justifies the need of more aggressive OPC, and RET techniques like Alt PSM and sub-resolution assist features are generated through the OPC infrastructure. From its first industrial utilization for 120 nm node to 32 nm prospectively, many evolutions have been seen for OPC. These include the generalisation to all lithographic layers, moving to pixel based simulation, usage of full chip simulation and verification, the incorporation of process window effects like Energy Latitude or Depth of Focus into the OPC algorithm, and inverse lithography approach. For RET, we have seen huge differentiation depending on the type of application, such as logic or memory. In conclusion, we need to consider design as a third party that is playing a key role in this RET–OPC synergy. To use more aggressive RET and reduce the cycle time of OPC recipe development, more regular designs are considered as a key enabler for the future: they will allow logic makers to consider RET options that are pushed as far as those used by memory makers. ***To cite this article: Y. Trouiller, C. R. Physique 7 (2006).***

Résumé

La technologie CMOS de 120 à 32 nm : le développement des OPC et RET au secours de la lithographie optique. A partir de la génération CMOS 120 nm jusqu'à la prochaine plateforme 32 nm, nous sommes entrés dans un nouveau régime lithographique : plus de changement de longueur d'onde (fixée à 193 nm), et de plus en plus près de la résolution théorique ultime. Dans ce contexte, des techniques visant à améliorer la résolution dénommées RET (Resolution Enhancement Techniques) ont été développées pour imprimer les formes correctement et réduire l'écart dans la résolution accessible. Le premières RET développées ont été des illuminations hors axe, des motifs diffractants sous résolus et des masques à décalage de phase. En outre, travailler proche de la résolution limite implique une plus grande distorsion entre l'image voulue du masque et l'image réelle sur la plaquette de silicium. A cet effet, des corrections d'effets de proximité appelées OPC (Optical Proximity Corrections) ont été utilisées largement pour précompenser sur le masque tous les effets non-linéaires, la diffraction optique et les effets d'interférence, la résine et la gravure. Ces technologies OPC et RET sont fondamentalement liées : les techniques RET comme l'illumination hors axe génèrent plus de distorsion et justifient donc le besoin d'OPC plus agressive, et en même temps bon nombres de techniques RET (comme les PSM

E-mail address: yorick.trouiller@st.com (Y. Trouiller).

1631-0705/$ – see front matter © 2006 Published by Elsevier Masson SAS on behalf of Académie des sciences.
doi:10.1016/j.crhy.2006.10.001

alternés et les motifs diffractants sous résolus) sont créés grâce aux outils informatiques mis en place en OPC. Depuis sa première utilisation industrielle pour le nœud 120 nm jusqu'aux perspectives en 32 nm l'OPC a vu bien des évolutions. Ceci inclut la généralisation à tous les niveaux lithographiques, le passage à la simulation au niveau du pixel, l'usage de la simulation et de la vérification de toute la puce, l'incorporation dans l'algorithme d'OPC des effets de fenêtre de procédé comme la latitude d'exposition ou la profondeur de champ, ainsi que l'approche par lithographie inverse. Pour la RET on a vu une forte différentiation en fonction du type d'application comme la logique ou les mémoires. En conclusion nous devons considérer la conception du circuit comme une tierce partie qui joue un rôle clé dans cette synergie RET–OPC. Pour utiliser une RET plus agressive et réduire le temps de cycle de développement des recettes d'OPC, des motifs plus réguliers sont considérés comme clés pour le futur : ils permettront aux fabricants de circuits logiques de considérer des options de RET aussi poussées que celles utilisées par les fabricants de mémoires.
Pour citer cet article : Y. Trouiller, C. R. Physique 7 (2006).

Keywords: OPC; RET; Optical lithography

Mots-clés : OPC ; RET ; Lithographie optique

1. Low k_1 imaging: the reason for OPC and RET

Before printing dimensions below one quarter micron, lithographers have mainly used the path of wavelength reduction to assure the scaling factor required by Moore's law used in semi-conductor industry. Moore's law corresponds to the shrink of all dimensions by 30% every 2 years. This wavelength scale allows for a gain in resolution without a large penalty in process window margin. This trend has been followed from the 1980s (g-line of mercury lamp—435 nm wavelength) to the beginning of 2000 (ArF laser 193 nm wavelength) (Fig. 1).

However, since the beginning of the twenty-first century, no wavelength scale has occurred because of limitations in classical diffractive lithography. These limitations include increasing absorption and birefringence in lens materials, insufficient laser power for smaller wavelengths, and the difficulty to find a resist with the proper contrast. The classical Rayleigh equation is always useful to understand optical limitations:

$$R = k_1 * \lambda/\mathrm{NA} \tag{1}$$

where R is the resolution (half-Pitch criteria), k_1 is the merit factor of the lithography, λ the wavelength and NA the Numerical Aperture. The Depth Of Focus (DOF) can be modeled with the altered Rayleigh equation:

$$\mathrm{DOF} = k_1 * \lambda/\mathrm{NA}^2 \tag{2}$$

As we see, NA can also improve the resolution, but only at the expense of Depth Of Focus (square dependency). Therefore lithographers have increased NA slowly, and tried to decrease the k_1 factor. The k_1 factor was 0.5 for 120 nm

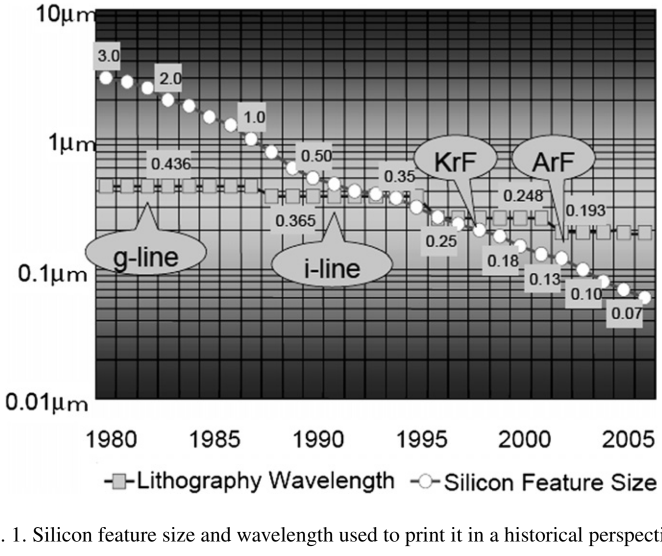


Fig. 1. Silicon feature size and wavelength used to print it in a historical perspective.

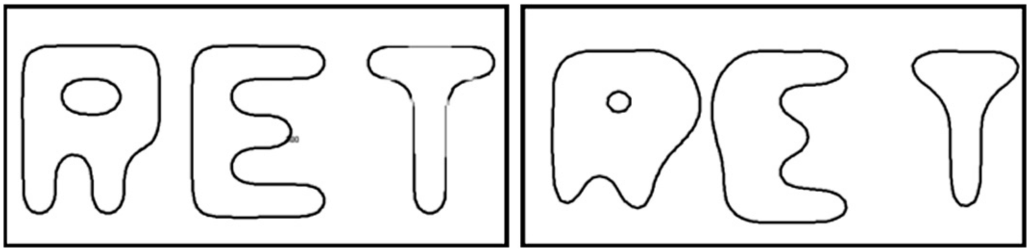


Fig. 2. Pattern printed with $k_1 = 0.5$ (left) and $k_1 = 0.35$ (right).

technology and will go below 0.32 for 32 nm node. Unfortunately decreasing k_1 factor implies higher distortions, as seen in Fig. 2.

To overcome this challenging limitation, broad usage of OPC and RET is required:

- OPC to pre-compensate on the mask for the anticipated distortion;
- RET to recover the resolution.

2. OPC

OPC started to emerge in the 1990s as a valid technique for sub micron lithography. The goal was to use Rule Based Correction to correct the most critical structures of the layout. A classical example is to use hammerheads to reduce line-end pull-back. However, at geometries below 150 nm, it appears that the rule based approach fails because the required rules are too complex to generate for a logic chip while taking into account all correctable cases. Therefore the *model based* approach has started to emerge. As seen in Fig. 2, the basic principles are:

- simulate the drawn shape on silicon;
- break the design polygons into moveable fragments;
- iteratively minimize the difference between the desired shape and the wafer image.

2.1. Optical, resist and etch simulations

Optical simulation has been developed through the Hopkins theory, which properly describes the optical phenomena in lithography (the Abbe approach is too costly and not required). To speed up the simulation, kernel decomposition with SOCs (Sum Of Coherent system) has been developed successfully [1]. In this approach, the aerial image is the sum of canonical kernels convoluted by the mask (Fig. 3). Depending on lithographic complexity and accuracy needs, 5 to 50 canonical kernels are used for the approximation of the aerial image.

Subsequent development of optical simulators has been dependent on the required NA regime:

- below NA of 0.75, the scalar approximation at wafer level and thin film mask approach have been used;
- for NA greater than 0.75, the vector approach at the wafer level is being utilized, and interference effects in the resist stack are comprehended;
- for NA greater than 1, made possible with immersion lithography (45 nm technology), light polarization effects and 3D mask effects need to be taken into account.

Resist simulation has been approached through a more empirical way than optical theory because the physics and chemistry of photo resist reaction are less understood. In this spirit two types of generic models are used (together or independently):

- *Gaussian convolution* of the optical aerial image to mimic all universal Fickian phenomena. Resist diffusion is dominated by acid diffusion with diffusion length below 50 nm and quencher (additive base) with a diffusion length around 300 nm. Some more advanced models are also using truncation effects.

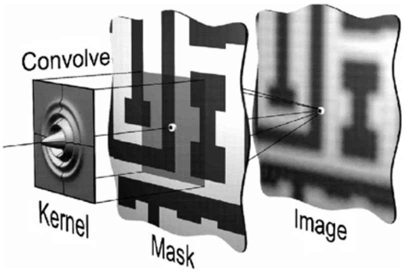


Fig. 3. Kernel decomposition methodology used for fast simulation in OPC.

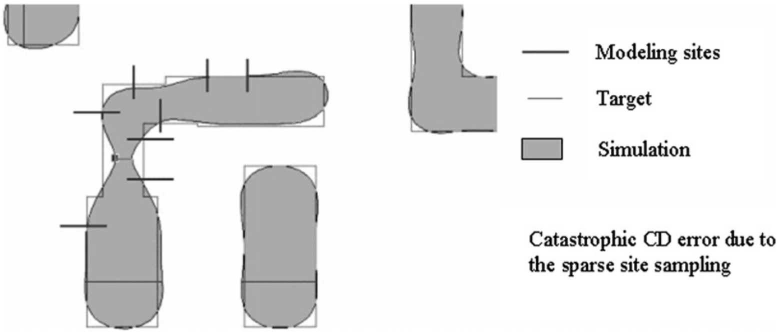


Fig. 4. An example of catastrophic OPC error in 65 nm technology when simulation sites are not properly placed.

- *Variable threshold* is based on the modulation of the threshold that is used on the aerial image curve to determine the dose to print in the resist. Without strong resist effects and with current chemically amplified resists that have sharp profiles and good contrast, a threshold around 0.3 is required to convert the optical intensity curve into dimensional measurements of the resist after development. The variable threshold principle is that we can replace this threshold by an effective threshold which is a function of the environment. By doing this as a function of $I_{\min}$, $I_{\max}$ intensity of the aerial curve, or 2D curvature terms and cross terms, we are able to mimic resist effects.

Etch effects are slightly less influential than resist effects, and therefore etch is often modeled by simple rule based techniques. At the first order, the etch bias for each polygon edge can be considered as a function of line and space measurements. Recently for sub 100 nm generation, people have moved to model based etch simulation because of tighter accuracy requirements. The Gaussian convolution principle is re-used with diffusion length around 1 μm, but an angular dependency needs to be taken into account as well as the mask open area.

To fit these complex models with real data on silicon wafers Scanning Electron Microscope measurements of test patterns are used. Commonly, 50 to 5000 patterns are defined for each level (poly, contact, metals) and each technology to be representative of all the possible design layouts given a discrete design rule Manual.

Finally the critical parameter, simulation space, needs to be taken into account: in theory, optics, resist or etch phenomena can have long range interactions (more than 500 μm). However, for practical reasons (reasonable simulation time), simulation is limited to a region smaller than 5 μm surrounding the simulation point. For wavelengths of 193 nm or 248 nm, an optical diameter from 1 to 4 μm is sufficient to describe interference and diffraction phenomena based on optical theory. Straylight (also called flare) is neglected. In the same way resist and etch effects are simulated within a 5 μm window, neglecting long range loading effects.

2.2. OPC engine description

The first operation of the OPC software is to dissect the initial layout into fragments, with a minimum of three fragments per edge to assure the proper convergence. This dissection operation is usually rule based, with the basic concern to add more fragment where the local curvature is high (for example close to the corners). During this process it is necessary to find a good balance between the OPC accuracy and the total number of fragments that will impact the mask writing time for the current e-beam mask writer tool. In the latest generation, sophisticated dissections algorithms have been deployed because rule based dissection is not able to find the correct fragment for complex RET. These algorithms first analyze the image produced with basic rule based dissection and then place the fragment at the best location to maximize the efficiency of the correction.

The second operation of the OPC software is the simulation site placement. Due to simulation time constraint, it is difficult to simulate everywhere. Therefore simulation is usually done only at locations with one cut line per fragment. The choice of the simulation site placement is very critical, as seen on Fig. 4.

The third operation is the OPC convergence step. At the beginning of this operation the image is simulated at each site, and the Edge Placement Error (EPE) corresponding to the difference between the simulated edge and the targeted edge is calculated. All fragments of the layout are moved by the –EPE amount (multiplied by an attenuator factor to assure the proper convergence). This is repeated iteratively to reach the final result. Usually between 4 and 20 iterations are required to reach OPC errors smaller than the OPC grid size at each simulation site. It is also important to note that fragment movements are enforced by Mask Rule Constraints (MRC) of minimum CD and space in order to allow accurate mask inspection. For recent technologies (90 nm and beyond), it is also very difficult to assure good OPC convergence without any blocked edges, This is a special concern when more than 3 edges are in conflict for a minimum space or width. For this convergence operation, sophisticated algorithms are also used to improve the OPC results and reduce the number of convergence steps:

- usage of PID (Proportional Integral Derivative) controller developed in the frame of the classical control theory [2];
- for each edge movement, take into account all the neighboring influences: rather than have a unique equation for each edge, a full matrix needs to be inverted to find the movement factor of each edge.

2.3. Hierarchical context and flattening operation

OPC software needs to handle the initial hierarchized description context of the database. For example a memory array is described as one cell mirrored and repeated. This initial hierarchy is often broken during each OPC step because the simulation window is bigger than the cell itself, and therefore some parts need to be flattened. Another way to simplify the problem is to erase the hierarchy description and to put all data at the same level. However, in that case, the database size increases dramatically as the OPC needs to simulate all large memory arrays. It is more intelligent to simulate a small part of the array and to take into account the inherent symmetry of the features. Therefore, OPC software uses very complex algorithms to manage the hierarchy at each OPC operation. In most cases, more than 10 times faster run-time can be achieved with this method, but at the expense of unpredictable cycle time and higher risk of errors when complex hierarchy need to be handle (like multi-overlapped cells).

2.4. OPC verification

As seen in the previous paragraphs, OPC is a very risky operation, because it has to be applied to billions of polygons for each level of each database and more than 30 different lithographic steps for 90 or 65 nm generations. For 120 nm node, the first generation to use model based OPC in industrial mode, this OPC operation has been controlled mainly by geometrical and topological checks. These checks include minimum line width and spacing authorized during the OPC, loss of polygons compared to the initial database, and preservation of angles 0/90/45 degrees. Nevertheless, it has been shown that databases can pass these integrity checks but fail on silicon with drastic yield loss. Therefore *full layout simulation check* after OPC has been developed to control risk of necking, bridging, and overlay between two layers like vias and metals. These types of checks have been deployed for 90 nm technology, and improved for 65 nm generation to simulate through the entire dose-focus lithographic process window [3].

2.5. Evolutionary OPC approach: pixel based simulation, PW aware OPC

After less than 10 years of industrial utilization, OPC can not yet be said to be a stable and mature technology. As OPC becomes one of the biggest challenges of the semiconductor industry, many researchers are working on some evolutionary and revolutionary scenarios to simplify and speed-up the OPC development processes at each generation, and improve the quality required for ever decreasing k_1 lithography. To date two kind of evolutionary techniques are emerging:

- Replacing the traditional site-based simulation by a *pixel-based approach* that is able to simulate over an entire layout. These types of algorithms are based on the Fast-Fourier Transform, and require the utilization of large computer farms or dedicated customized and accelerated boards. This approach provides the capability to find the worst location in terms of bridging and pinching;
- The primary focus of OPC is to reproduce the intended design layout as closely as it is possible. However, in low k_1 imaging, this approach is failing because it does not take into account the optimization function of the final pattern generated in terms of a robust lithographic process window (dose, focus). To overcome these issues,

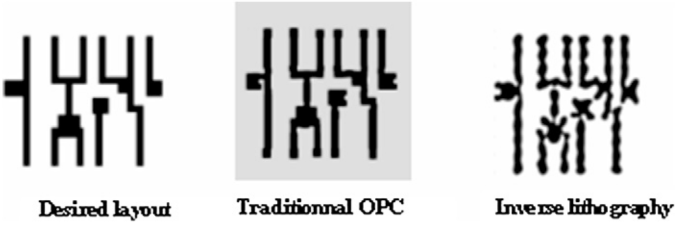


Fig. 5. Comparison of mask after OPC produced by traditional OPC and inverse lithography.

algorithms that include imaging from the corners of the process window in the OPC cost function, have been developed. This is called *process-window-aware OPC*, and it allows for slight adjustments to the intended design layout when it leads to improvements in robustness of the silicon pattern.

2.6. Revolutionary OPC approach: inverse lithography

Of course, more revolutionary flows have been suggested and tried by some successful (and unsuccessful) spin-off companies. Standard OPC approaches require a set of many non-trivial parameters (typically more than 500 variables, and more than 3000 customized coding lines) that often need more than 5 real tests on mask before convergence to the proper set of parameters is reached for each new process. Moreover, some new structures discovered after the R&D development phase are missing from the OPC script. These structures can create new failures that require adjustment of the OPC script during production phase. The only way to simplify the problem drastically is to completely re-think the OPC as an inversion problem rather than an optimization problem with multiple and sequential constraints [4]. The starting point for this approach is the desired shape and, as most of optical or basic resist diffusion terms are invariable, it is possible to find the ideal mask. This method has proven to lead, by construction, to a more robust pattern on the wafer and better pattern fidelity, but at the expense of mask complexity (Fig. 5). Actual development shows that this method has a real competitive advantage for extreme low k_1 imaging (k_1 below 0.33).

3. RET

Resolution Enhancement Techniques (RET) have been seen in the 1980s as a way to extend the life of classical optical lithography. From this perspective we can define two kinds of RET:

- RET improving resolution, like off axis illumination, phase shift mask, polarization;
- RET improving the process window robustness (mainly Depth Of Focus), like scattering assist features or focus drilling techniques, and water immersion.

We can also distinguish those techniques that are cumulative (like attenuated Phase Shift Mask + off axis illumination + assist feature) from those that are non-cumulative. The cumulative techniques can be used for an evolutionary platform, while the non-cumulative techniques need disruptive lithographic processes like Alternated Phase Shift Mask. In this article, we will concentrate on those that have a strong inter-dependency with OPC techniques, i.e. off axis illumination, assist features, and alternating Phase Shift Mask. The others are also well described in the literature.

3.1. RET evolutionary flow: off axis illumination

Off axis illumination is a simple way to improve the lithographic resolution. To come back to the Fourier theory, a lithographic system can be described as a simple frequency filter that allows only orders of diffraction with frequencies smaller than NA/λ to pass. High contrast resist requires at least 1 order of frequency (excluding zero order) to have a modulation in order to print a fine resist line and space pattern. As seen on Fig. 6, with the use of an oblique source ray arriving on the mask, we can divide the resolution by two by imaging only with the 0 and −1 orders, and sacrificing +1 order. This leads to a theoretical minimum k_1 of 0.25.

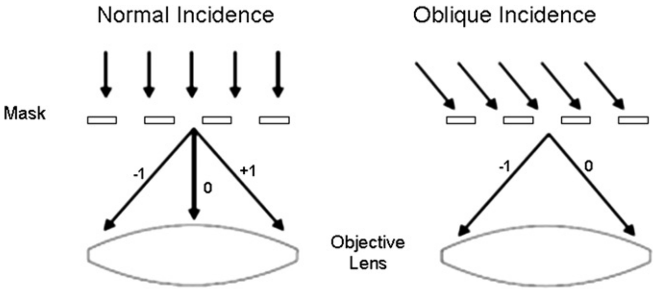
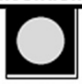
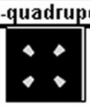
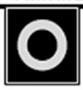
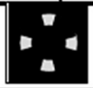
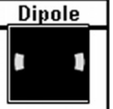


Fig. 6. Normal incidence and oblique incidence principle.

Table 1
Ratio of first order in the entrance pupil and zero order as a function of k_1

	Conventionnal	45°-quadrupole	Annular	Cross-pole	Dipole
					
k1=0.4	20%	70%	55%	50%	100%
k1=0.33	10%	15%	35%	45%	95%
k1=0.3	5%	10%	25%	45%	65%

Of course, actual illumination sources should not be described with only an angle but with a solid angle. With this description conventional illumination is a cone of light arriving on the mask, and annular illumination is a ring. In the same way we can describe other illumination sources that have angular dependencies

- on axis quadrupole or cross-pole;
- 45 tilted quadrupole;
- Dipole.

Table 1 summarizes the efficiency of first order captured. This parameter is directly linked to the final contrast and achievable resolution on the wafer. It is evident that across the illumination options, from conventional illumination, which is the closest to normal incidence, to dipole illumination which is the closest to oblique incidence, that each is effective at a particular k_1 factor.

It is also important to understand that these extreme oblique illuminations cause higher distortion to be induced on the wafer and therefore more aggressive OPC techniques are required (Fig. 7).

As a consequence, lithographers have carefully adjusted illumination conditions depending on the k_1 factor required. In parallel with this activity, they have increased OPC aggressiveness as illumination becomes more extremely off-axis.

3.2. RET evolutionary flow: Sub Resolution Assist Feature (SRAF)

The primary focus off-axis illumination was to increase the resolution of dense pitches. This is of course a key objective, but we should not forget that designers can not draw chips with an only one dimensional dense feature. It is clear that if a gain in performance can be achieved with a given RET on the densest pitch, no gain will result on some bi-dimensional structures or isolated line (and often degradation is incurred). To avoid this critical issue, lithographers have invented sub-resolution assist features that are placed throughout patterns to create a dense-like environment. SRAF are able to mimic dense patterns without printing on the wafer when the size is properly chosen. These patterns are usually generated during the OPC step by very complex rule based scripts that are able to take into account all possible 2D configurations and mask rule constraints. An example of catastrophic error is seen on Fig. 8 for a non optimum SRAF placement.

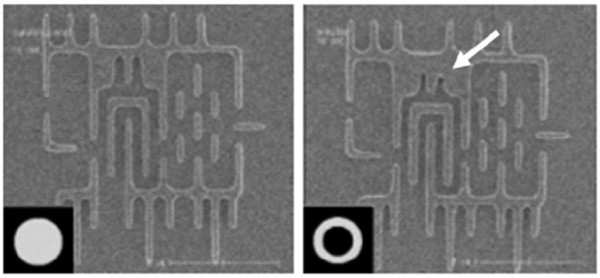


Fig. 7. 120 node gate printed with conventional illumination (left) and annular illumination (right).

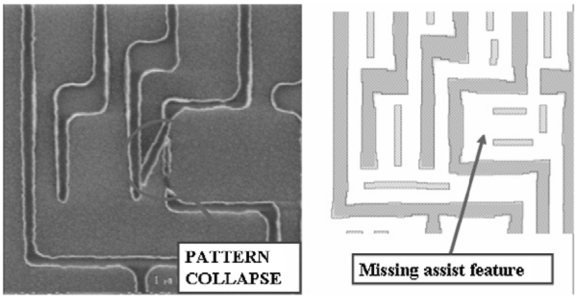


Fig. 8. Example of pattern collapse issue in defocus condition due to non optimal SRAF placement.

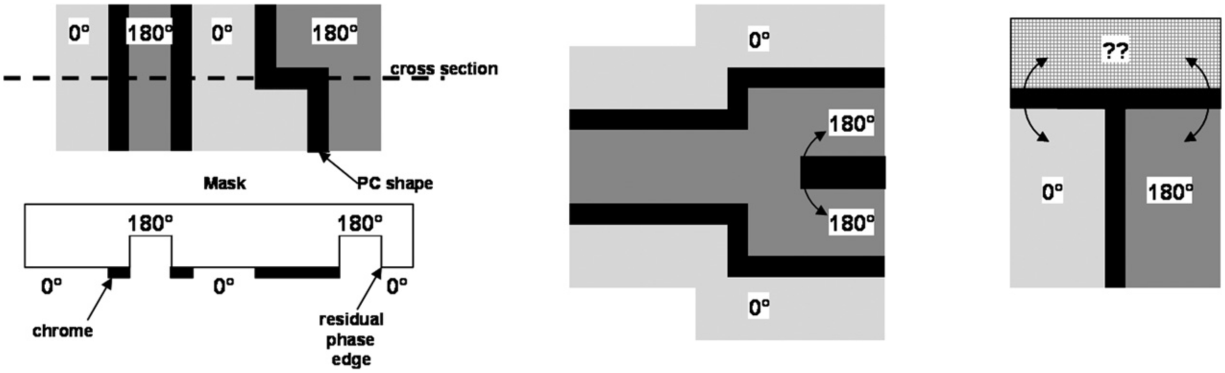


Fig. 9. Alternated PSM mask layout (left) and phase conflict examples (right).

In summary it is important to mention that the use of SRAF has been adopted in production environments as a standard RET when OPC software is available to generate this type of pattern for complex designs.

3.3. RET revolutionary flow: alternated Phase Shift Mask

Alternated phase shift mask has often been considered as the grail of lithographers because it is the RET that leads, in theory, to the best resolution with the best process control. Invented in the 1980s by Mark D. Levenson, it has caused the generation of thousands of publications, but unfortunately it has still a very rare industrial application. The reasons why are numerous, but all related to technical complexity [5]. The idea behind alternated PSM is to change phase in the transparent area at one side of the opaque pattern. One of the biggest challenges is that phase conflict can appear depending on the layout (Fig. 9). This can not be solved without having very complex algorithm (usually developed by OPC vendors) and phase conflict compliance check done by the designers.

4. RET, OPC and design

OPC and RET are, in their nature, technologies that are design dependent. In fact, the RET strategy should be adapted to the type of circuit desired. We can distinguish three different families, in an ascending order of design regularity. They are, System on Chip (SOC), microprocessors and memories (DRAM and Flash).

Together with RET evolution, there is a global effort to promote a better communication path with designers. This is true for RET and OPC aspects, but more globally in the deployment of Design For Manufacturing (DFM).

Table 2
RET and OPC differences for memories, microprocessors and SOCs

	Major economical constraint	Design characteristics	OPC constraint	k_1 constraint	Preferred RET
Memories	• wafer cost	• full regular in the array • not regular in the periphery but with relaxed rules		• very aggressive	• strong off-axis • polarization
Microprocessors		• regular only for gate level			• Alternated PSM
SOCs	• wafer cost • mask cost	• not regular	• OPC run time • OPC robust for every new design		• weak off axis • SRAF • Attenuated PSM

4.1. *Different strategies for logic SOC makers, microprocessors and memories*

To explain the different strategies associated with the three product families, it is important to recall the economic constraints, layout characteristics, and OPC constraints of each one. Table 2 offers an overview of the primary RET choices related to these specificities.

As seen in Table 2, memories are able to take advantage of strong RET. This is because of layout simplicities and relatively small number of products. To the other hand, SOCs have more constraints on the OPC and the mask cost, and therefore use more conservative RET solutions. Microprocessors are in an intermediate position.

4.2. *Design For Manufacturing (DFM)*

DFM has appeared to be a general trend of the semiconductor industry since 2000s [6]. It appears that OPC and RET have been seen as a catalyst for many DFM actions:

– the intensive usage of RET and OPC generates a lot of layout marginalities, and emphasize the need for design anticipation;
– all the virtual silicon full chip simulation tools created for OPC verification can be re-used at the design level to emulate fab behaviour and correct patterning hot spots;
– the development of more regular layouts in the frame of DFM for microprocessors or SOCs can allow the use of more aggressive RET, pushing the required k_1 factor further by the current wavelength and NA limitations.

5. Conclusions

In conclusion, OPC and RET have became key techniques in the semiconductor industry for sub 100 nm technology, because of the technical limitation of classical lithography driven by the wavelength and Numerical Aperture scaling. With this introduction we have entered into a new paradigm where the lithographic performances are fully design dependent. Therefore a new infrastructure within the DFM effort is required to anticipate at design level the OPC-RET impact with virtual patterning simulation tools.

References

[1] N. Cobb, Fast optical and process proximity correction algorithms for integrated circuit manufacturing, PhD thesis 1998, University of California, Berkeley.
[2] B. Painter, et al., Classical control theory applied to OPC correction segment convergence, Proc. SPIE 5377 (2004) 1198–1206.
[3] A. Borjon, et al., High accuracy 65 nm OPC verification: Full process window model vs. critical failure ORC, Proc. SPIE 5754 (2005) 1190–1201.
[4] D.S. Adams, et al., Fast inverse lithography technology, Proc. SPIE 6154 (2006) 61541J.
[5] A. Tritchkov, et al., Lithography enabling for the 65 nm node gate layer patterning with alternated PSM, Proc. SPIE 5754 (2005) 215–225.
[6] L. Liebmann, et al., Reducing DfM to practice: the lithography manufacturability assessor, Proc. SPIE 6156 (2006) 61560K.