Review Article

Peter De Bisschop*

How to make lithography patterns print: the role of OPC and pattern layout

Abstract: This paper will review some of the methods that have been devised to bring lithography-generated patterns as close to the desired target patterns as possible, while making them also robust against inevitable deviations from the ideal conditions during the printing process. Optical proximity correction (OPC) is the first step in this process. Various ways have been developed for efficient creation of accurate process window aware OPC models. Also, the use of the actual OPC step, to transform the target patterns into actual lithography mask patterns has seen significant progress. A computational verification step then checks whether the predicted pattern shapes meet the quality requirements and identifies any residual failures or weak patterns ('hotspots'). Once the mask is available, a second verification step, now looking at patterns on printed wafers, is performed to make sure that all critical patterns print to within the requested tolerances. Each of the steps in this flow can – and usually does – lead to corrective iterations to one of the previous steps. As the task of ensuring sufficient process margin is gradually becoming more difficult, with the ever decreasing pattern sizes, constraints are being increasingly defined on the type of patterns that can be allowed in the target layout itself (design restrictions), leading to a tendency toward more regular designs, an evolution that needs to be facilitated by the patterning technology and materials used. So the problem of ensuring good printability now also involves both layout and technology, and we will look into this aspect of the optimization problem as well.

Keywords: DTCO; geometrical layout restrictions; OPC; optical lithography; unidirectional layouts.

DOI 10.1515/aot-2015-0023
Received March 15, 2015; accepted April 14, 2015; previously published online May 20, 2015

***Corresponding author: Peter De Bisschop,** Advanced Lithography Department, IMEC, Kapeldreef 75, B-3001 Leuven, Belgium, e-mail: peter.debisschop@imec.be

www.degruyter.com/aot
© 2015 THOSS Media and De Gruyter

1 Introduction

Semiconductor devices are built as stacks of patterned layers, where the lithography step is the one creating the basic patterning of each of the layers. This is done by transferring patterns from a mask to a photosensitive layer (resist) that has been spun on the wafer, using an optical projection system. The resist pattern is then subsequently transferred into the underlying layers by an etch or an implantation step [1]. Repeating this cycle, layer by layer, creates a three-dimensional stack that eventually becomes the semiconductor device.

Whether or how well this device will work, depends (among others) on the quality of the patterns in each of the process layers, i.e., how close they are to the intended or target pattern, even when the printing condition deviates from the ideal conditions, e.g., because of (inevitable) errors in the focus or dose of the scanner or errors in the mask pattern. Therefore, methods and tools of increasing sophistication have been developed over the past decades to optimize both pattern fidelity and robustness, as much as possible.

Among these, optical proximity correction (OPC) plays a key role. The need for OPC is caused by the fact that the projection lens in the lithographic exposure tool (the 'scanner') – due to its finite size (NA) – cannot recombine all the diffracted light into the reconstructed image at the wafer side [1, 2]. This implies that this image will not be identical to the pattern on the mask. This image deviation becomes more important with decreasing target-pattern size, but it is also highly pattern dependent. OPC uses a combined optics-resist model (and if desired also etch model) to correct for this proximity effect, by converting the pattern on the mask from the target pattern into a different one, such that what is printed on the wafer will, in fact, approach the target pattern as closely as possible [3, 4].

Figure 1A illustrates the basic OPC concept for a small example clip of a 20-nm logic node (N20) Metal1 layer, that is already split into two separate layers or 'colors' (double patterning). The etch model converts the design

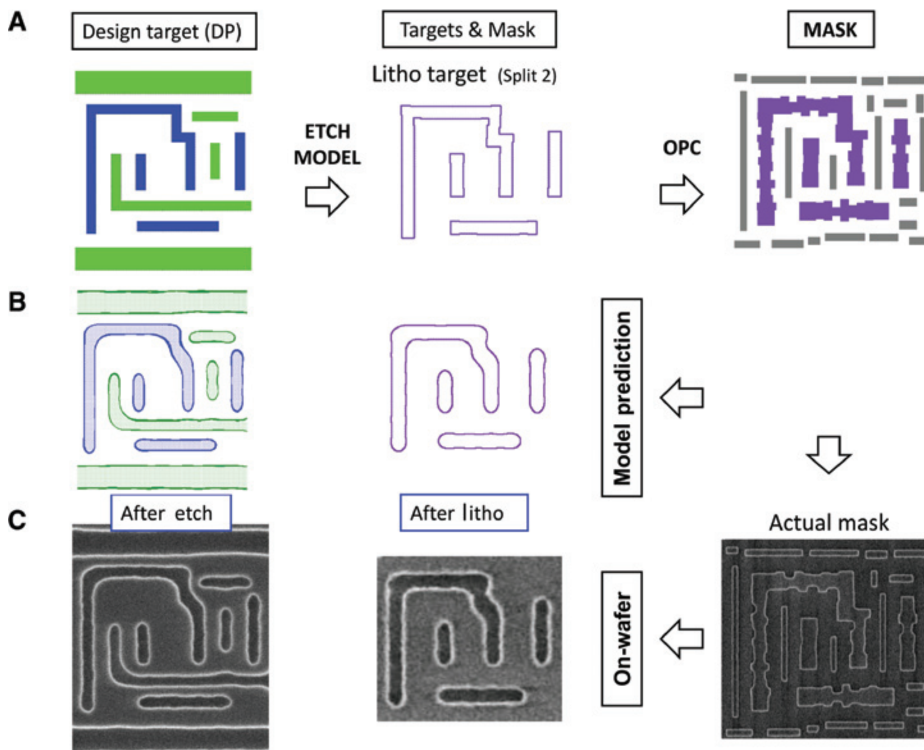

Figure 1: Illustrating OPC for a double-patterning N20 Metal1 example (NTD resist process). (A) OPCed mask pattern generation; the gray shapes are the SRAFs. (B) Predicted pattern after litho, i.e., in resist and after etch, and (C) actual wafer pattern after litho and etch. Also shown is a SEM image of the OPCed mask pattern itself.

target pattern into the litho target pattern, and the optics-resist model, then further transforms it to the pattern that is actually placed on the mask. For critical layers, this OPCed pattern also contains subresolution assist features, or SRAFs (gray polygons): these are not supposed to print, but help make the printing of the final pattern more robust against certain process variations [1]. Figure 1B and C illustrate the expected pattern (as predicted by the OPC software) and the actually observed pattern of the wafer, after the litho and the litho and etch step, respectively. (The etch step in this example is also used to trim down the dimensions of the litho pattern.)

Several electronic design automation (EDA) companies – e.g., Mentor Graphics, Synopsys, Brion (currently an ASML company), and Cadence – offer software packages that are capable of making such OPC corrections for a complete mask within an acceptable time.

The software that makes such an OPC correction to the design pattern is of course based on models that describe how a specific pattern on the mask is expected to print on the wafer. The success of the OPC correction is then evidently related to the accuracy of these models. In Section 2, we will discuss how such OPC models are built and how their quality can be evaluated. Application of these models in the actual OPC mask correction – an art in itself – and the computational verification step that is done immediately after that will be discussed in Section 3. Once the masks are available, an on-wafer verification of the pattern quality and robustness is done. This step will be discussed in Section 4. In the final section of this paper, we will show how geometrical simplifications in the allowed design patterns are becoming an important additional weapon in the battle for printable patterns. Figure 2 shows the flow from the target layout to the on-wafer patterning and refers each of these steps to the various (sub-)sections of this paper.

Before we start, the following remark must be made. The subject of 'optimization of lithographic printability' is, in fact, a very broad one, and covering all of it in a single paper is not feasible. This complete optimization flow includes things like choosing the resist and process stack, defining the stepper conditions (including the scanner source shape), defining an antireflection strategy, etc. [1]. This paper steps into the larger flow at the point where all these 'operational conditions' have been fixed and when at least some basic ground rules for the process layer under consideration are known. This paper will focus on today's state of the art, while also looking

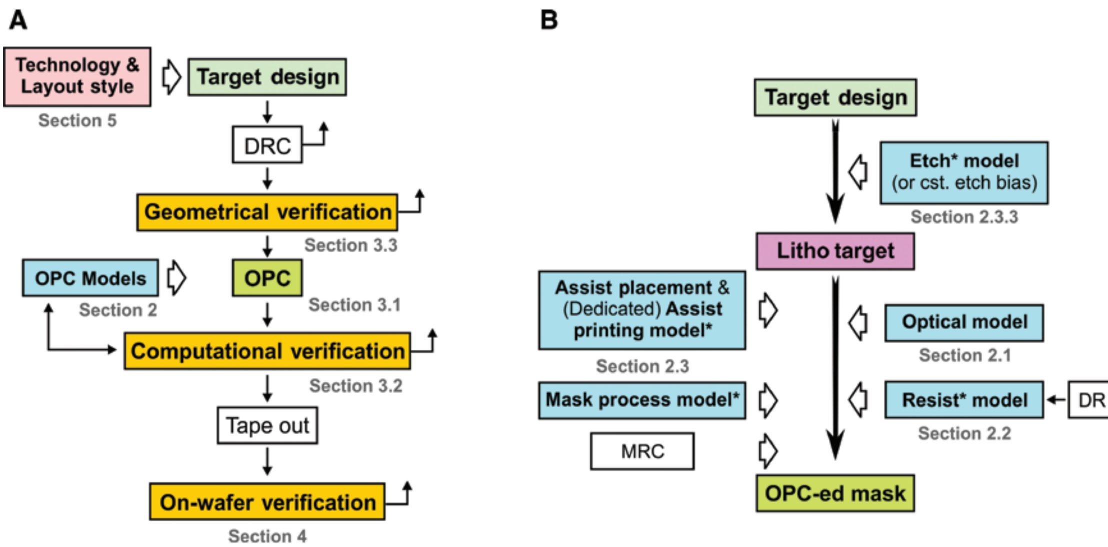

Figure 2: Simplified schematic of the steps that need to be taken to take the target layouts to robustly printing wafer patterns, making also reference to the sections in this paper. (A) General flow; (B) models involved in the OPC step (models marked with an asterisk need to be calibrated on wafer data). For most steps in the flow, iterations to previous steps usually take place (symbolized by the upward pointing arrows) before all patterns in the chip print to within the required quality specifications.

back into recent (published) history, to see how we got where we are today.

The Appendix contains a summary of most of the abbreviations used in this paper.

2 OPC models

The software that makes OPC corrections on the design pattern is based on models that describe how a specific pattern on the mask is expected to print on the wafer. In this section, we will look into these models, see how they are built (calibrated) and point out some of the mechanisms that could/should be included. The purpose is to review where potential sources of remaining inaccuracies or error could be, as these will transfer into inaccuracies in the OPC corrections applied and, hence, the quality of the printed wafer patterns.

Lithography models consist of two – and if we include also etch, of three – major steps [1]:

1. In the first step, the 'optical image' is calculated, using an 'Optical Model'. The term 'optical image' refers to the light-intensity distribution that is created by the projection lens inside the volume of the resist film, I(x, y, z).
2. The second step calculates how this optical image eventually gives rise to a resist pattern, after the bake/development/rinse step. This calculation includes mechanisms that describe how light interacts with the resist, interactions between the different resist components (such as acid-catalyzed deprotection, acid-quencher interactions, acid and quencher diffusion and development). At the end of this step, the variations in the optical image intensity have been converted into areas where the resist has been removed by the developer and areas where resist still remains. We will refer to models behind this step in the calculation as the 'Resist Model'.
3. In the final patterning step, the resist pattern is transferred into some material layer (stack) below, often using some dry-etch step. This step is then described by an 'Etch Model', i.e., some mathematical formalism that tells whether and how the etch process further changes the resist pattern.

We will discuss each of these separately.

2.1 Optical model

It is, in principle, possible to calculate the image intensity distribution inside the resist film very accurately (assuming that all the optical properties of the projection system are known, which in general is, to a large extent, the case), though not at an affordable price if this needs to be done for large areas, such as the full mask. OPC, as in the case of any complicated calculation, must find a

balance between the conflicting requirements of accuracy and speed. Simulators referred to as ‘rigorous lithography simulators’, such as Sentaurus Lithography (or S-Litho, from Synopsys, Mountain View, CA, USA), Prolith (from KLA-Tencor, Milpitas, CA, USA), Hyperlith (Panoramic Technology Inc., USA) or Dr. LiTHO (Fraunhofer IISB, Erlangen, Germany), simulate the lithographic process for a very small mask area only (typically a few µm by a few µm at the most) and can, therefore, go for the highest accuracy while keeping the calculation time within reasonable limits. The much larger mask area that OPC simulators need to cover necessitates approximations in order to keep simulation times realistic. The first simplification that is being made is that the image intensity inside the resist is only calculated (at least in general) in a single xy plane (parallel to the wafer surface), at some selected z-position (z_0), instead of in the entire volume of the resist film. It is from this two-dimensional image intensity, I(x, y), that the essentially three-dimensional developed resist pattern will be deduced with the Resist Model. We will come back to this simplification later.

When practical OPC came into use in the early 1990s, this intensity function I(x, y) was calculated with what is referred to as the ‘scalar approach’, which means that light was treated as a scalar wave, ignoring its polarization, i.e., its vector-wave character. Also, the mask was treated as infinitely thin, so that it can be described by a simple, binary mask function, M(x, y), that has values of either 0 or 1 depending on whether or not the mask is transmitting at position (x, y). This assumption is referred to as the ‘Kirchhoff’ or ‘Thin Mask’ approximation (TMA). An equation for I(x, y) under these assumptions has essentially been known since the 1950s and has been described in several publications, e.g., [1, 2]. Approaches to efficiently apply this formalism to entire mask areas express this intensity function essentially as [1, 4, 5]:

$$I(x, y; z_0) \approx \Sigma_n c_n |M(x, y) \otimes \phi_n(x, y; z_0)|^2 \quad (1)$$

i.e., as a sum of convolutions of the mask function M(x, y) with so-called ‘optical kernel’ functions, ϕ_n. It is important to note that both these kernels as well as the coefficients c_n are determined directly from optical properties of the illuminator, the mask, the lens, and the resist, so that Eq. (1) does not stand for a fitting operation. Also, the c_n coefficients decrease with increasing n, such that the sum can be truncated at some convenient maximum value of n. As the kernels do not depend on the mask pattern (i.e., on M), they can be calculated once (for the optical settings used) and then applied to the intensity calculation at positions (x, y) covering the entire mask. This is what makes this process so efficient.

In its use for doing OPC for advanced-node masks today, Eq. (1) has the following mechanisms built into c_n and ϕ_n:

1. The so-called ‘*thin-film*’ *effects*: light reflections at effectively every interface between the different layers of the wafer-resist stack modify the intensity inside the resist film, creating a standing wave in this layer [6].
2. The ability to include lens aberrations – which includes an offset from the best-focus plane (a ‘defocus’), or lens apodization – if desired. Especially, the ability to include defocus as well as dose as variables into the model is important. This has led to the so-called process window (PW) models, i.e., models that can also predict how mask patterns will print in off-nominal focus-dose conditions. Such PW models are being used today not only to print the wafer patterns close to the design target when the scanner is operating at its optimum (nominal) focus and dose setting, but also to minimize the *variation* of the printed patterns if some (unwanted but inevitable) focus or dose errors occur. This approach has been called process window OPC or PWOPC see, e.g., Ref [7]).
3. The use of a ‘vector model’ [8, 9], which is essential when applying Eq. (1) to lenses with the large numerical apertures (NA) that are common today, for which the scalar (or paraxial) approximation is not valid any more. Such a vector model takes the polarized nature of light into account. Today, scanners used to print the critical layers of the advanced nodes are all high-NA systems, so this vector approach has become standard.

There are, however, a number of mechanisms that are not included in the optical model as we have described it here so far, some of which are considered by many to be equally important as the ones we just highlighted. Today’s OPC software packages have been or are being extended with advanced options to also include these additional effects. We will now briefly discuss the more important ones.

2.1.1 Mask topography or mask-3D (M3D) effects

The absorption layers on the masks used in lithography are thin (a few tens of nm) but not infinitely thin, so the Kirchhoff (or Thin-Mask) approximation that diffraction effects that occur at mask level can be calculated simply as the Fourier transform of M(x, y), is not exact. It has become clear that this approximation is often not justified when applied to the type of mask structures that are

being used today, and it is now considered by many as essential to include M3D effects in advanced OPC calculations. Since the early years 2000, EDA companies started working out approximate ways to include M3D effects in their OPC packages. Although this task is now largely completed, it has proven to be a nontrivial one.

M3D effects were first studied in the context of what we previously already called 'rigorous lithography simulators' (e.g., S-Litho or Prolith): including M3D effects in a fully correct way essentially requires solving the Maxwell equations which is slow, but can be done with appropriate mathematical techniques if the mask area for which the calculation needs to be done is small. Examples of such studies, with some discussion on the methods used, can be found, e.g., in references [10–12]. From the combined literature on this topic, it has become clear that M3D effects induce pattern-specific CD changes (and, hence, affect the proximity behavior) as well as changes in the through-focus behavior (Bossung distortions and/or best-focus shifts) of certain structures. To make matters more complicated, it was also found that a correct description of M3D effects requires that the Maxwell equations need to be solved separately for light incident on the mask at different angles of incidence [12], as opposed to assuming that the amplitudes and phases of the diffracted light are independent of the incident angle (something which has been called the 'no-Hopkins' vs. the 'Hopkins' approach), when solving the Maxwell equations.

Apart from having an impact on the proximity behavior, itself, M3D effects are also relevant because of the best-focus shifts they induce on some structures. We illustrate this in Figure 3 with the example of three structures we used in the OPC calibration/verification data set for our 20-nm node Metal1 OPC model.

Let us now return to OPC intensity simulations and see how EDA companies have tried to implement M3D effects. One key point to realize is that rigorously solving the Maxwell equations cannot be done on the scale of an entire mask, which means that approximate methods had to be developed that are fast enough to permit full-mask application. It seems that one of the first breakthroughs in efficient M3D calculations was an approximate treatment called the 'domain decomposition method' or DDM [13], later extended to take into account the 'non-Hopkins' oblique incidence effect [14]. Other EDA companies have developed their own approach to solving the M3D puzzle; see, e.g., references [15–18]. A more recent extension of the DDM approach is described in references [19, 20]: DDM essentially treats the impact of topography edges on the mask as independent from each other, but this is an approximation that starts to become invalid as the distance between these edges gradually becomes smaller with decreasing target sizes.

Owing to a lack of published details on each of these implementations, potential differences between the EDA M3D models are difficult to assess. Also, no benchmarking

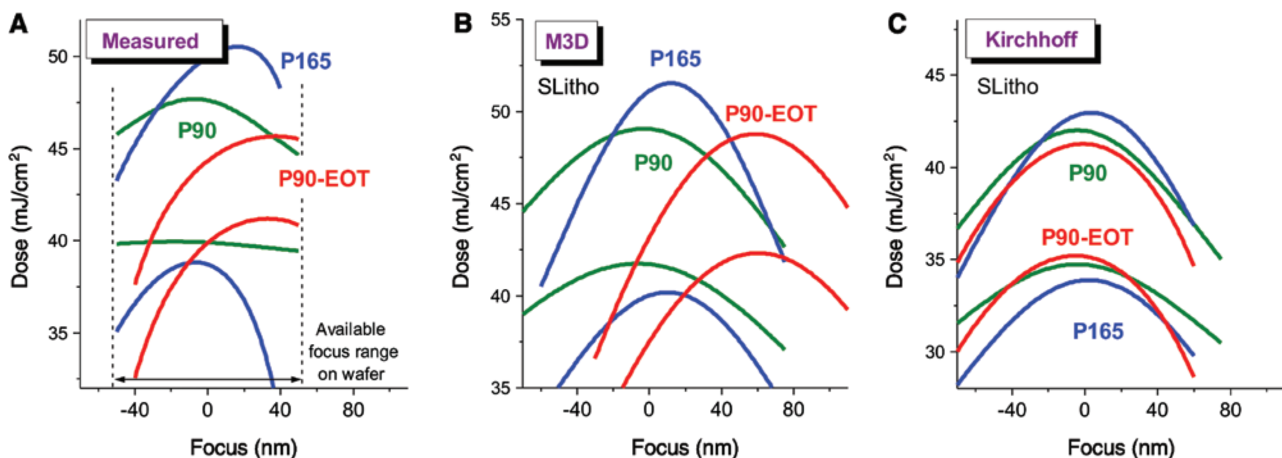

Figure 3: Illustration of the M3D effect on the Process Window of three structures: a 90-nm and 165-nm pitch trench case and a small-gap End-of-Trench (EOT) structure in a 90-nm trench array. (A) Measured from our 20-nm Metal1 OPC model calibration data set, (B) and (C) S-Litho simulated (using an NTD resist model), with a M3D and Thin-mask ('Kirchhoff') model, respectively. The (qualitative) agreement between (A) and (B) shows that M3D effects contribute to the observed best-focus shift. Especially, the EOT case is affected. Note, however, that the amount of focus shift decreases if either the trench pitch or the EOT gap size increases. In our actual 20-nm M1 Logic-cell block, the M3D impact on best focus was much smaller than in the examples shown in this figure. (We allowed the 90-nm pitch in our N20 design rules, but in the double-patterning split Logic patterns, it does not occur). The small best-focus differences in the Kirchhoff case are due to resist effects.

has been published, and their performance cannot be compared. But most of the publications cited report improvements in OPC model quality when M3D models are used, so the current methods have proven the potential to provide benefit. Most important is the ability to account for M3D-induced modifications of the through-focus behavior of certain structures (e.g., focus shifts) [16, 20, 21], as this is a purely optical effect that should not become 'absorbed' into the parameters of the resist model. Because of the growing importance of these M3D effects, we expect to see further development and validation in the coming years.

2.1.2 Mask process corrections (MPC)

The patterns on real masks, of course, always deviate to some extent from the intended pattern (see the example in Figure 4). A mask has errors, and these will affect the patterns that are printed with them. Mask errors range from global mean-to-target (MTT) deviations, corner rounding, mask-proximity effects, over-etch depth variations, across-mask non-uniformities, etc. An early example of the impact of mask errors and a recent review of the various mask errors that are relevant for OPC can be found in references [22, 23], respectively.

If these mask error effects are constant for a given mask process, and if they can be modeled into what is called a 'mask-process model', it is possible to take them into account during the OPC phase of masks that will be generated later with the same mask process, i.e., it is possible to correct for these errors or at least take them into account during the OPC process. Mask errors then need to be first measured from a test mask, from which a correction model can be derived for that specific mask process. This means that the ideal mask function, M(x, y), in Eq. (1) is replaced during OPC by a corrected mask function, let us call it $M_{corr}(x, y)$, that is closer to what the actual mask will look like. The generation of this corrected mask function can be done in different ways. An overall MTT deviation (or global mask bias) or corner rounding that is constant for all corners can be implemented by a simple operation on M that requires only a few variables as input. These variables can be obtained from a few measurements on a mask made previously with the same process and does not require any extensive mask model building. Figure 4 illustrates this basic concept with a simple example. It shows a mask SEM image and a mask contour that was derived from this image. This contour is then compared (overlayed) with a mask shape that was obtained from a simple bias- and corner-rounding operation on the ideal OPCed shape. Although there are some differences with the measured shape, this simple mask model is clearly already a step in the right direction.

Compensating for more complicated effects, requires a more sophisticated approach that is based on more extensive mask measurements. Depending on which effects one wants to correct for, a set of measured mask error data, ΔCD_M, will be fitted to a model. A number of model functions that have been mentioned in the literature [24, 25] could be combined into the following equation

$$\Delta CD_M(x, y) \approx \Sigma_i a_i M(x, y) \otimes K_i^S(x, y) + \Sigma_j b_j D(x, y) \otimes K_j^L(x, y) + P(x, y) + MTT \quad (2)$$

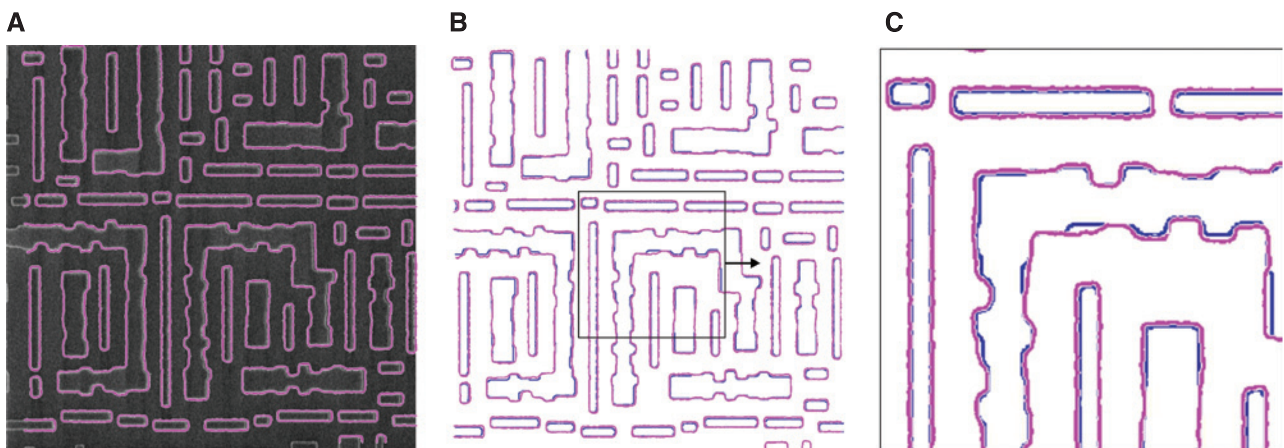

Figure 4: (A) Actual mask SEM image of the OPCed pattern shown in Figure 1, together with a mask-edge contour (pink) derived from it. (B) and (C) This measured contour is compared to a simple model shape (blue) obtained by applying a global bias and corner rounding on the ideal after-OPC shape shown in Figure 1.

The first term represents 'short range' mask effects, as a sum of convolutions of M with a set of kernels that are assumed to describe these effects. The second term represents 'long-range' effects, having their own kernels that are now convoluted not with M but with a more coarse 'mask-density function' (D) that is derived from M. The third term is some appropriate polynomial that represents the overall (i.e., structure independent) across-mask non-uniformity. An important difference with Eq. (1) is that, here, the kernel functions K^S, K^L, and P are not derived from some rigorous theory but have been adopted by the EDA because they were found to fit well with experimental data in earlier exercises. The parameters (a_i and b_j) in the two sum terms as well as in the polynomial now have to be fitted to experimental ΔCD_M data. Once the mask process model of Eq. (2) – or a similar expression – has been experimentally determined, it can be used to predict what the actual mask contour will look like for a given intended M(x, y), and this function can then be used instead of M in the calculation of the optical images of Eq. (1).

It does not seem as if the MPC correction approach as sketched in this section is standard practice today. In the past, mask errors were usually not taken into account explicitly in OPC model building, which means that they were effectively 'absorbed' into the fitted parameters of the resist model. Although this approach seems to have worked well for 'older' technology nodes, incorporating mask effects into the resist model must lead to some error, and with diminishing process margins everywhere, it can be expected that some sort of MPC approach will become increasingly beneficial.

2.1.3 Resist 3D

Traditional OPC modeling considers the image intensity in a single xy-plane only (see Eq. 1) and then uses its resist model to derive the printed resist pattern as a contour in that plane. We will refer to this as the 'Resist-2D' (or R2D) approach. But in reality, resist patterns are three-dimensional shaped, with (structure dependent) effects like variable side wall angle (SWA), resist top loss (RTL), resist footing or undercut, etc., and none of this complexity can be captured in a 2D model. As the etch process that usually follows the litho step is expected to behave slightly differently in the presence of SWA or RTL, and can even lead to patterning failures depending on what the 3D resist profile actually looks like, there is a rapidly growing interest in what is now called R3D models for OPC, i.e., models that, at least to some extent, incorporate some information on how the printed resist pattern changes in the z-direction. Other useful applications for R3D models would be SRAF printing (which sometimes takes place at the resist surface or bottom only) or other effects that preferably happen at the resist surface or the interface with the underlying substrate (such as footing or scumming).

Rigorous lithography simulators like S-Litho and Prolith have always included this third dimension and, in fact, deliver a 3D resist profile. The example in Figure 5 illustrates this for the case of one if our N20 Metal1 patterns that we found to be weak after etch (i.e., having a limited process window), something which could be attributed to the resist profile being sloped at the position of the weak spot [26]. A similar literature example can be found in [27], where a rigorous simulator was also used as a 'forensic tool' to understand the cause of failures observed after etch. Although this use of rigorous simulators is quite valuable, one would like the OPC model to flag suspected resist profile-related hotspots *before* the mask is built, i.e., during the mask printability check that is done immediately after OPC ('computational verification', see Section 3.2). This, however, requires that the OPC model be extended into the third resist dimension, i.e., it requires R3D OPC models.

Extending the OPC models, themselves, to R3D has not been all that difficult. Indeed, the original R2D

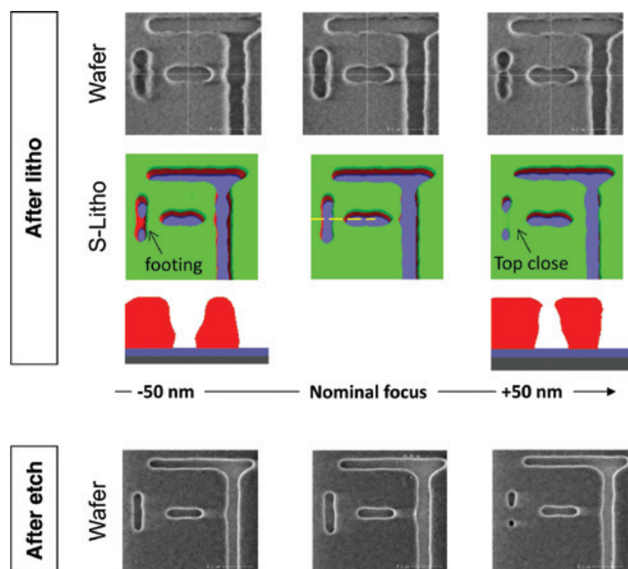

Figure 5: Example of a potential R3D hotspot from our Tespa20 Metal1 test vehicle (i.e., trench printing), at nominal focus and at a +50 nm and -50 nm defocus [26]. The S-Litho-simulated resist profiles show a clear difference in resist profile for positive and negative defocus, in agreement with the observed after-litho top-down SEM images. After etch, these profile differences translate into a quite different result: at positive defocus, the short vertical trench disappears almost completely.

models – although confined to a single (x, y) plane at a certain vertical position z_0 inside the resist – can be built at *any* value of z_0. Building multiple R2D models, each for a different z_0 [28], was indeed the first step in the direction of R3D. (An example will also be shown in Figure 16). What is missing in this initial approach, however, is vertical diffusion, i.e., the fact that chemicals inside the resist are not confined to a single xy-plane (as R2D models implicitly assume) but also move in z, thereby effectively coupling the xy-planes together. Today, all major EDA companies have extended their models with some vertical-diffusion mechanism [29–31], and some initial validation of their models has been published, e.g., by comparing their predictions with the results of a rigorous simulator [29] or with printed wafer data [34].

If the mathematical framework for R3D models seems essentially ready or at least known, the real problem with R3D models is how to calibrate them, as this requires reliable wafer data on actual resist profiles, preferably for a reasonably large set of structures and preferably including different types of structures. That is not obvious.

One technique that has the potential of yielding resist-profile information and that has a short measurement time (which means that measuring a lot of structures is feasible), is scatterometry, also called optical CD (or OCD) metrology. Application of OCD for OPC purposes has, indeed, been tried for both 1D and 2D structures (see, e.g., [28, 32, 33]), and although some of the initial results were promising, it seems that the momentum to use OCD for resist-profile measurements has largely faded away. The number of variables that need to be included in the OCD models (that fit the measured spectra and convert them into resist profile information), simply seems too large to obtain reliable, absolute values for, e.g., SWA or RTL. This is even more so if the wafer stack used contains multiple layers beneath the resist or if 2D patterns need to be measured, in which case the number of geometrical parameters increases rapidly.

Cross section SEM (XSEM) or atomic force microscopy (AFM) has, of course, been often used to measure resist profiles. But both techniques are slow and not applicable to all the structures that are usually included in resist-model calibrations for OPC, such as 2D structures (XSEM) or narrow isolated spaces (AFM). Nevertheless, an R3D model that was recently built from XSEM and AFM data [35] has shown some promise. Whether enough XSEM/AFM data can be collected to build a model that can be applied to a wide variety of structures still remains to be demonstrated.

Another approach that has recently been taken [29, 34] is to use rigorous simulators to generate resist-profile data for many structures and use that data as input for an R3D model calibration. Rigorous simulators do not require a large set of measured structures to build a model from, and once a calibrated resist model is available, it can be used to calculate resist profiles for as many structures one wants, both 1D and 2D. The Achilles' heel in this approach is, of course, the question how accurate these simulated profiles are. Also, this approach would, therefore, require further validation.

The conclusion of this section – at least today – seems to be that there is an increasing demand for R3D models, but that there is no proven nor accepted path to generate the calibration data it requires. We, therefore, expect continuing efforts in this field. An evolution that will relax, to some extent, the need for accurate R3D models or the variety of structures they should cover, is the trend toward simpler and more regular layouts (which we will discuss in Section 5).

2.1.4 Effect of a non-uniform wafer substrate: Wafer 3D

Wafer topography effects or Wafer 3D (W3D) effects are not an issue for most process layers: the top surface of the wafers is usually planarized, and the use of some antireflective strategy effectively 'hides' any buried topography below the surface. There is (or was), however, one exception, namely, the implantation steps that occur in the front-end processing. The use of a Barc adds to cost and also poses technical difficulties in these cases, which is why a Barc was traditionally not applied for implantation layers: the resist is spun directly on the post-Active or post-Poly wafer topography. This means that the implantation exposure is affected by reflections from the substrate (which are different on Si, oxide, and poly-silicon), as well as by the wafer topography of Active or the combined topography of Active and Poly. These implantation layers were long considered to be noncritical layers for lithography, such that W3D effects could be ignored or could be dealt with by using simple rule-based corrections, but with recent nodes, this has gradually been changing. W3D effects have become quite severe, and as correct implantation profiles are key to the device operation, an increased interest to include actual W3D modeling into the OPC flow can be seen in the literature since about 2009.

Rigorous simulation solutions to wafer topography effects have existed for quite some time (see, e.g., references [36–38], but these again use methods (e.g., finite element, FEM, or finite-difference time domain, FDTD) that are far too time consuming to be applied full chip.

So, over the last few years, a number of approaches have been developed each of which provide a way to incorporate W3D effects by adding extra terms to the intensity calculation, to take into account the reflected light from the various patterned areas on the wafers, as well as (at least in the most elaborate schemes) the interference between this reflected light and the incoming light.

The exact details of the different methods are not public, but they all start from traditional model(s), i.e., models assuming a flat, uniform substrate. The light intensity calculated with these models is then 'corrected' with additional terms, representing the nonuniform reflection and – sometimes – also interference terms. An initial approach to obtain these extra terms [39–42] was to define one or several 'virtual' or 'fictive' masks, derived from the Active (and Poly, in case of a post-Poly implantation) mask, from which these extra intensity terms were then calculated. One way to include interference terms between incoming and reflected light is to define a set of appropriate kernel functions for mixing the incoming and reflected waves that are defined such that the resulting model fits measured CD data from actual topographical wafers (see, e.g., reference [43]). Two other approaches that tried to tackle the problem in, perhaps, a more physical way can be found in reference [44], where a fast EMF-solving algorithm (called the 'generalized source method') is used, or in [45], where the wafer image is calculated with a method using a library of rigorous EM fields.

Most of these recent papers (and the references they contain) use actual wafer data to verify and demonstrate the concept they propose. Also, several EDA companies now have a W3D module in their OPC software packages. However, it also seems clear that the traditional approach to do the implantation litho steps without Barc or another planarizing approach is becoming untenable. With shrinking dimensions and increasing topography that comes with the introduction of Fins instead of STI (Active), it is increasingly difficult (to impossible) to keep avoiding the use of Barc or another planarization approach in the implantation steps. Reference [46] shows wafer data and rigorous simulations of combined Fin-Poly topography from a 10-nm node example, where the topography is such that it is not possible any more (at least at some positions) to get enough light intensity down to the bottom of the resist layer to clear the resist. Figure 6 shows a similar example for a planar technology SRAM case.

That is something OPC cannot correct but needs to be solved by a modification of the patterning process. For a while, the hope has been that wet-developable Barcs would solve this issue, but this idea seems to be abandoned today, which means that more traditional planarizing approaches have to be used. If this becomes the mainstream approach for the most advanced nodes, the need for an interest in W3D modeling is likely to fade away again.

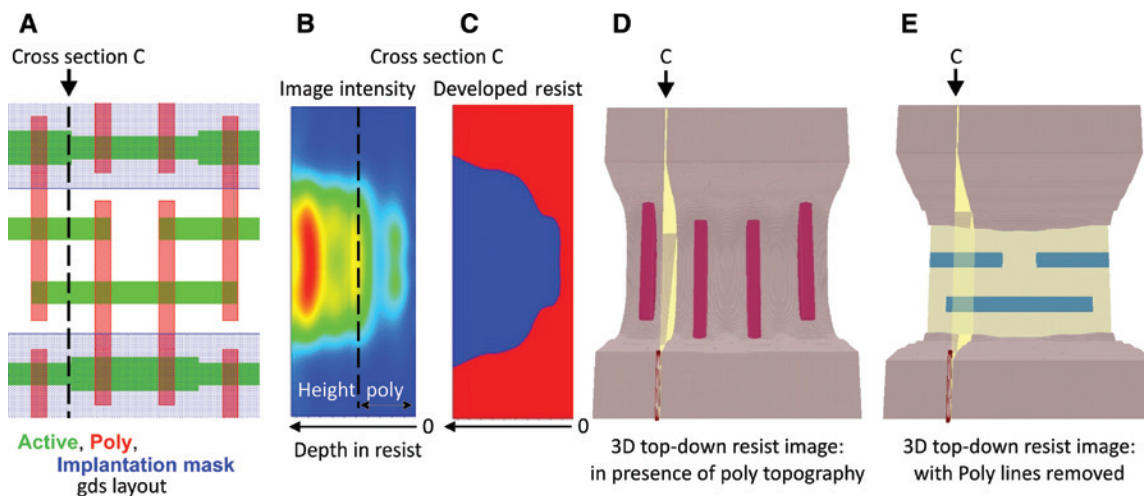

Figure 6: Prolith simulations for a post-Poly implantation step on a planar SRAM case (28 nm node, i.e., Poly pitch of 110 nm), (A) Different mask layers in gds; (B) Prolith-predicted image intensity (cross section along plane C); (C) predicted resist profile (in the same cross section plane); (D) 3D top view of the predicted resist pattern in the entire simulated area. The Poly lines, themselves, emerge from the resist, but in between the poly lines, the resist is not cleared. (E) The same resist simulation in the absence of the Poly lines: then, the resist clears nicely (showing the active areas at the wafer surface), so it is really the Poly topography that 'prevents' the light from reaching the resist bottom. In case of a FinFet-type Active (instead of planar), this issue only becomes worse.

2.1.5 Laser bandwidth

Especially in 193-nm exposure systems, the projection lens has a non-zero longitudinal chromatic aberration, characterized by a quantity $F_\lambda \equiv dF/d\lambda$, which gives the shift of best focus with a laser wavelength change. The laser itself is non-monochromatic but has a finite bandwidth, and it has been shown that this induces CD changes for every structure (changes compared to the hypothetical monochromatic laser case) that can be approximated as [47]

$$\Delta CD \approx M_2 F_\lambda^{\,2} Q_F \propto E95^2 F_\lambda^{\,2} Q_F \qquad (3)$$

where the second moment, M2, and E95 are metrics that characterize the laser bandwidth [47], and $Q_F \equiv \frac{1}{2}\, d^2CD/dF^2$ is the 'quadratic focus sensitivity' of the printed structure under consideration. So this equation tells that the finite laser bandwidth modifies the optical proximity, an effect that scales with Q_F. Most often, the OPC software assumes a monochromatic laser, which means that the ΔCD effects of Eq. (3) will be absorbed into the resist model.

It is, however, possible to incorporate a laser bandwidth effect into the optical OPC models by calculating the intensity in the image plane as a superposition of contributions from separate laser wavelengths, each generating a best-focus plane at a slightly different position (consistent with the value of F_λ). An example of such an approach can be found in reference [48].

2.1.6 Stray light (or flare)

Stray light, also called flare, occurs due to light scattering along the optical path between the mask and the wafer and results in a blurring of the optical image. This blur is usually described by a stray light or flare point spread function (PSF_{SL}). It has been shown that flare adds an additional intensity at position r=(x, y) that can be approximated as [49]

$$\Delta I_{SL}(r) \approx PSF_{SL}(r'-r) \otimes I(r')\big|_{r'<R} + PSF_{SL}(r'-r) \otimes D(r')\big|_{r'>R} + cst \qquad (4)$$

The first term in this equation is a 'short-range' contribution that gives the scattered intensity at r from position r′ not too far away from r: this contribution convolutes the no stray light intensity function I(r′) with PSF_{SL}. Scattering contributions from positions r′ farther way from r (second term) are calculated with a similar expression, except that now, the convolution with the PSF can be done with a more coarsely calculated 'image', for which one can use some grayscale or mask density function, D, to speed up the calculation. The transition distance R can be chosen somewhat arbitrarily. The PSF_{SL} function can often be described by a fractal or double fractal function [49–51]. From Eq. (4), it follows that the stray light-induced intensity change will be largest from the mask with a high transmission and negligible for very dark-field masks.

Stray light levels are usually small enough in modern optical lithography scanners (i.e., operating at 193 or 248 nm), such that OPC typically does not need to take it into account. The situation is different, however, in EUV lithography: the amount of light scattered from a surface with a given roughness is approximately inversely proportional with the square of the wavelength. Whereas in optical lithography, flare is often caused by contamination of certain optical surfaces, in EUV, the smoothness of the mirrors themselves, is the determining factor. Although flare levels of the latest generation EUV tools have gone down significantly compared to the earliest EUV tools, OPC for EUV will, in general, try to correct for it, especially in the case of masks with a high reflectivity or with high reflective areas. More detailed discussions of the role of flare in EUV lithography and on how it can be measured can be found in references [50, 51]. Today's OPC tools for EUV applications are well equipped to take flare into account during the calibration of OPC models as well as during the OPC mask-correction phase [52]. For a given mask and PSF_{SL}, they essentially pre-calculate $\Delta I_{SL}(r)$ of Eq. (4) for the entire mask (this function is often called the 'flare map') and then add this intensity to the no-flare intensity expression of Eq. (1) during the OPC calculations.

2.1.7 EUV mask shadowing and other EUV-specific effects

EUV lithography presents a few additional proximity-related challenges that do not exist in optical lithography. The most important of these is the so-called shadowing effect [53–55]. EUV masks are reflective masks, which implies that the average angle of incidence of light from the illuminator cannot be at 90° to the mask surface but needs to be offset by a small angle ε (today ε=6°). This implies that mask structures with a different orientation will be illuminated differently, the result of which is that both the printed CD and the pattern placement will not be the same. Figure 6 illustrates the CD difference between horizontally (H) and vertically (V) oriented trench array structures. The CD and placement differences also change cross the slit. These effects need to be corrected for.

The initial approach to solving the H-V printed CD difference was to apply a constant H-V bias on the mask, i.e.,

shift the horizontally oriented absorber edges with a fixed amount with respect to their design target position. It was, indeed, found from rigorous calculations that the H-V bias needed to completely eliminate printed H-V CD difference was almost constant for 1D structures except at the smallest dimensions (mask space width below ~20 nm at 1×). Figure 7 illustrates this.

The constant H-V bias leaves, however, some residual CD differences and even breaks down for small 2D structures [55]. With the tight tolerances that need to be put on the dimensions for which EUVL can be expected to be used, a more accurate approach is desirable. Fortunately, efficient M3D solvers are now also available within the OPC packages for EUV [52, 56]: in principle, these should capture the variation of the shadowing effect for different structures and through the exposure slit and, hence, make the constant bias approximation unnecessary. Reference [56] demonstrates that the M3D model used produces results that are very close to those of a rigorous simulator.

The situation of the shadowing-induced pattern shifts is somewhat different. Although these pattern shifts are, in reality, several nm, the alignment of the mask before a wafer exposure will detect and correct the pattern shift of the alignment marks, before the exposure begins. This means that the OPC tool only needs to deal with structure- and slit position-dependent placement error *differences*. Reference [56] shows that M3D models of today's OPC tools can also correct for these.

OPC for EUV has been demonstrated for a number of cases; see, for example, references [57] and [58]; the latter also includes a discussion of residual error mechanisms and, hence, where further improvements in the OPC modeling are required or desirable.

If EUV tools move into higher NA values, additional effects will grow in importance. One is the angle of incidence dependence of the multilayer reflectivity of the mask [59]. Also, this type of effect will then need to be addressed in the OPC modeling. EDA companies are currently incorporating this effect in their software.

2.2 Resist model

The approach to generate the resist model is fundamentally different from the approach to generate the optical model. Whereas the optical model is based on first principles, though with approximations made to meet the desired balance between rigor and speed, the resist model is essentially an empirical model that is fitted to measured wafer data. Even in rigorous lithography simulators, not all the processes that play a role in converting the 3D image intensity in the resist film into a developed, three-dimensional resist pattern are completely understood, let alone correctly modeled in all their complexity, and yet, the resist part of the (rigorous) simulation takes a large fraction of the simulation time.

OPC modeling has, therefore, taken a different approach to the resist model: it applies some 'mathematical conversion' to the calculated intensity, I(x, y), that eventually predicts where the resist will be developed away and where it will stay. Every EDA company has its own specific approach. Common is that this 'conversion operator' contains enough degrees of freedom as well as a set of free parameters that can be fitted to a set of CD data that are measured from a printed wafer: the calibration data set. (Examples of such empirical model forms can be found, e.g., in [4, 60].) The challenge is then to construct

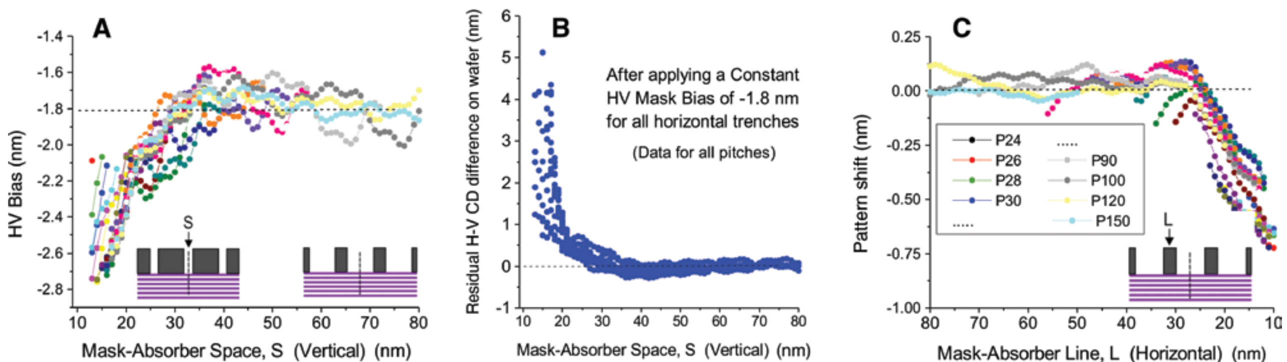

Figure 7: Shadowing simulation example: trench printing (NA=0.33, Annular 0.9/0.65, Image-in-Resist model). Simulation for pitches (P) ranging from 24 nm to 150 nm. (A) Simulated HV Bias for each individual structure. (B) Residual H-V wafer CD difference, if all horizontal structures on the mask are corrected with a fixed HV-Bias of -1.8 nm. (C) Pattern displacement of horizontal trenches (assuming that the pattern shift of very large structures has been automatically corrected for during the reticle alignment in the scanner).

this calibration data set in such a way that the resist model that is fitted from it is also capable of predicting the resist-printing behavior of structures that were *not* included in the calibration. The art of generating an accurate resist model, therefore, consists of

– defining an appropriate calibration-structure set,
– generating high-quality CD data for these structures,
– calibrating the resist model by fitting it to the calibration data,
– verifying that the model can, indeed, be reliably used to be applied to all the structures that can occur in the targeted application. This verification step makes use of a second set of wafer measurements, independent of the calibration set. The model quality is then assessed by its ability to predict the CDs of these verification structures.

The metrics that are being used to quantify the quality of either the calibration or the verification steps are, therefore, based on the difference between the measured and modeled CD of each individual structure j:

$$\Delta CD(j) \equiv CD_{modeled}(j) - CD_{measured}(j) \tag{5}$$

Although these individual model residuals are ideal for detailed analyses (looking for trends, for example), it is also convenient to reduce such a large number of values into more compact metrics. For example, specifying how many percent of a certain structure type (e.g., 1D, Line-Ends, …) have a $|\Delta CD(j)|$ below a certain target value, ΔCD_{min}., a quantity that we call the 'In-Spec Percentage' (or ISP). Often used is also the RMS value over all the $\Delta CD(j)$ values [4],

$$ErrRMS \equiv \sqrt{\Sigma_j \Delta CD_j^2 / N} \tag{6}$$

even though such a single-value metric can be deceptive: a low RMS value does not exclude that the model fails for a very small group of structures. Figure 8 shows an example

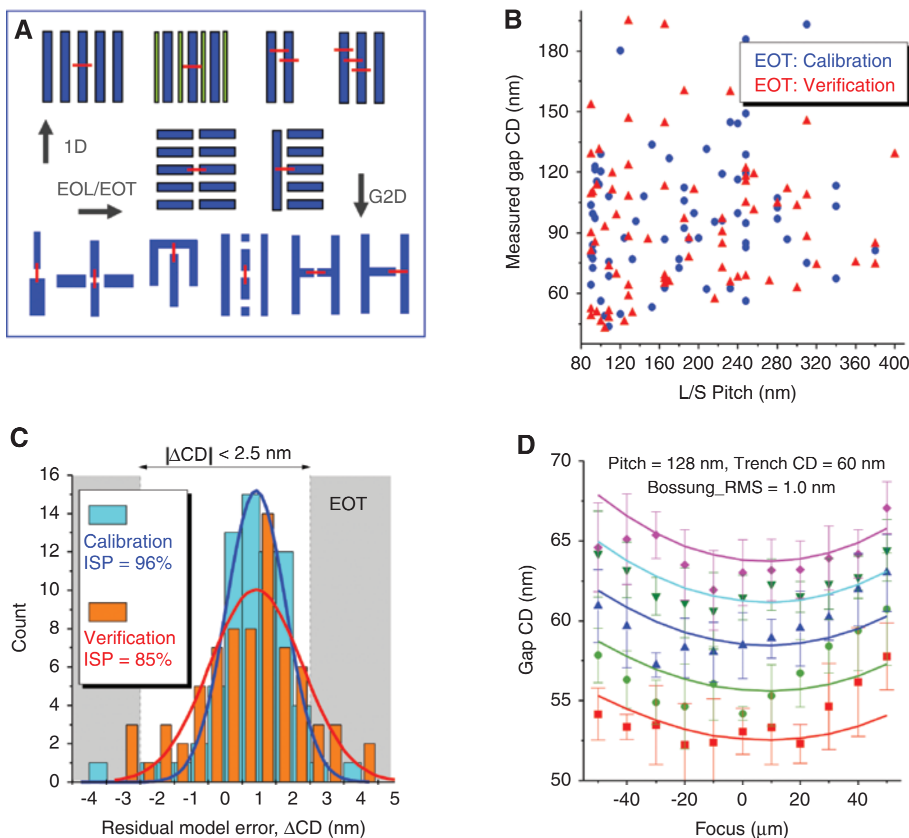

Figure 8: Resist model for OPC on our 20-nm node Metal1 test vehicle (NTD resist), illustrating (A) the type of calibration structures we used, (B) Pitch-gap combinations for the End-of-Trench (EOT) structures in the calibration and verification structure sets, (C) ΔCD histogram and ISP (in-spec percentage) values of the EOT CDs of calibration/verification, (D) example of a measured EOT Bossung curve (part of the calibration structures) and its model prediction (line curves). Line-end control is very important for metal layers; hence, the emphasis on the line end OPC model performance in this example. The Bossung_RMS value is calculated as in Eq. (6), but using only the $\Delta CD_j(F, E)$ values of this particular L/S structure j.

of some of the elements we just mentioned, for the case of our 20-nm node Metal1 OPC resist model.

In the following subsections, we will discuss each of the steps that should lead to a (hopefully) good resist model separately. Important to note, however, is that even when a model of the required accuracy is obtained, its validity is, in general, only guaranteed for the specific conditions (wafer stack, resist process, scanner settings, mask type) at which it was calibrated. If any of these conditions is changed, the model at least loses accuracy or becomes invalid altogether.

2.2.1 Selection of calibration structures

The key requirement for defining the calibration structure set is that they need to be 'representative' for the target application, which means that the model created with these calibration structures should be able to predict all the structures that can occur in the final design to which the model will be applied. Another way of formulating this, is saying that the calibration structures must be consistent with the design rules (DR) of the target application. At the same time, the size of the calibration structure set needs to be small enough to keep the time needed to measure all these structures within reasonable limits. It is because of these two competing requirements that people have been working on methodologies for efficient selection of calibration structures.

An initial approach, that has – with many variations – become quite commonly used, uses the approach of sampling a structure parameter space (PS) of some kind. This method originated from Mentor Graphics' VT5 method [60] in which the printing of a given mask-polygon edge is related to a number of properties of the image intensity around that edge: Imin, Imax, Slope, and Curvature. This led to the idea [61–63] to represent potential calibration structures themselves, as points in a multidimensional image-parameter space (IPS, four dimensional, if the four parameters used are Imin, Imax, Slope, and Curvature. The concept can be extended to more or to other parameters). The assumption then is that structures that are close to each other in this IPS are 'similar' to each other, from which the idea follows to select structures for resist model calibration such that they cover the entire space, e.g., by sampling this space along some uniform grid (see Figure 9A). The IPS concept is not exclusively linked to the four 'original' VT5 parameters, nor to the use of a specific EDA tool. We have, for example, used S-Litho to calculate the parameters for a contact-hole EUV model, using a five-parameter space [57], selecting structures from a few hundred thousand candidates. Figure 9 shows another example case from our own work (resist-line printing, with 193i), in which a selection of ~150 structures out of an original 4000 calibration structure set was found to deliver a model that was only slightly worse than the original 4000-structure model (i.e., the total RMS is only slightly higher, and the ISP values slightly lower: see Figure 9C and D).

Several further developments of these original concepts have been reported as well. Physical image parameters such as Imin, Imax, and Slope are clearly not independent parameters, so Singular Value Decomposition (SVD) techniques have been used [64, 65] to generate a new, orthogonal parameter set from the original (non-orthogonal) image parameters, thereby, further reducing the number of calibration structures needed. Extra parameters have been added to include through PW variations or sensitivity to mask errors, and sampling of the parameter space has been based on the identification of clusters of structures, rather than applying a uniform sampling [65]. One EDA company has developed a parameter space based on orthogonalized CD sensitivities [64]. Parameter space sampling methods appear very common now and have been successfully used by many.

It has, however, also been claimed that an approach based on image parameters only does not always lead to a representative structure set [66], and even though there can be other causes in the case of a failing model than just the content of the calibration set, it is true that the meaningfulness of any parameter-space approach depends on the relevance of the parameters chosen. There is, for example, no proof that the four original image parameters, or any alternative set, are the only parameters that matter in the conversion of an intensity distribution into a resist pattern. In that sense, it is the success of its use that justifies its applicability. Also the turn-around time for calibrating a model remains a concern. So improved structure sampling methods are still being tried; reference [67], for example, proposes a new mathematical formulation based on certain objective functions that can be specified according to the requirements of the process layer under consideration. This, the authors claim, leads to a further reduction of the amount of calibration structures needed.

If the trend toward more regular design styles is continued (see Section 5), it can, however, be expected that structure selection becomes easier again (at least for some layers) because of the much reduced variety of structure types in the target layouts. Let us elaborate on this for a moment. In Section 5, we will show that there is a trend to what is called unidirectional layouts for many layers, including the metal layers. If we now look at the OPC

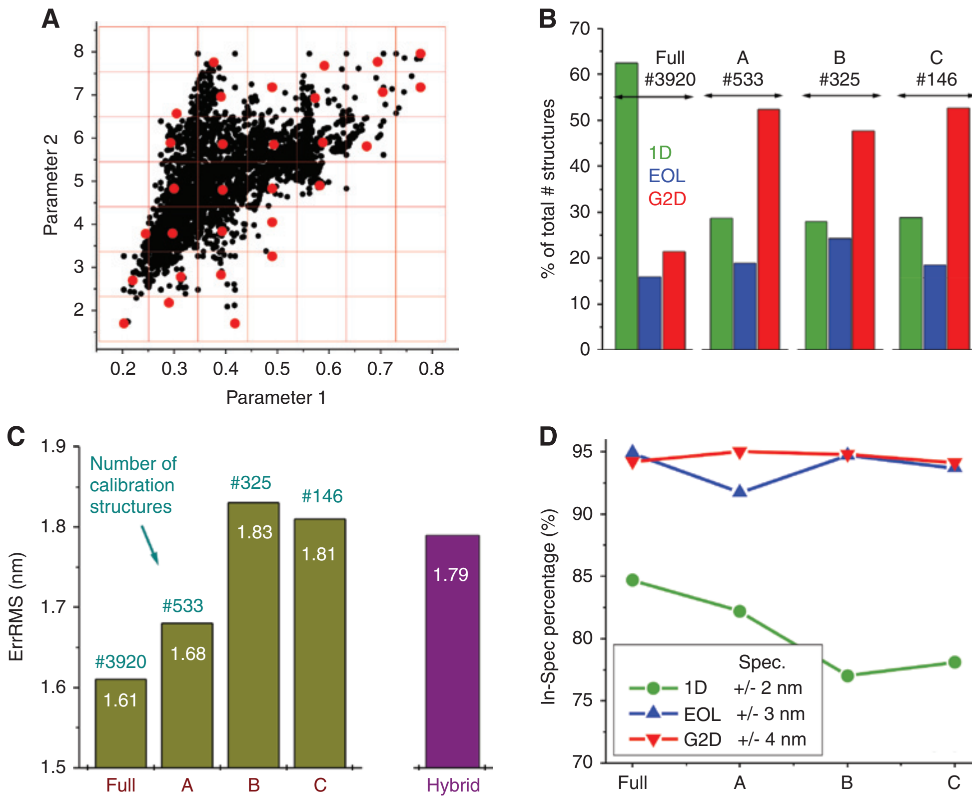

Figure 9: Illustration of the IPS (Image-Parameter Space) concept for a line-printing model case (internally called 'OPC6': PTD, 193i, NA=1.2, min. pitch 100 nm). (A) The concept of IPS sampling, assuming – for simplicity of the drawing – a two-dimensional IPS space. The black points represent all candidate structures; the red points are the calibration structures selected by picking out one structure from each grid cell. (B) Four calibration sets were generated, starting from a very large set of 4000 structures and then using the IPS sampling technique to generate three smaller selections, labeled A to C. This plot shows the relative content of each of these structure sets. (C) ErrRMS using the full 4000 structure set as verification set for the models generated with the four calibration sets. Even the smallest structure sets offer a verification ErrRMS that is only ~10% larger than the one from the original model. (The Hybrid model is a model based on mixed use of CDs and SEM contours; see Section 2.2.3.) (D) In-Spec Percentage values obtained with the four models.

models of both Figures 8 and 9 and assume that bidirectional calibration structures are no longer needed (for modeling unidirectional layers), the size of the model-calibration data sets is reduced by more than 50%, which is quite considerable. The remaining structures would then still support a wide variety of structure sizes and pitches. Unidirectional layers, however, very often contain a limited set of pitches and sizes only, such that the amount of calibration structures then becomes very small indeed.

2.2.2 CD data quality

Even if one has succeeded in defining the perfect calibration data set, the resist model that is calibrated using it can only be 'as good as' the quality of the measured CD data. The necessity of good quality data is so obvious that this subject does not often get explicit attention in the literature. Especially with model calibration, experience teaches that even a few percent of 'bad data' can compromise the accuracy of the entire resist model. So we will use this paragraph to discuss some of the measures we have taken in our own work to avoid 'bad data'.

We find working with design-based metrology [68] indispensable. It allows fast and efficient generation of measurement recipes, especially if one builds some additional software on top of it (see also Section 4.1), in which the measurement position and settings are predefined for each structure type of interest. This ensures that no mix between different measurement conditions occurs, and complex measurement jobs of previously unmeasured structures can be created in minutes. Even more important is that it allows to inspect and (often) correct each individual measurement on an off-line station, after the SEM recipe has been executed (see illustration in Figure 10). Small mistakes of, e.g., the placement of the measurement boxes can be repaired, and even though this is tedious and time consuming, we find this well worth the

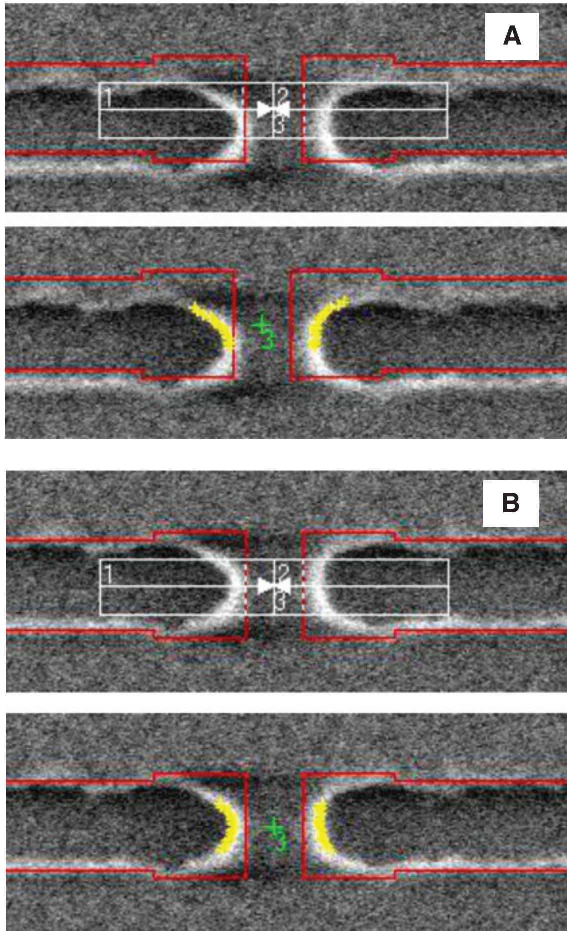

Figure 10: (A) Example of a misalignment between the measured SEM image and the design pattern (red lines), as a result of which the measurement boxes (white) and the edge detection are also off. (B) On the off-line Design Gauge Analyzer station, such errors can be (manually) repaired, thus, ensuring the quality of all CD data.

effort if something precious as the calibration of an OPC model is at stake. (We adore automation, but not everything can or should be automated).

Measurement statistics is the next concern [69]. Measurement of 2D structures (e.g., line ends) has a higher noise level and, therefore, requires a larger number of individual measurements to average. This requirement can be taken into account, to some extent, when designing the calibration structures on the test mask, by providing large-enough test structures such that multiple measurements can be made, one close to the other. This makes a separate pattern recognition and image refocus unnecessary for each individual measurement and, hence, makes doing more measurements less expensive without loss of accuracy.

Data filtering is equally important, but needs to be done with care. We typically use a histogram filter that verifies that the measured data constitute a single distribution and only delete data points that clearly do not fit into that distribution. When the distribution itself looks suspicious or when it is too wide (in comparison with similar structures), the entire structure is rejected from the calibration set.

We also check the exposure latitude for all structures: structures with an exposure latitude below a selected threshold are again rejected.

Finally, we visually inspect SEM images for all structures, to eliminate those that visibly do not print well, but have managed to survive all the automatic data filtering.

Apart from the random noise, systematic effects that affect the measured CDs are an increasing concern. The choice of the measurement algorithm is one; e-beam-induced resist shrinkage is another [70, 71]. Although it is possible to estimate the amount of shrinkage and, hence, correct the data for it (by measuring at multiple e-beam doses and extrapolating the data to dose zero), it is not clear to what extent this is being done in practice. To our knowledge, the impact of SEM shrinkage on OPC model accuracy has not been discussed in the literature. A thorough discussion of the subject of SEM-induced resist shrink is beyond the scope of this paper; for a recent discussion, see [72] and the references within. It may become necessary at some point to develop SEM-induced resist shrinkage models for OPC-modeling purposes.

2.2.3 CD vs. contour metrology

A CD-SEM image (especially of 2D structures) would appear to have a so much richer content than the few CD measurements that one usually extracts from it. This explains the attractiveness to use SEM image contours, representing the complete shape of (part of) the image for OPC modeling purposes, as an alternative to the traditional CD-based model calibration approach [73]. Expected benefits could be that less SEM time is needed, and/or more accurate models might be obtained. Between this simple concept and its realization, there is, however, an impressive series of hurdles that must somehow be dealt with. Among these are potential SEM image distortion, orientation-dependent edge quality differences, resist shrinkage, alignment of measured contours, statistics (i.e., contour ‘averaging’) and elimination of ‘low-quality’ data (fliers), selecting representative structures (again) while avoiding redundancy in the image content, matching of contour data with CD data, etc. For a more detailed discussion of some of these challenges, see, e.g., references [74, 75]. Progress is being made, however, and

judging from recent literature (e.g., references [75, 76]) the use of contours for OPC modeling purposes is gradually gaining interest and acceptance, even though it has not (yet?) become mainstream.

We will illustrate this by examples from our own experience in this field (also in Sections 3 and 4). Calibrating a model using contours relies on the measured Edge-Placement Errors (EPE) between the measured contours and the design target (or an ideal contour). This implies that the absolute position of the measured contour must be precisely known: any placement error goes immediately into the EPEs and compromises a good model quality. We have shown before that the contours delivered by the SEM vendor occasionally have a residual placement error [77]. Today, using more modern SEM tools, the situation has improved, but we still occasionally find residual alignment errors >1 nm (rising up to a few nm in more rare cases; Figure 11 shows a case where a relatively large alignment correction was needed), which is why we are still using our in-house software to re-align each measured contour, and – as a next step in the flow – calculate an averaged contour from a statistical set of contours, obtained from identical structures.

For each individual contour as well as for the averaged contour, one can quantify the deviation from the predicted contour using what we call the Contour_RMS [77]:

$$\text{Contour_RMS} \equiv \left(\Sigma_j D_j^2 / N \right)^{1/2} \tag{7}$$

D_j are the point-to-point distances between the measured and the predicted contours, measured along the entire contour or along a selected clip.

Figure 11 also illustrates contour averaging and shows how a measured PV band can then be generated from averaged contours at multiple FE conditions (see Figure 11B).

Reference [75] describes an (also) in-house tool for analyzing (and, hence, selecting) potential calibration structures form an image-space parameter perspective. Reference [76] shows cases where the use of contours leads to increased model accuracy. It also applies recent technology [70, 71] for correcting measured contours for SEM shrinkage.

As improvements and experience builds up, it can be expected that SEM contours will increasingly find their way into the model calibration field, as they already do when it comes to model verification [78–80], as we shall see in Section 3.

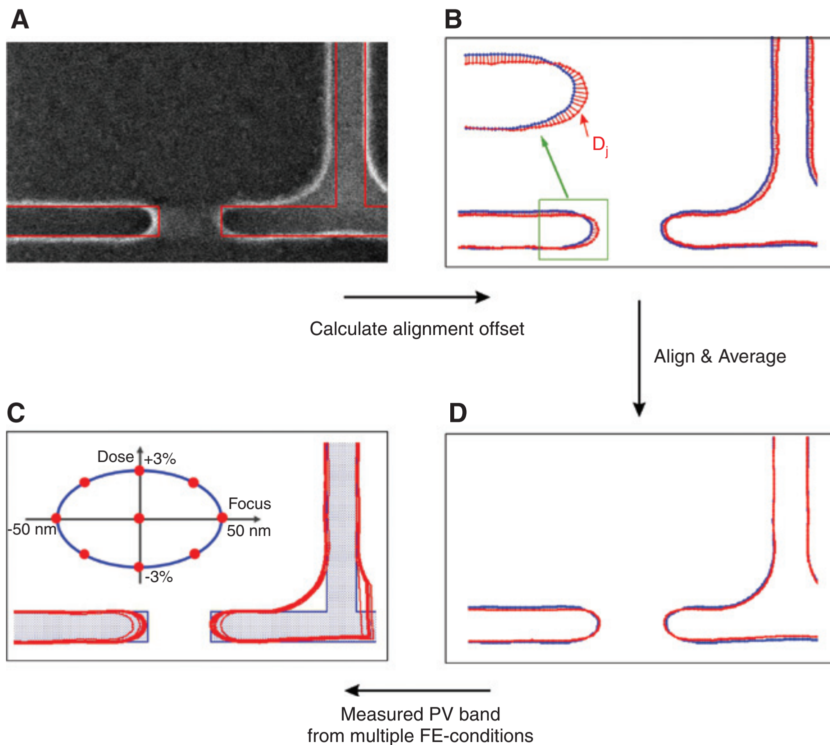

Figure 11: Example of SEM image contour use. (A) Measured SEM image (after etch) in its original alignment to the mask gds; (B) extracted SEM contour (red) together with a predicted contour (blue) that is used for re-aligning the measured contour. (C) Result of the averaging of 15 re-aligned contours (red) overlayed with the predicted contour (blue); (D) measured PV band, obtained from averaged contours at multiple FE conditions (ellipse 'touching' ±50 nm defocus ±3% dose offset).

2.3 Additional models

Apart from the optical and resist models described in the previous section, the OPC engine can/must be fed with additional input. An important one is the specification on how to create Sub-Resolution Assist Features (SRAFs or simply Assists). These structures are put on the mask and should not print themselves but, when properly placed and sized, should make printing structures ('main features') more robust against process variations. In this era of limited DOF, SRAFs are now used, in particular, to increase DOF. The user will need to decide how these SRAFs should be placed and needs to take measures against assist printing. These are the issues of Assist placement and printing.

We already mentioned the possibility of also specifying a model that describes mask imperfections, the mask processing correction model (MPC). This model is optional: if no MPC is specified, the OPC engine will assume that the mask is perfect. Mask errors will then be 'absorbed' into the resist model, which does, in principle, compromise the predictive power of the OPC model. Nevertheless, MPC models does not seem to be commonly used.

Patterning does not end after litho: usually an etch step follows, and during this etch step, the patterned dimensions change again. So the OPC software should also include an 'etch model', i.e., a model describing how the pattern in resist (after litho) will be transformed into a pattern after etch, such that it is this final, after-etch pattern that is close to the design-target pattern.

Already in Figure 2, we showed an overview of the models that are being used (or can be used: not all are mandatory) in the OPC process. We will now discuss the assist- and etch-related models. At the end of the section, we will also add a note about stochastic effects.

2.3.1 Assist placement: model or rule

A first approach to assist placement is to use a 'rule-based' placement approach. Such a rule specifies how many assists, of which size and at which positions have to be inserted in-between two main features in the target design, depending on the space between these main target design features. For line- or trench-like patterning layers, such a rule is not very difficult to make and is usually based on the observed printability of a large number of test structures. We have made such rules in the following way, targeting at i) an optimized DOF for all structures, and ii) guaranteed assist nonprinting.

– Print a test mask that contains a large number of 1D structures (lines or spaces at a variety of pitches and with varying mask-line- or mask-trench widths) on which a (large) number of trial assist rules have been placed. Mark all the cases for which the assists are not printing, while the main feature prints approximately at the intended target CD. Do this inspection at nominal focus, as well as at under- or overexposed dose condition (for a bright-field mask/NTD combination or a dark-field mask/PTD combination, respectively). Cases with printing assists are eliminated from the list. The fact that assist nonprinting is ensured from printed wafers and not from some model is an important advantage of this procedure.
– Run the remaining assist-rule cases through a rigorous simulator (preferably with a calibrated resist model) and determine for every pitch which assist rule provides the largest DOF.
– Measure the best performing rule cases from the previous step on wafer, this time through focus, to make sure the intended DOF is reached. (If not, try more rules.)

The assist-placement rules generated in this way can be very successful in many cases. The requirement that assists should not print, can usually be met, as it is built into the rule-generation procedure and is obtained from printed wafers. Also, the assist placement is usually very good for patterns with regularly shaped spaces (between main features), but when the space is nonregular, the rule may not lead to a unique solution, but allows different implementations. The assist placement solution that is adopted may not be the real optimum, which can result in weak through-focus performance. Also, generating rules in this way is time consuming.

This is why people have tried to devise *models* for assist placement, the goal of which is to ascertain that also for complex main feature configurations, an adequate assist coverage is obtained. Many approaches have been tried, and the literature on this topic is quite extensive. Confining this discussion to the more recent history, the following approaches can be mentioned.

1) Generation of some pixel map that suggests assist locations. This concept first defines an objective function (defined on all xy positions where no main feature exists) that assigns a value to the usefulness of placing an assist at position xy with respect to the printability of the main features in its neighborhood. What this objective function looks like and which main feature printability metric is to be optimized does not need to be the same for all who have used

this type of approach (and is often not publicly available). But this map is then used to insert assists at the most promising positions, after which a step to simplify the shape of the resulting assists is inserted, as well as a check that the assists will not print. Examples of such an approach can be found in references [81–83], where this pixel map is called the 'SRAF Guidance Map', the 'Cost Covariance Field' and the 'Gradient Map', respectively. Each of these approaches has demonstrated some level of success, and is relatively fast. Whether they succeed in making most or all target structures sufficiently robust, depends, of course, also on the soundness of the objective function they assume and the reliability with which they can estimate assist printability.

2) A different type of approach is called Inverse Lithography Technology (ILT). Originally, the ILT concept aimed at inverting Eq. (1), i.e., instead of calculating the image from a given mask function, it tries to calculate the mask function that leads to the required (or 'ideal') image [84, 85]. Note that this ILT approach does not only generate assist, but also the ideal shape of the main features, so it performs the assist placement and OPC step simultaneously. The difficulty, however, is that the ideal mask solution that results from such an inverted calculation usually has a continuously varying transmission and phase and would, hence, not be manufacturable. This means that this ILT-generated mask at least needs to be simplified to be consistent with the mask manufacturability requirements. This simplification can then be expected to lead to some loss in performance.

3) A more practical approach to the ILT concept is the pixel-inversion approach: for pixels outside the main features, the effect of inverting the mask from the design target transmission to the opposite (i.e., assist) transmission is evaluated, from which a decision to convert this pixel to the assist tonality or not is taken. Also this technique leads to complex-shaped assists, such that again, a shape simplification step needs to be added, as well as an assist-printability check. It is this version of ILT that is being used today.

ILT-type approaches are expensive from a calculation time as well as a mask-making point of view. Instead of applying them to the entire mask, they are typically used in a more 'focused' way, e.g., for generating better assist-placement rules, or as local fixes of hotspots on layouts that were OPCed with more traditional approaches (see, e.g., [86]).

Each of the assist-generating methods that are being used today has reported successes, but each is also based on some assumptions, either in the underlying concepts (e.g., cost functions) or in their technical implementation. So probably, none of them can guarantee success in all cases, and further progress in this field can be expected.

We conclude this section with two illustrations of the complexity of assist placement. Figure 12 shows two 20-nm node Metal1 patterns, one in which the rule-based approach was found to be better than the model-based approaches we tried, and one where a model-based approach is desirable. For contact or dot layers, the problem of assist placement is even more complex.

2.3.2 Assist printability model

Any model-based assist placement engine must also try to make sure that the assists will not print. Resist models calibrated from main features, as discussed in Section 2.2, are generally not reliable in making predictions about assist printing: predicting the behavior of assists would be a significant extrapolation from the structures that were used to calibrate the model. Also, using a simple threshold

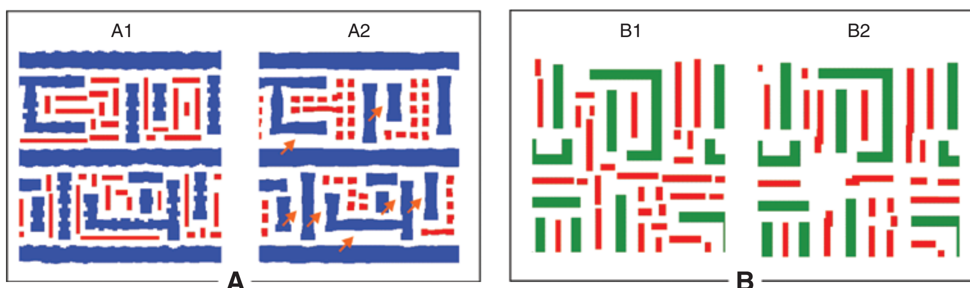

Figure 12: Examples of different assist-placement scenarios. (A) Rule-based and model-based assist placement example, on the same M1A clip. In this particular example, the model-based case leaves some of the narrow spaces without assist (spaces indicated with arrow). As we have showed elsewhere [93], this limits the DOF in those areas; see also Figure 16. (B) Two implementations of the same basic rule for assist placement on the same clip of M1B (before OPC of the main features). Because the spaces between the target features are less regularly spaced, the actual assist arrangement can be quite different according to the details of the assist-placement setup.

model on the intensity at the assist positions is too simple an approach to be reliable enough. This means that a dedicated assist-printing model would have to be calibrated on measured wafer data from printing and nonprinting assists.

One difficulty with calibrating such a model is that the input data cannot be CD data: the most important condition that requires good prediction is the *onset* of assist printability. In that case, one cannot assign any reliable CD value to the assist. That is why the input data of such a model is often defined in a binary way, e.g., assigning a value 0 to nonprinting cases and 1 to printing cases (even though assigning a continuously varying printability value is possible as well, and should leads to easier modeling). Also, the onset of assist printing starts happening either at the bottom or at the top of the resist film (in clear-field or dark-field imaging, respectively). That is why it makes most sense to base assist-printability predictions on image intensities in a plane either close to the bottom or close to the top of the resist, i.e., in a different plane as the one used for the main feature model. This is the approach that was taken in references [87, 88]. The authors report an improved, though not yet perfect, assist printability with this type of approach.

A maybe better approach could come from R3D type models (see Section 2.1.3): in reference [30], a good reproduction of assist printing was obtained for the cases shown with an R3D compact model that was calibrated from rigorously simulated resist profiles. The real challenge of R3D models is – as we discussed above – how to calibrate them from wafer data.

2.3.3 Etch model

Patterning does not end after litho, it is usually followed by an etch step. Etch usually transforms the shape of the printed patterns in a way that is slightly different for each pattern, i.e., it contributes to the final proximity effect. Figure 13 illustrates this point for the case of our 20-nm node Metal1 etch process. The etch process was intended to shrink the litho dimensions – in this case, with 15 nm – an approach that is often taken to achieve small after-etch dimensions while keeping the litho process windows high. Figure 13 shows that the actual difference between the after-litho and after-etch CDs, which we call the 'etch bias', is actually structure dependent. Formulated in terms of pattern edge placement: not all edges shift with the same amount during the etch process.

For accurate final pattern control, this effect of etch needs to be included in the correction process. This is done in the way that was already shown in Figure 1: once the dependencies as in Figure 13 are known, the design target is converted into a litho target, displacing each edge (fragment) according to the effects that etch is expected to have on this edge. After that, the combined optical-resist model further converts this litho target into the actual OPCed mask.

The effect of etch can be described by a rule table or by a model. Although model forms to describe etch effects have been known for a long time (see, e.g., references [89, 90], until recently, there seems to have been a preference in the industry for a rule-based etch correction (for a discussion on this point, see reference [91]). One of the reasons for that was that etch effects have a longer range than optical and resist effects, which puts additional burden on the calculation time. But today, these computational issues can be efficiently dealt with, and with the ever increasing need for more accurate pattern control, a model-based approach is increasingly required. References [92, 93] show examples, where a model has been built on an etch-bias data set of which Figure 13 shows a

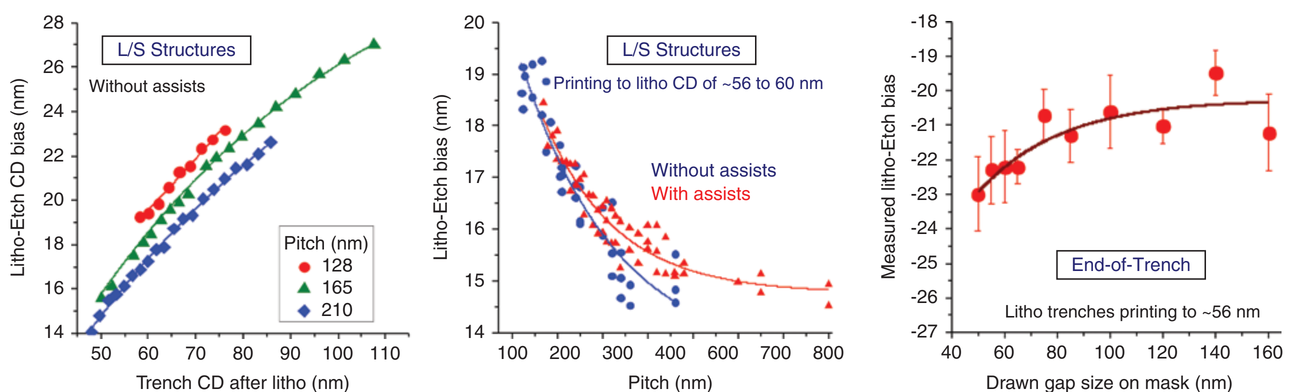

Figure 13: Examples of measured etch-bias dependencies for the case on an NTD resist on a SOC/SOG stack (case of our Metal1 LELE process). CD shrink during etch targeted at 15 nm (at ~isolated structures). The lines in the plots are exponential-fit curves through the data.

selection and afterward applied to 2D structures. We have used this type of modeling in the staged approach that is shown in Figure 1A for a successful combination of litho-etch proximity correction. So it seems that today's etch-modeling infrastructure is well capable of doing the job it was designed for.

2.3.4 Stochastic effects

Also stochastic effects have an impact on the shape and especially the variability of printed patterns. Traditional lithographic modeling has always implicitly assumed that the resist is a continuum, but matter, as well as light, consist of particles, and at the dimensions, we are working today, this starts showing. LWR/LER are well-known manifestations of this, but close to the printability limit, stochastic effects can lead to random printing failures ('stochastic failures'). Randomly missing contacts are just one example of this.

Today, we have only a partial understanding of all the mechanisms behind stochastic effects [1, 94], and although some of its contributions are incorporated in rigorous simulators, the current model description is not complete. It is, in any case, not possible to *correct* for stochastic variability or failure with OPC as it can never be predicted where exactly failures will occur. The only thing that OPC engines could potentially do is to flag structures that are expected to be especially stochastic-effect sensitive. This would require a model that predicts stochastic-effect sensitivity, something that today does not exist. As stochastic effects may well turn out to become the ultimate limiters to lithography, we believe it is important that such models are developed.

3 OPC correction and computational verification

Once the OPC models are available, the actual OPC correction can be applied to the target design, from which the OPCed mask will be the result. Before having this mask manufactured, however, additional computational tests are run to verify that the predicted printing contours meet the spec for all the process conditions at which this is required. This inspection usually leads to the identification of locations where these specs are not met, so-called 'hotspots', leading to modifications in either the OPC recipe, the OPC models or the design target layout itself, to remove these hotspots, or the OPC engineer can choose to waiver these local defects. This post-OPC computational step is called the 'OPC Verification' step.

In this section, we will touch upon this phase of the flow. We, however, have chosen to not attempt a review of all the art and subtleties that are involved in these critical steps, one reason being that such a discussion would have to zoom in on some of the EDA-specific techniques and tricks, some of which are not public. Also, this field is so broad that a complete review cannot be fitted into one paper, together with the other topics we try to cover. So in the rest of this section, we will highlight only a few selected topics and show some examples.

3.1 Applying the OPC correction to the design target

The general OPC correction flow was already shown in Figure 1, and a few specific topics such as the problem of assist generation and the use of inverse lithography have already been mentioned in the previous section. In general, the OPC step has to decide how to divide the edges on the target pattern into separate fragments, how to iteratively move these to bring the predicted printed contours gradually closer to the target shape, which objective function (or cost function) is to be used, and which process conditions are to be included in this process.

The ambitions and power of this process have increased tremendously, as automated full-chip OPC was first introduced around the mid 1990s [3–5]. Without the aim of being complete, we would like to mention here:

- 'Matrix OPC', being a technique where the influence of neighboring fragments on each other is being taken into account, rather than moving each fragment independently of each other (which is the more traditional approach). This technique, introduced in the early 2000s, has now come of age [95], thanks to advances in mathematical techniques and computer power.
- 'Process Window OPC' (PW-OPC), already mentioned before, which minimizes the deviation between the entire PV band to the target pattern, rather than the contour for the nominal printing condition only [96]
- 'Interlayer aware OPC', which means that the OPC of a certain layer takes into account the existence of another layer [96]. This can take various forms. One example is coupling the OPC for a metal layer to a connecting via layer. The target here is to maximally 'defend' the overlap area between the two layers. Another example is related to double patterning [97, 98], where the stitching overlap between two split layers can be optimized or where the expected shape

of one split is taken into account while doing OPC on another, to avoid, e.g., bridging (or too narrow spaces) between both layers.

- ‘Model-based fragmentation’ [99]. For a good control of bridging, necking, ringing, line-end control, etc., it is very important to apply optimized fragmentation settings, but how to define good rules for fragmentation is very much dependent on the skill of the OPC engineer. Therefore, model-based approaches to fragmentation have been developed. These not only choose the best fragmentation, depending on the shape of the predicted contour, but also continuously adapt it during the OPC iteration. This leads to easier OPC recipe setup and also improves the final result.
- ‘Inverse Lithography’ (ILT). We already mentioned ILT in the context of assist placement, but as it defines the shapes of the main structures as well as the assist features, it offers, of course, a complete OPC solution. Traditional OPC operates by moving fragments of the pattern edges. ILT is different in that it divides the mask into pixels and then decides for each individual pixel which mask tonality it should best have (after which clean-up and shape simplification steps follow) [84, 86]. ILT is considered to offer the best solution [100], but is expensive in computation time and mask complexity. In the next section, we will see that it is now often used to locally fix hotspots, while staying with traditional OPC everywhere else.

3.2 OPC verification and hotspot detection and fix

Before taping out a mask, the OPC solution is first checked over its entire area in the ‘OPC-Verification’ step. This is a computational step in which the predicted shapes are calculated at a number of focus-, dose- and mask-error conditions, from which a PV band is obtained [101, 102]. This result is then automatically checked against the requested specification for a number of pattering-quality metrics that the user can define. These metrics are designed to flag potential failures or locations with a small process window, the so-called hotspots. A few typical detectors that would be used in, e.g., a metal layer patterning are shown in Figure 14. Each EDA company has a different way of visualizing the result; Figure 14 shows an example in which the number of a specific hotspot type found for each process condition is graphically represented in a focus-dose matrix.

The observed hotspots are then usually classified into different types, and the EDA software offers an easy way for the OPC engineer to locate and visually inspect each individual hotspot.

If too many hotspots are observed, the next step is to try and repair them. Traditionally, this is being done by tuning the OPC recipe, e.g., by changing the fragmentation rules, the assist placements, or some other parameter in the recipe. Today, more advanced techniques are also available. Without the ambition to be complete, we would like to mention:

- Local hotspot fix with ILT [86, 100, 103]. The Verification flow can now be set up in such a way that the OPC at the hotspot positions is locally redone using ILT. The ILT solution is then re-inserted into the original OPC solution of its environment. This means that the OPC does not have to be redone for the entire mask (as in the traditional approach) but is repaired locally. So both from a run-time as from an accuracy point of view, this local ILT fix is the better solution.
- Hot Spot Fix by local layout modification (‘retargeting’). If allowed by the functionality of the structure, hotspots can sometimes be removed by changing the size, shape, or position of one or more polygons in the original layout. Doing that manually for all hotspots is unpractical, so recently, ways to automate this process have been proposed [104, 105].
- Even if all hotspots can be eliminated using the above-described methods, the question still remains whether the Verification tool detected all hotspots (and whether it also perhaps found ‘false’ hotspots).

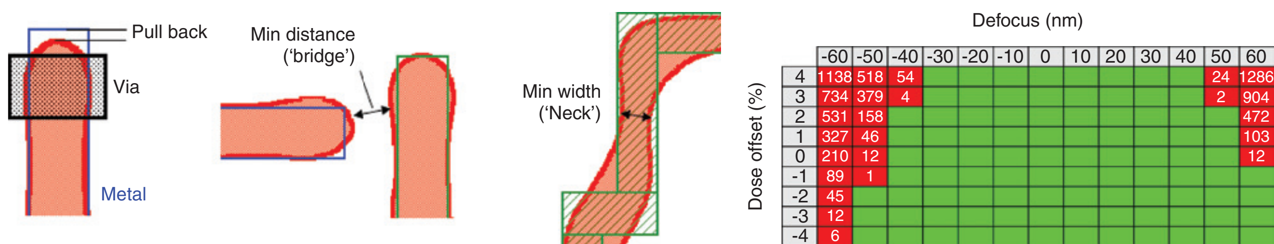

Figure 14: Left: Examples of OPC-verification metrics for the case of a metal layer: line-end-pullback, bridge, and neck. Right: representation of the #hotspots (of a certain type) per PW condition. Green means that no hotspots (of a certain type) were found at this FE condition.

Verification is based on models, and no model is perfect. We have seen earlier that the calibrated resist model is based on intensities in a single plane inside the resist only, very often close to the middle of the resist. Hotspots sometimes occur rather near the top or the bottom of the resist. Therefore, separate models for hotspot detection have been built with an adapted vertical position of the image plane. We have met such models earlier when discussing the issue of assist printability prediction, and a similar approach has been also reported for hotspot prediction [87] (see also Figure 16 in Section 4).

– Another approach for increasing the predictive accuracy during the OPC Verification phase is the 'Embedded Rigorous simulations' for selected critical areas [106]. This means that a rigorous simulator, like S-Litho, is automatically called by the OPC Verification engine, to make a full three-dimensional resist simulation in (small) areas where, e.g., a hotspot is suspected or where a very critical pattern is located. The results are reported back to the OPC engine that then decides whether or not a problem does exist in these areas.

3.3 Geometry-based structure screening: recognition of known hotspot clips

The computational verification-repair iteration described in the previous section has become an essential step before the final mask tape out, but it is also a rather time-consuming process due to the multiple full-chip calculations it requires. For this reason, people have been looking for ways to do a fast initial screening of the design target itself, trying to detect geometrical patterns that look 'similar' to already known defects [107, 108].

This approach requires that one can start from a library of existing hotspot structures. These can be based on earlier masks from the same technology node or on experience from earlier nodes. As one cannot limit this technique to searching only identical copies of one of the structures in the hotspot library ('pattern matching'), but must try to find also 'similar' structures, the exercise becomes one of first defining and quantifying what similarity means and then performing some kind of pattern recognition search throughout the entire design layout.

Several approaches have been tried [109]. One of these is to represent each clip in the design as a point in a 'structure-parameter space' (a concept not unlike the image-parameter space we used earlier). A 'support vector machine' (SVM) decision system then fits a boundary surface around the location of the known hotspot clips from the library, from which it can then decide for any other structure clip whether it falls in the hotspot or non-hotspot part of this space [109, 110].

As this approach does not do any attempt to calculate the printed shape of any structure, it can never be as accurate as the computational verification of the previous section, but it is much faster. If this geometrical screening step succeeds in identifying a significant amount of the hotspots (say >80%), while not generating too many 'false alarms' (e.g., <100 false hits/mm^2) [108], such that the target layout can be cleaned of these hotspots before sending it to OPC for the first time, less OPC and Verification cycles will be needed to finally arrive at a hotspot-free end result. This, then, speeds up the entire process. To simultaneously achieve all requirements, however, still seems a challenge today.

4 Quality assessment of printed wafer patterns

Once the OPCed mask is available, the on-wafer verification step can start. The computational verification step of the previous section is based on models, and no model is perfect, as we already stressed before. That means that extra failures or hotspots usually show up during inspection of printed wafers. These, then, lead to a modification in one or more steps of the mask-OPC flow of Figure 2 (e.g., tuning of model, OPC settings, layout, …) and, hence, a new iteration of the complete mask-generation cycle.

One of the questions to answer when starting this printed wafer inspection is which structures should be measured. The computational verification phase usually yields a list of structures with a (suspected) weak performance, and these are then obvious candidates for wafer inspection. But that needs to be complemented by additional measurements to screen for potential additional weak- or hotspots that were not predicted by the computational verification. Selecting such extra structures can be based on prior experience or can be done through a statistical screening of the entire layout. Fast defect-review SEM systems have been built specially for this purpose.

4.1 Developing infrastructure for the on-wafer verification step

It will be clear that detailed on-wafer evaluation of suspected hotspots requires measuring a long list of different

structures and, therefore, requires an efficient and automated approach. This is where, once again, the Design-based SEM metrology proves invaluable. Generation of all the required SEM measurement recipes can be largely automated and – as we already mentioned above – the measurement results can be inspected and even corrected on off-line work stations, after the measurement [68, 111].

Figure 15 shows the flow that we often use for this type of work, in which we have added further automation around the Design-based SEM metrology. On top of the SEM-vendor software and hardware (blue field in Figure 15; we work with Hitachi), we have built additional in-house software (green field) to further speed up the generation of the input for these measurements and improve and analyze the output. This additional software performs the following tasks:

– Recipe Director input-file creator software: this package creates and sends the required input files for Recipe Director from (in the most 'advanced' version) nothing more than a list of gauge (i.e., measurement structure) names, and a manual selection of the dies – and possibly sub-dies – that need to be measured. This requires that for each mask (i.e., gds) of interest, these gauge names have been previously identified and linked to the correct coordinates in the gds and to predefined SEM metrology settings. These structure identifications we derive directly from the scripts that were used to generate the gds itself, while the exposure maps are read from the scanner lot reports. Once this infrastructure is set up for a given mask, very complex measurement jobs can be generated and available on the CD-SEM within minutes, passing through Hitachi's Recipe Director software as an intermediate step.
– Once the measurements are done, we import the results in the Design Gauge Analyzer software, where the data can be inspected and cleaned, as already mentioned. When requested, this software outputs not only CDs but also SEM image contours. Such contours can be extremely useful in the verification stage, as they allow Edge-Placement Error (EPE) metrology from the wafer measurements [111]; we will give an example in the next section.
– These measurement outputs, CDs and contours, are then further treated with a second set of in-house data-analysis packages. Especially when EPE-analyses are intended, it is essential that the measured contours are perfectly aligned to the original design-intent layer. We, therefore, use our own software to perform re-alignment of each individual contour and also do an averaging of contours measured from nominally identical structures on the wafer, for statistical purposes [77]. An example is already given in Figure 11.
– The final analysis of these averaged contours (EPE analysis, PV-band calculation, detection of necking or bridging cases, etc.) is then finally done with standard EDA software (or other *ad hoc* in-house tools).

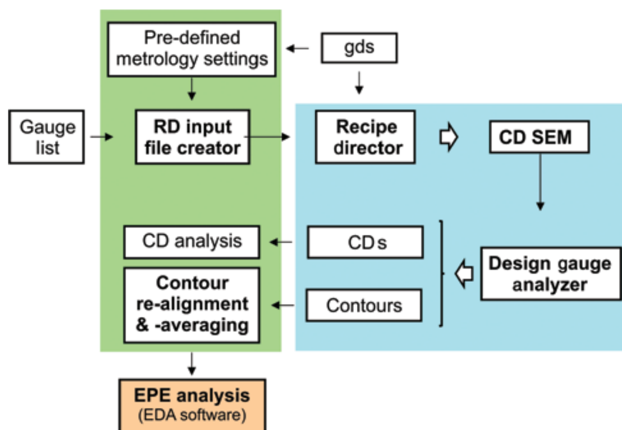

Figure 15: In-house flow for on-wafer pattern-printing verification. Components in the blue area are provided by the SEM vendor (Hitachi); components in the green shade area are out own in-house tools. Together, they make fast and efficient recipe creation and data analysis tool box.

Each fab will have its own version of the flow shown in Figure 15; often more complicated, as it may also automate the feedback to the OPC flow itself, or include monitoring of hotspot cases (see, e.g., [112]).

4.2 A few wafer-verification examples

In this section, we will look at a few examples from the on-wafer verification phase, chosen to illustrate some of the claims in this paper, without trying to be complete.

Figure 16 shows a first example of such an after-litho wafer inspection. It concerns two OPC versions (the two cases shown in Figure 12A called A1 and A2, differing mainly in assist placement) of the same target structure [93]. Around best focus, there is little difference, but out-of-focus, we see trench scumming occurring in the A1 case but not in the A2 case. This difference was not predicted in by the standard OPC model: these predicted contours are shown in the right graphs of Figure 16 as the green lines. Using a modified OPC model that evaluates the image intensity in a plane closer to the bottom of the resist than the standard model does, however, yields the red line contour prediction. These are quite close now to the observed wafer performance. So, this case also illustrates

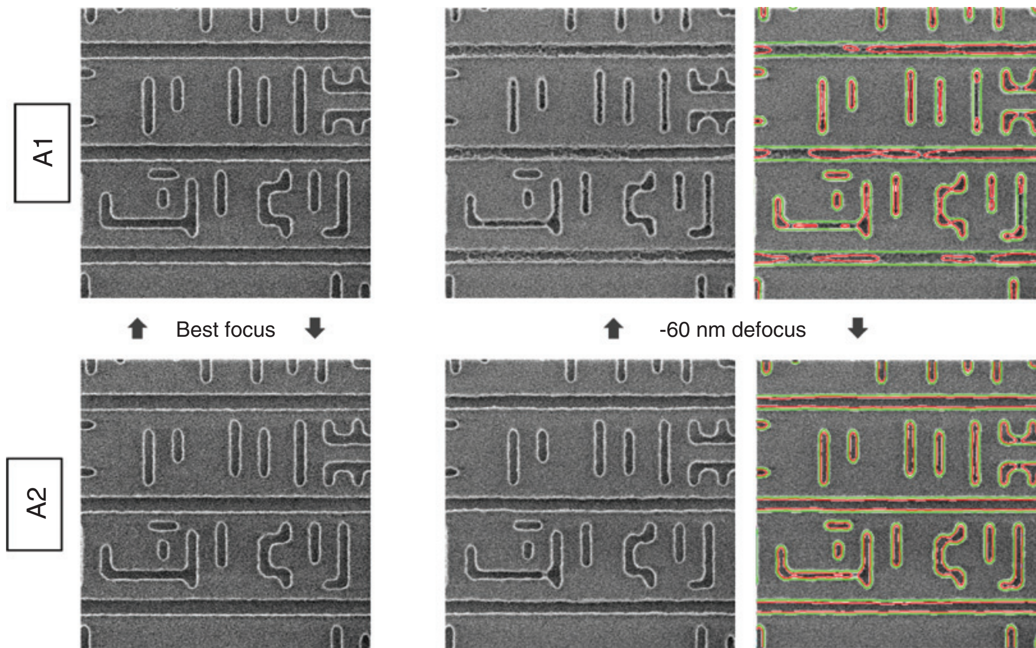

Figure 16: Measured printing performance (in resist) of the two OPC versions A1 and A2, shown in Figure 12A. Top-down SEM images at best focus (left) and at a -60 nm defocus (middle). The defocus images are shown a second time (right), now overlayed with predicted contours from the standard OPC model, using image intensities from a z-plane in the middle of the resist (green lines) and predicted contours from a second model, using image intensities from a z-plane closer to the bottom of the resist (red lines). The bottom-plane contours predict the trench scumming much better. This is an example of the type of hotspot model that was discussed in Section 2.1.3.

the usefulness of an R3D model for hotspot prediction, rather than using the two separate models used here.

A second example illustrates the usefulness (even necessity) of measured contours and EPE metrology: see Figure 17. It shows a measured PV band for the example we started this paper within Figure 1, deduced from the measured contours after etch, when both the first and second litho exposures were done as an FEM. The result is quite close to target and shows good robustness against FE variations. Interesting exceptions, however, are the tip-line configurations marked as 1 and 2 on the figure, where we see a significant line-end pull back (of almost 10 nm). At the same time, the opposing line 'pushes out' from its target position, such that the tip-line CD itself turns out to be very close to the target value (which is an aggressive 70 nm in this case; this configuration was actually put on the mask to test whether 70 nm could be an acceptable design rule). So based on the tip-line CD metrology only, this structure would pass. Complementing the analysis with contour metrology, and the EPE values that can be derived from it, this 70-nm tip-line structure does not pass: the line-end pull back is clearly too large in this case.

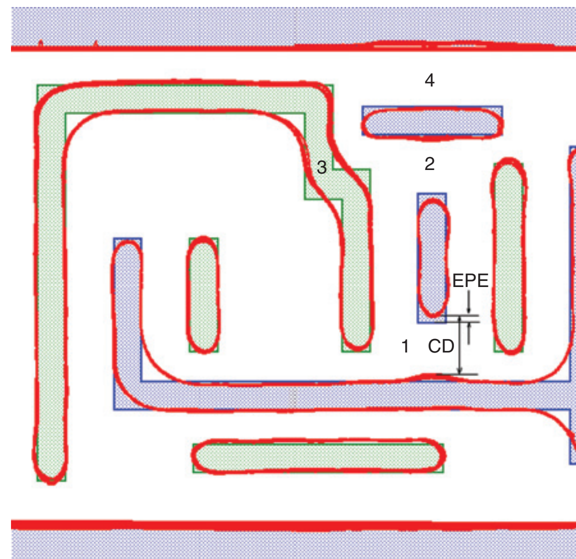

Figure 17: Measured process variation (PV) band example (same FE conditions at litho as in Figure 11) of both M1 splits after the complete LELE process. Note a line-end pull-back at positions 1 and 2, line pinching at 3 and a 'repulsion' of two closely spaced trenches printed in the same split at 4. The line-end/gap case at position 1 shows how CD and EPE metrology can lead to opposite conclusions: the CD value is actually very close to the target value, in spite of a significant line-end pull back. This example illustrates the value of EPE metrology.

Figure 18 shows two types of measured variability band that are different from the 'traditional' PV band. In Figure 18A, we look at the variability in the printing

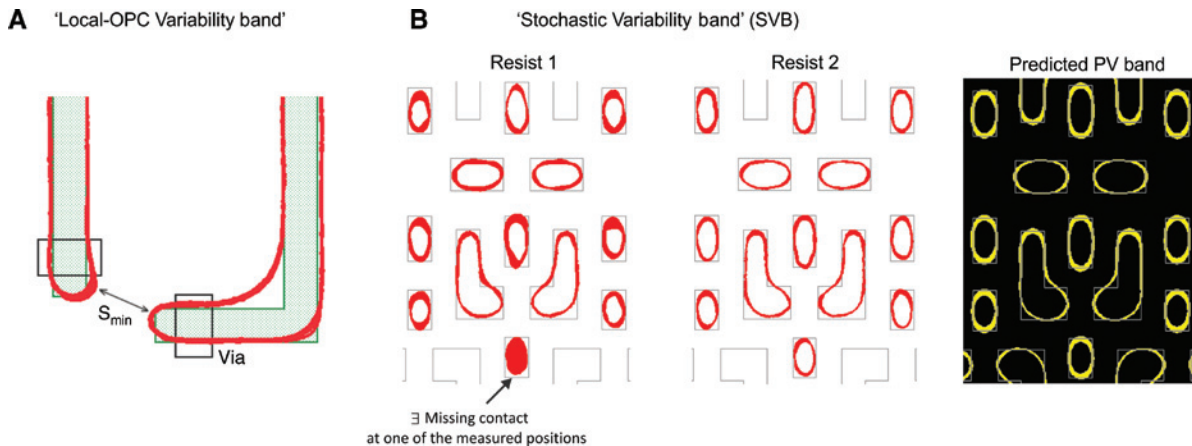

Figure 18: Two examples of variability bands that are different from the traditional PV band. (A) 'Local-OPC Variability band': measured contours (at nominal FE settings, after etch) from the same structure, placed in a different local environment. (B) 'Stochastic-Variability band' or SVB (EUV case): overlay of measured contours measured at different locations (but very close together) in an SRAM array, at nominal FE conditions, for two different resists. The variability for resist 1 is so large that one contact is missing at some of the measured locations. Also shown is the 'traditional' computational PV band prediction. It shows that the stochastic effect-induced variability is of the same order of magnitude as the FE-induced variability throughout the entire process window.

performance of a certain structure at nominal FE conditions caused by changes in its environment (i.e., the surrounding cells) and, hence, the local OPC solution. This structure is a tip-to-tip case from two oppositely oriented lines and is again part of a design-rule testing set of structures, where we want to establish what the smallest acceptable drawn corner-to-corner distance is. Each printed line end must offer enough overlap with the connecting Via, but in this case, one must also be careful that no 'short' (i.e., a value of S_{min} that is below what is considered to be the safe distance) occurs. When the two tips are too close, they 'attract' each other, an effect that becomes larger when the target corner-corner distance is smaller.

Figure 18A shows an overlay of 25 measured contours, all measured at nominal FE conditions, but from different locations in the logic block. These positions are, however, close enough (within a few hundred micron) such that the printing variability is not likely to be due to mask error or scanner variation. What does change from position to position – except for random effects – is the local OPC solution as the local environment is different each time. Two OPCed examples for this particular structure were already shown in Figure 12B (where we labeled them as B1 and B2). So the measured variability band in Figure 18 could be called a '*local-OPC-variability band*'. Evaluating the width of this variability band (also at off-nominal FE conditions) helps to judge whether the particular corner-to-corner design rule that was applied in this case is acceptable or not.

We mentioned stochastic effects briefly in Section 2.3.4. It refers to randomness in the image shapes that is essentially caused by the particle nature of matter and light. Even in the absence of any OPC solution difference, identical mask structures, e.g., from a regular contact array will not print identically. There can even be missing contacts. Figure 18B shows an example of such measured variability for the case of a 10-nm-node SRAM pattern printed with EUV. Here, we show an overlay of contours measured at different locations (but very close to each other) in the SRAM array, something which we will call here the '*stochastic-variability band*' (SVB). We show this SVB for two different resists. In one case ('Resist 1') one of the contacts is, indeed, missing at one of the locations, which is not the case for the other resist ('Resist 2'). Even for the best of the two resists, we find that the width of the SVB is about as large as the width of the predicted PV band, illustrating that the variability caused by stochastic effects is becoming an important effect.

5 Design restrictions, layout style, and technology

Most of the patterning examples in this paper relate to a 20-nm node logic application, where – as we have seen – the printability quality of even complex 2D patterns can be ensured using the tools described above. But as every lithography paper underlines: things become tougher

every node. One of the main reasons for this is that the wavelength and the maximum available NA are not changing any more (if, for a moment, we leave EUV aside), which means that the k1 factor will decrease with each node. ArF lithography (λ=193 nm) has been in use since the early 2000s. The introduction of immersion lithography (about 10 years ago) increased the max. NA by a factor 1.44 (refractive index of water), but no further decrease of λ/NA has happened since then and is not to be expected in the future (again with, of course, the exception of EUV). The minimum pitch that can theoretically printed directly with today's 193i tools lies somewhat below 80 nm. So, to print at higher densities than that, double patterning was introduced.

There are essentially two different ways in which double (or multiple) patterning is being used:

– Splitting the patterns of the design layer into two or more separate layers [113].
– Sidewall-spacer double pattering, also called Self-Aligned Double Patterning (SADP, and by extension, SAQP) [114]

Both techniques are now being used in production, but their application puts constraints on the design-target pattern shapes they can be applied to. This implies that the target-pattern layout is no longer independent from the patterning optimization, but needs to be conformal with the limitations (rules) imposed by the intended multi-patterning approach.

In a way, this is not new. Simple pattern shapes (e.g., 1D shapes) have always been more easy to print than more complex pattern shapes (e.g., 2D shapes). The increasing difficulty of printing especially the 2D shapes has often been underlined and has led to the concept of restricted designs (see, e.g., [115, 116]). Restricted design means that the geometrical complexity of the target patterns is restricted and/or that a higher geometrical regularity is imposed on the target layouts. In SRAM, for example, the shapes in (almost) all layers were reduced to rectangles-only, already about a decade ago [117]. Implementing such geometrical regularity in random logic has been a more gradual process. It started in the front-end layers, where from about the 28-nm node (i.e., the last single-patterning node; gate pitch=110 nm) active-area shapes were limited to (nonconnecting) rectangles-only, and Poly became a single-pitch unidirectional layer (with a separate Poly-cut layer). Figure 19A illustrates this type of layout for the simple case of a two-input gate NOR cell (NR2D1). The Metal1 layer still contains 2D shapes. In the example of this simple cell, they are not too complex, but for larger cells, Metal1 becomes more complex, as in the examples we have already seen in this paper.

At the 20-nm node (gate pitch 90 nm, minimum Metal1 pitch 64 nm), the Metal1 pattern needs to be split in two layers, and at the 10-nm node (minimum Metal1 pitch ~48 nm), it needs to be split in three layers. To ensure, however, that the Metal1 design target layer can be split ('decomposed') into two or three layers without any remaining 'coloring conflicts', the layout of each individual cell needs to be already decomposable in two or three splits, i.e., compatibility with double or triple patterning needs to be ensured at the cell-design level [118]. Afterward, the EDA tools used for cell placement also need to observe this compatibility.

But layer splitting does not eliminate all printability problems, and it introduces new ones. Even though the minimum pitch of each split layer will, of course, be within the pitch resolution limit of the exposure tool, the target width of the individual-split patterns keeps decreasing with every node. For example, if the Metal1 width in the 20-nm node (double split) was perhaps ~32 nm, it becomes ~24 nm in the 10-nm node (triple split). Litho

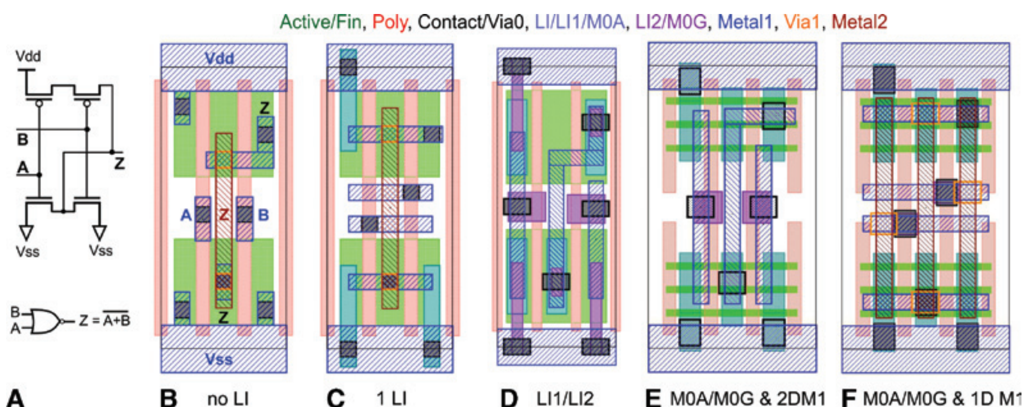

Figure 19: Different layout versions of a NR2D1 Logic cell (nine-track cells; different nodes, but drawn to the same vertical size in this figure).

will not try to print these dimensions directly, it prints to larger CDs, and a post-litho process (often etch) trims the litho CDs down to the target values. But this bias trick cannot be stretched infinitely, so litho needs to try to print as close to its CD limits as possible. Especially then are 2D patterns becoming problematic: in the examples above, we have seen tendencies to pinching and bridging, which are typical for any 2D (Metal1) type pattern.

The 'new' problem introduced by layer splitting is overlay: non-zero overlay errors will occur between the different split layers, and these will 'interact' with the overlay errors of the different splits of the layers connecting to it (e.g., via the layers connecting to the metal layers), leading to a rapidly increasing number of layer-to-layer overlay-error combinations, as the number of split layers increases. This will affect variability and, ultimately, device performance.

SADP is a partial answer to this overlay problem: due to the self-aligned nature of the spacer process, the scanner alignment and overlay performance does not enter any more into the pattern-placement-error budget of the SADPed structures. (Pattern placement errors can still arise due to CD errors of the SADP-core pattern though). But SADP leads to difficulties and printing artifacts when applied to 2D patterns [119], and today, nobody considers SADP for 2D-type Metal1 applications any more.

So already for a long time have people predicted that also the back-end layer shapes would have to be made more regular, more simple (see, e.g., [115, 120]: 'simplify to survive'), the unidirectional pattern being the ultimate dream.

But how to achieve that is not straightforward. Let us illustrate why, taking again the simple NR2D1 cell of Figure 19 as an example. Figure 19A shows the connections to active and gate that need to be made. Among others, both input gates need to be connected, as well as a connection between two output source/drain (active) points, crossing the B- gate. It is clear that this cannot be done by one unidirectional layer. Figure 19B shows the 'old' 2D solution. Simply splitting this 2D solution in two unidirectional components is not an option (especially not in more complex cells than the NR2D1 example we are considering here): this would require stitching between horizontally and vertically oriented lines, and stitching requires stitching overlap, for which, in many cases, there is no room. So it seems Metal1 was doomed to remain 2D (as many people, indeed, believed, until recently).

The way out of this dilemma came with the introduction of additional process layers, and more specifically started with the introduction of Local Interconnect (LI). LI is inserted between the process layers that were traditionally called front-end and back-end of line layers (i.e., between Gate and Contact) and is, therefore, sometimes called the 'middle of line' or MOL. (The different ways of contacting Active and Gate that are used in the cell layout solutions of Figure 19 are further illustrated in Figure 20.)

An initial idea [121, 122] on how such an extra LI layer could drastically simplify the Metal1 patterns was to use LI for making active-active connections and to connect active to the power rails (Vss and Vdd), as shown in Figures 19C and 20B. (LI is only used for the power rail connections in this simple example cell). Metal1 can then become unidirectional (oriented 'horizontally', i.e., perpendicular to Poly), not only in the case of the simple NR2D1 cell, but for all logic cells, including the very complex ones (e.g., the Flip Flops). Such unidirectional Metal1 patterns can be printed in a single pattering [123] step down to pitches of ~80 nm (and gap sizes of ~70 nm), so it would seem a perfect solution. Once applied, however, this LI scenario was found to lead to reliability issues: the probability for an electrical 'leak' between the LI and the transistor junction, buried under the surface at the active-STI boundary turned out to be too large. This illustrates that any connection scheme that looks good on the layouter's table still has to prove manufacturable before it becomes a real solution. Which is where the term Design-Technology co-Optimization (DTCO) essentially comes from (more or less replacing the older DFM term): the technology used

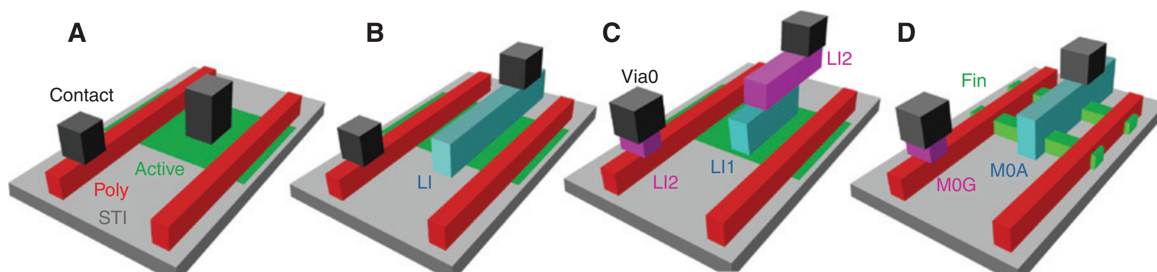

Figure 20: Different scenarios for contacting Active and Gate. (A) No local interconnect (LI), (B) single-layer LI, (C) dual-layer, LI1/LI2, (D) M0A/M0G scenario, for use with FinFets. (Dimensions not on scale.)

for integration of the entire application needs to allow for whatever connection scheme is attempted. In the case of the 'single layer' LI scenario ('1LI') of Figure 19C, this turned out to be not the case. Another problem with this solution is also that the active-active LI connections (not used in the simple example of Figure 19C) would run close to a gate contact, leading to a risk for a short or leak between this LI and gate contact (especially when the gate pitch becomes smaller). So the 1LI scenario of Figures 19C and 20B is not a good option.

A solution to the problems presented by the single-layer LI approach is the two-layer LI approach shown in Figures 19D and 20C, which we will call here LI1 and LI2:

– LI1 connects Active to LI2, but is not allowed to run over STI
– LI2 connects to LI1 to Via0 and Poly to Via0 ('Via0' is a renaming in this scenario of what was previously called 'Contact'). The stacked LI1-LI2 construction allows to connect Active to the power rails, while avoiding the junction leaks.

We can see from Figure 19D that Metal1 has simpler shapes than in the case without an LI (e.g., in Figure 19B, Metal1 extensions to the power rails are needed to connect to Active), though at a relatively high cost: two LI layers are needed here, each of which would have to be printed with a multi-patterning approach.

The introduction of FinFet as a replacement of the 'planar' type Active allows for a simplified LI approach, shown in Figures 19E and 20D. The reason is that LI can be allowed now to run over STI (the transistor junctions are somewhat deeper below the surface and the risk for electrical leaks is gone). It is again a dual-layer approach, and to distinguish it from the previous approach, we will now use the term Metal0 (M0) instead of LI. The two M0 layers are [124]:

– M0A: connects Active to Via0 and is allowed to run over STI.
– M0G: connects Poly to Via0

This approach leads to a simplified patterning for M0: less masks will be needed than in the LI1-LI2 cases. But the Metal1 pattern does not change very much. Both the LI1-LI2 and M0A-M0G approaches lead to the type of 2D Metal1 patterns that we have seen in the earlier sections of this paper.

The last step in our story finally brings us to 1D metal patterns: Figure 19F. This happens by 'dividing' the connections made by Metal1 in all the previous cases (Figure 19B–E) over Metal1 and Metal2 (interconnected by Via1). In the solution presented in Figure 19F

– The M0 layers are essentially used as before
– Metal1 is now horizontal only
– Metal2 is vertical only

Figure 21 shows the example for another, somewhat more complex Logic cell, comparing the metal 2D vs. 1D solutions for the FinFet type cells (i.e., the equivalents of Figure 19E and F); see also reference [125] for a similar comparison. In all cases, the 2D to 1D transition does not increase the cell area, but of course, it does add an extra via and metal layer (and hence cost). On the positive side, the 1D solution offers a better 'pin access' to the cells, i.e., the metal lines that make up the input and output signals of the cells (A, B, and Z in the example of the NR2D1 cell) are longer, and hence provide more degree of freedom to the cell router on where exactly this connection will be made. This could lead to area gain in the routing process. Figure 21C and D also show that for SRAM as well, there is a layout solution with Metal1 horizontal, i.e., the same Metal1 orientation as in the Logic cells. This is, of course, important with respect to the connectivity between SRAM and its environment (periphery).

We need to stress here that the examples shown in Figures 19 and 21 are just that, examples, and we do not claim that the solutions actually used in foundries and by ICMs literally corresponds to what we are showing here.

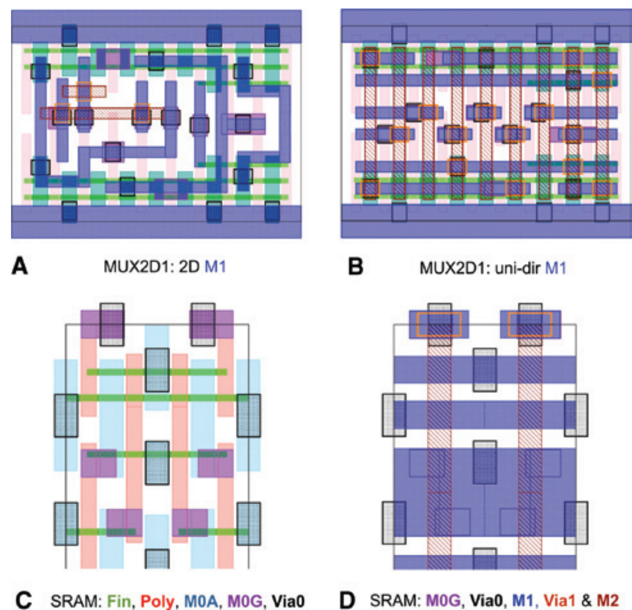

Figure 21: Example of a more complex logic cell, MUX2D1, using the M0A/M0G scenario with 2D Metal1 (A) and 1D Metal1 (B). SRAM layout with horizontally oriented 1D metal1: (C) Front-end, MOL and Via0 layers, (D) M0G, Via0, Metal1, and Via1. Metal2 (not drawn) runs vertically.

Also, the cells we show in this paper all correspond to what is called nine-track cells, which means that the vertical dimension of the cells is 9× some selected pitch value (often the pitch of Metal1 or Metal2). When cell libraries of a different track height need to be made, different solutions will probably be more efficient. Other (local) interconnect scenarios are possible as well; the ones shown here are given to illustrate the possibilities and a type of evolution that has taken place the past few logic nodes, and how that is indeed leading to a geometrical simplification of a number of layers.

It is also good to underline that this geometrical simplification was not the only (or not even the most important) motivation behind the introduction of, e.g., local interconnect or FinFet. These were in the first place oriented to improving the transistor performance. (For a review of some of the recent technology changes in the front end, see, e.g., references [126, 127]). But all these changes combined have now finally led to the geometrical simplifications that were already in some people's minds more than a decade ago.

Returning to the topic of patterning options, the metal patterning in the case of the 1D-layout type cells can now also be done using SADP (or SAQP). Apart from the benefit of being less overlay sensitive, SADP also offers the possibility of smaller tip-to-tip design rules. If the unused metal tracks can be replaced by dummy metal lines, the metal-block mask only needs to generate the tip-to-tip gaps in the metal lines. Figure 22 shows an example. Looking just at this small example, it is clear that the patterning burden for 1D layers that are printed with SADP/SAQP goes almost entirely to these cut or block layers. The size of these cut/block rectangles can be very small, and they can be located in relatively random configurations. Most of these cut/block layers will have to be multi-patterned, themselves. Because of their pattern size and complexities, cut/block layers are attractive candidates for the introduction of EUV lithography.

6 Conclusion

The past decade has shown significant changes in the path to printing critical patterns. Despite the fact that scanner wavelength and NA are no longer the drivers that enable scaling, as was the case before, scaling has continued. This has been made possible, of course, because of

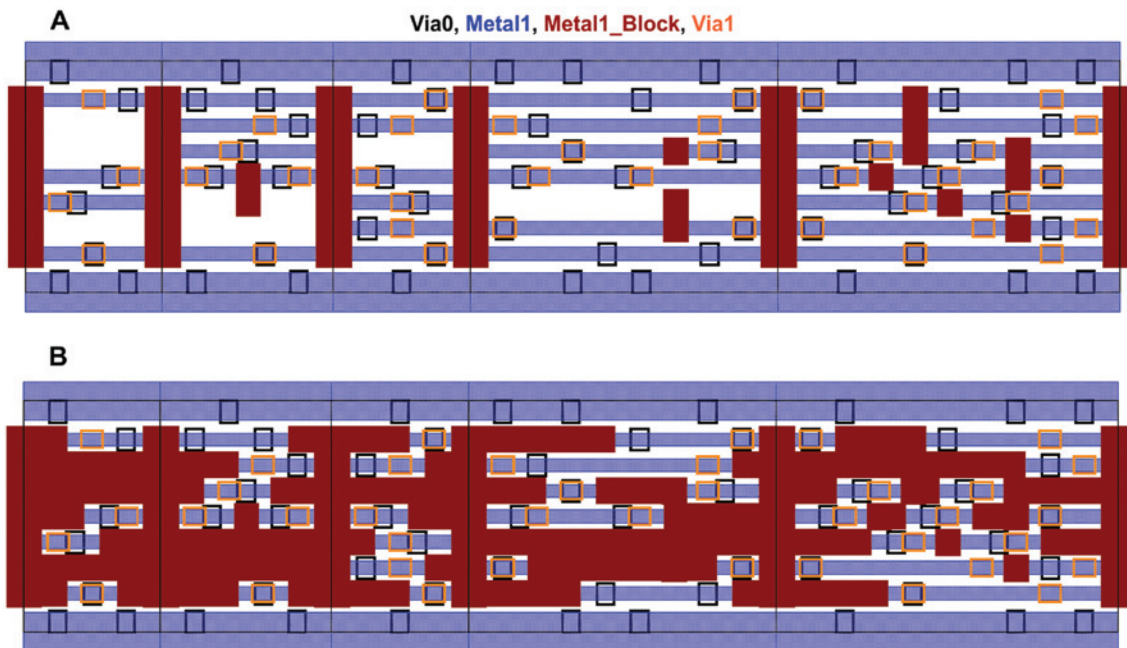

Figure 22: Row of five Logic cells. Compared to the layout examples of Figures 19F and 20B, an extra Metal1 line has been added to make the Metal1 patterning compatible with SADP (uniform space width between the Metal1 lines). (A) Minimum number of Metal1_Block shapes. After SADP patterning, the 'missing' Metal1 lines will be filled with dummy lines. As a result, the capacitive coupling between adjacent Metal1 lines can become unacceptably large. (B) Extended Metal1_Block content, to eliminate all unnecessary Metal1. It is clear that the patterning of Metal1_Block in both cases would be quite challenging, as it consists of a combination of small islands and larger (in case B very complex) shapes.

important changes in the technologies and materials used (e.g., metal gate, FinFet, local interconnect, high-k dielectrics, etc.), but also the tools used in the patterning flow itself, have grown enormously in accuracy and sophistication. In this paper, we have shown how the modeling tools are incorporating ever more physical effects (e.g., focus-dose effects, M3D, models for resist- and wafer 3D, etch, etc.). The OPC correction tools have become much more powerful, and the remaining hotspots can now be improved selectively by using local ILT. Contour metrology allows more accurate on-wafer evaluation, leading to better models and to better evaluation of the final printed results. Multiple-patterning techniques have matured and are now being actively used. In the last section, we showed how also the layout has become a variable in the total optimization exercise.

This does not mean that patterning has become easier. All these improvements have made that scaling still has some future ahead of it. The end of the optical lithography roadmap has been predicted many times in the past, but it seems that today, we have not yet reached the end of that road quite yet. We can continue to 'march to the beat of Moore's law' [117], for at least a while longer.

Acknowledgments: Many thanks to the following people: Jeroen Van de Kerkhove, Julien Mailfert, Staf Verhaegen, Darko Trivkovic, Werner Gillijns, Monique Ercken, Geert Vandenberghe, Patrick Wong, Vicky Philipsen, Alessandro Vaglio Pret, Mustafa Badaroglu, Julien Ryckaert, Karim Mercha, Jürgen Bömmels, Naoto Horiguchi, David Rio, Bart Laenens, Hua-yu Liu, Mu Feng, Germain Fenger, John Sturtevant, Kostas Adam, James Word, Andres Torres, Weimin Gao, Kevin Lucas, Ulrich Klöstermann, Chris Cork, Gerry Luk-Pat, Lena Zavyalova, Daisuke Fuchimoto, Hiroyuki Shindo, Daisuke Hibino, Doni Parnell, Grozdan Grozev, Shinji Tarutani, Mike Smayling, and Andreas Erdmann.

Appendix: summary of the main abbreviations used

DDM, Domain-Decomposition Method; DR, Design Rule; DRC, Design-Rule Check; DTCO, Design-Technology co-Optimization; EDA, Electronic Design Automation; EOL/EOT, End-of-Line/End-of-Trench; EPE, Edge-Placement Error; FE, Focus-Exposure; ILT, Inverse Lithography Technology; IPS, Image Parameter Space; ISP, In-Spec Percentage; LI, Local Interconnect; M3D, Mask 3D (or topography); M0, Metal0; M1, Metal1; MRC, Mask-Rule Check; MTT, Mean-to-Target; NTD, Negative-Tone Development; N20, 20 nm (Logic) node; OPC, Optical Proximity Correction; PTD, Positive-Tone Development; PV, Process Variation; PW, Process Window; R3D, Resist 3D (or resist profile); SADP, Self-Aligned Double Patterning; SRAF, Sub-Resolution Assist Feature; STI, Shallow-Trench Isolation; VB, Variability Band.

References

[1] C. Mack, in 'Fundamental Principles of Optical Lithography', Wiley, Chichester, UK, (2007).
[2] A. Wong, in 'Optical Imaging in Projection Microlithography', SPIE Press, Bellingham, USA, (2005).
[3] N. Cobb, A. Zakhor and E. Miloslavsky, Proc. SPIE 2726, 208–222 (1996).
[4] J. Sturtevant, J. Vac. Sci. Technol. B30, 030802 (2012).
[5] J. Stirniman and M. Rieger, Proc. SPIE 2197, 294 (1994).
[6] M. S. Yeung, D. Lee, R. Lee and A. R. Neureuther, Proc. SPIE 1927, 452–463 (1993).
[7] J. T. Azpiroz, A. Krasnoperova, S, Siddiqui, K. Settlemyer, I. Graur, et al., Proc. SPIE 7274-36 (2009).

Vector Effects

[8] M. Yeung, Proc. SPIE 922, 149 (1988).
[9] K. Adam, Y. Granik, A. Torres and N. Cobb, Proc. SPIE 5040, 78 (2003).

Mask 3D (M3D) effects

[10] A. Wong and A. Neureuther, Proc. SPIE 1809, 222 (1992).
[11] A. Erdmann, in 'Handbook of Photomask Manufacturing Technology', Ed. by S. Rizvi, CRC Press, Taylor and Francis, Boca Raton, FL, USA, (2007).
[12] A. Erdmann, P. Evanschitzky, J. T. Neumann and P. Gräupner, Proc. SPIE 9426-16 (2015).
[13] K. Adam and A. R. Neureuther, Proc. SPIE 4691, 107 (2002).
[14] K. Adam and M. C. Lam, Proc. SPIE 6924-49 (2008).
[15] P. Liu, Y. Cao, L. Chen, G. Chen, M. Feng, et al., Proc. SPIE 6520-27 (2007).
[16] P. Liu, Proc. SPIE 8166-149 (2011).
[17] J. Cheng, J. Schramm, D. Q. Zhang, Y. M. Foong, C. Zuniga, et al., Proc. SPIE 8326-62 (2012).
[18] H. Zhang, Q. Yan, E. Croffie, L. Zhang and Y. Fan, Proc. SPIE 8880-78 (2013).
[19] M. Lam, K. Adam, D. Fryer, C. Zuniga, H. Wei, et al., Proc. SPIE 8683-48 (2013).
[20] D. Fryer, M. Lam, K. Adam, C. Clifford, M. Oliver, et al., Proc. SPIE 9052-42 (2014).
[21] B. Yenikaya, C. Chuyeshov, O. Bakir and Y. Han, Proc. SPIE 9052-43 (2014).

Mask Process Correction (MPC)

[22] T. Brist and G. Bailey, Proc. SPIE 4691, 1373(2002).
[23] J. Sturtevant, E. Tejnil, T. Lin, S. Schultze, P. Buck, F, et al., Proc. SPIE 8683-6 (2013).
[24] E. Tejnil, Y. Hu, E. Sahouria, S. Schulze, M. Tian, et al., Proc. SPIE 6924-125 (2008).
[25] T. Lin, T. Donnelly, G. Russell, S. Jung and J. Jeong, Proc. SPIE 7971-6 (2011).

Resist 3D (R3D) Effects

[26] W. Gao, U. Klostermann, I. Kamohara, T. Schmoeller, K. Lucas, et al., Proc. SPIE 9052-11 (2014).
[27] C. Sarma, A. Abdo, T. Bailey, W. Conley, D. Dunn, et al., Proc. SPIE 7641-9 (2010).
[28] A. Dave, O. Kritsun, Y. Deng, K. Yoshimoto, J. Li, et al., Proc. SPIE 7274-40 (2009).
[29] P. Liu, Z. Zhang, S. Lan, Q. Zhao, M. Feng, et al., Proc. SPIE 8326-9 (2012).
[30] C. Wu, J. Chang, H. Song and J. Shiely, Proc. SPIE 8880-89 (2013).
[31] C. Zuniga and Y. Deng, Proc. SPIE 9052-81 (2014).
[32] S. Suh, I. S. Kim, E. M. Lee, Y. S. Kang, S. J. Lee, et al., Proc. SPIE 5752, 9 (2005).
[33] A. L. Charley, M. Dusa, T.-B. Chiou, P. Leray, S. Cheng, et al., Proc. SPIE 8324-86 (2012).
[34] Y. Fan, K. Koh, Q. Yang, W. Hoppe, B. Kuechler, et al., Proc. SPIE 8683-43 (2013).
[35] A. Szucs, J. Planchot, V. Farys, E. Yesilada, L. Depre, et al., Proc. SPIE 9052-7 (2014).

Wafer Topography (W3D) Effects

[36] T. Matsuzawa, A. Moniwa, N. Hasegawa and H. Sunami, IEEE Trans. CAD 6, 446 (1987).
[37] K. Tadros, A. Neureuther, J. Gamelin and R. Guerry, Proc. SPIE 1264, 322 (1990).
[38] A. Erdmann, C. Kalus, T. Schmöller, Y. Klyonova, T. Sato, et al., Proc. SPIE 5040, 101 (2003).
[39] H. Song, J. Shiely, I. Su, L. Zhang and W.-K. Lei, Proc. SPIE 7488-126 (2009).
[40] H.-J. Stock, L. Bomholt, D. Krüger, J. Shiely, H. Song, et al., Proc. SPIE 7640-64 (2010).
[41] E. Sungauer and F. Robert, Proc. SPIE 8326-89 (2012).
[42] J.-C. Le-Denmat, C. Martinelli, E. Sungauer, J.-C. Michel, E. Yesilada, et al., Proc. SPIE 8683-39 (2013).
[43] M. Oh, H. Youn, N. Chung, J. Maeng, S. Lee, et al., Proc. SPIE 8326-103 (2012).
[44] J.-C. Michel, J.-C. Le Denmat, A. Tishchenko and Y. Jourlin, Proc. SPIE 9052-77 (2014).
[45] J.-C. Le-Denmat, J.-C. Michel, E. Sungauer, E. Yesilada, F. Robert, et al., Proc. SPIE 9052-12 (2014).
[46] G. Winroth, A. Vaglio Pret, M. Ercken, S. Robertson and J. Biafore, Proc. of SPIE 9052-34 (2014).

Laser Bandwidth Effects

[47] P. De Bisschop, I. Lalovic and F. Trintchouk, J Micro Nanolithogr MEMS MOEMS 7, 033001 (2008).
[48] C. Zuniga, K. Adam, M. Lam, I. Lalovic and P. De Bisschop, Proc. SPIE 7274-60 (2009).

Stray Light (Flare)

[49] Y.-C. Kim, P. De Bisschop and G. Vandenberghe, J Micro Nanolithogr MEMS MOEMS 4, 043002 (2005).
[50] G. F. Lorusso, F. Van Roey, E. Hendrickx, G. Fenger, M. Lam, et al., J Micro Nanolithogr MEMS MOEMS 8, 041505 (2009).
[51] L. Sun, O. Wood, E. Verduijn, M. Singh, W. Wang, et al., J Micro Nanolithogr MEMS MOEMS 12, 042001 (2013).
[52] J. Word, C. Zuniga, M. Lam, M. Habib, K. Adam, et al., Proc. SPIE 8166-14 (2011).

EUV Shadowing and OPC

[53] M. Sugawara, M. Ito, T. Ogawa, E. Hoshino, A. Chiba, et al., J Micro Nanolithogr MEMS MOEMS 2, 27 (2003).
[54] M. Bienert, A. Göhnemeier, O. Natt, M. Lowisch, P. Gräupner, et al., J Micro Nanolithogr MEMS MOEMS 8, 041509 (2009).
[55] G. McIntyre, C.-S. Koay, M. Burkhardt, H. Mizuno and O. Wood, Proc. SPIE 7271-48 (2009).
[56] P. Liu, X. Xie, W. Liu and K. Gronlund, Proc. SPIE 8679 (2013).
[57] G. F. Lorusso, J. Van de Kerkhove, P. De Bisschop, E. Hendrickx, J. Jiang, et al., Proc. SPIE 8322-56 (2012).
[58] T. Coskun, T. Wallow, G. S. Chua, K. Standiford, C. Higgins, et al., Proc. SPIE 9048-31 (2014).
[59] V. Philipsen, E. Hendrickx, E. Verduijn, S. Raghunathan, O. Wood, et al., Proc. SPIE 9235-18 (2014).

Image-Parameter Space (IPS)

[60] Y. Granik, N. Cobb and T. Do, Proc. SPIE 4691, 377 (2002).
[61] T. Roessler, B. Frankowsky and O. Toublan, Proc. of SPIE 5256, 222 (2003).
[62] K. Patterson, Y. Trouiller, K. Lucas, J. Belledent, A. Borjon, et al., Proc. SPIE 5756, 294 (2005).
[63] M. Niehoff, S. Shang and O. Toublan, Proc. SPIE 6156-50 (2006).
[64] Y. Sun, Y. M. Foong, Y. Wang, J. Cheng, D. Zhang, et al., Proc. SPIE 7973-92 (2011).
[65] D. Vengertsev, K. Kim, S.-H. Yang, S. Shim, S. Moon, et al., Proc. SPIE 8522-45 (2012).
[66] A. Abdo and R. Viswanathan, Proc. SPIE 7640-48 (2010).
[67] N. Casati, M. Gabrani, R. Viswanathan, Z. Bayraktar, O. Jaiswal, et al., Proc. SPIE 9052-18 (2014).

Metrology

[68] H. Morokuma, A. Sugiyama, Y. Toyoda, W. Nagatomo, T. Sutani, et al., Proc. SPIE 5752, 546–558 (2005).
[69] M. Sendelbach, et al., Proc. SPIE 9050-55 (2014).
[70] T. Ohashi, T. Sekiguchi, A. Yamaguchi, J. Tanaka and H. Kawada, Proc. SPIE 8681 (2013).
[71] T. Ohashi, S. Hotta, A. Yamaguchi, J. Tanaka and H. Kawada, Proc. SPIE 9050-19 (2014).
[72] B. Bunday, C. Montgomery, W. Montgomery and A. Cepler, Proc. SPIE 8681-18 (2013).

SEM Contour Metrolog

[73] D. Hibino, H. Shindo, Y. Abe, Y. Hojyo, G. Fenger, et al., Proc. SPIE 7638-67 (2010).
[74] D. Fischer, G. Han, J. Oberschmidt, Y. W. Cheng, J. Y. Maeng, et al., Proc. SPIE 6922-8 (2008).
[75] F. Weisbuch, K. K. Koh and K. Jantzen, Proc. SPIE 9052-78 (2014).
[76] K. Hitomi, S. Halle, M. Miller, I. Graur, N. Saulnier, et al., Proc. SPIE 9052-31 (2014).
[77] P. De Bisschop and J. Van de Kerkhove, J Micro Nanolithogr MEMS MOEMS 9, 041302 (2010).
[78] I. Santo, A. Higuchi, M. Anazawa, H. Bandoh, Proc. SPIE 9050-93 (2014).
[79] D. Fuchimoto, T. Ishimoto, H. Shindo, H. Sugahara, Y. Toyoda, et al., Proc. SPIE 9050-30 (2014).
[80] Y. Toyoda, H. Shindo, Y. Hojo and D. Fuchimoto, Proc. SPIE 9050-102 (2014).

SRAF Placement and Printing

[81] J. Ye, Y. Cao and H. Feng, US patent 2008/0301620 (2008).
[82] L. Barnes, A. Poonawala, B. Painter, A. Jost, T. Takei, et al., Proc. SPIE 7274-38 (2009).
[83] S. Jayaram, P. LaCour, J. Word and A. Tritchkov, Proc. SPIE 8683-17 (2013).

[84] A. Poonawala, B. Painter and C. Kerchner, Proc. SPIE 7488-38 (2009).
[85] J. Zhang, W. Xiong, Y. Wang, Z. Yu and M.-C. Tsai, Proc. SPIE 7488-35 (2009).
[86] A. Villaret, A. Tritchkov, J. Entradas and E. Yesilada, Proc. SPIE 8683-14 (2013).
[87] Y. Tang, C. Chou, W. Huang, R. Liu and T. Gau, Proc. SPIE 8326-45 (2012).
[88] R. Viswanathan, J. Tirapu Azpiroz and P. Selvam, Proc. SPIE 8326-46 (2012).

Etch Modeling
[89] Y. Granik, Proc. SPIE 4346, 98 (2001).
[90] S. Sato, K. Ozawa and F. Uesawa, Proc. SPIE 6155-3 (2006).
[91] I. Stobert and D. Dunn, Proc. SPIE 8685-3 (2013).
[92] L. Zavyalova, L. Luan, H. Song, T. Schmoeller and J. Shiely, Proc. SPIE 9052-76 (2014).
[93] J. Mailfert, J. Van de Kerkhove, P. De Bisschop and K. De Meyer, Proc. SPIE 9052-25 (2014).

Stochastic Effects
[94] P. De Bisschop, J. Van de Kerkhove, J. Mailfert, A. Vaglio Pret and J. Biafore, Proc SPIE 9048-8 (2014).

Advanced topics (matrix OPC, PW, Double Patterning, fragmentation).
[95] J. Lei, L. Hong, G. Lippincott and J. Word, Proc. SPIE 9052-22 (2014).
[96] M. Terry, G. Zhang, G. Lu, S. Chang, T. Aton, et al., Proc. SPIE 6520-28 (2007).
[97] X. Li, G. Luk-Pat, C. Cork, L. Barnes and K. Lucas, Proc. SPIE 7274-39 (2009).
[98] M. Gheith, L. Hong and J. Word, Proc. SPIE 7274-44 (2009).
[99] Y.-S. Woo, W.-H. Choi, B.-S. Seo, Y. Kim, V. Liubich, et al., Proc. SPIE 7640-49 (2010).
[100] S.-W. Lee, T. Cecil, G. Xiao, M. Lee, J.-H. Choi, et al., Proc. SPIE 8880-68 (2013).

Computational Verification
[101] J. Kim and M. Fan, Proc. SPIE 5256, 919 (2003).
[102] N. Cobb and D. Dudau, Proc. SPIE 6154-18 (2006).
[103] C. Beylier, N. Martin, V. Farys, F. Foussadier, E. Yesilada, et al., Proc. SPIE 8326-42 (2012).
[104] M. Kajiwara, S. Kobayashi, H. Mashita, R. Aburada, N. Furuta, et al., Proc. SPIE 9053-15 (2014).
[105] N.-Y. Chung, P.-S. Kang, N.-R. Bang, J.-D. Kim, S.-J. Lee, et al., Proc. SPIE 9053-16 (2014).
[106] S. Shim, S. Moon, Y. Kim, S. Choi, Y. Kim, et al., Proc. SPIE 7973-90 (2011).

Geometrical Target Pattern Screening
[107] H. Yao, S. Sinha and X. Hong, Proc. ICCAD (2006) p. 625.
[108] J. A. Torres, IEEE/ACM International Conference on Computer-Aided Design (ICCAD), (2012) p. 349.
[109] S. Mostafa, J. A. Torres, P. Rezk and K. Madkour, Proc. SPIE 7974-6 (2011).
[110] B. Yu, J.-R. Gao, D. Ding, X. Zeng and D. Pan, J Micro Nanolithogr MEMS MOEMS 14, 011003 (2015).

On-Wafer Verification
[111] C. Tabery, H. Morokuma, A. Sugiyama and L. Page, Proc. SPIE 6152-51 (2006).
[112] K. Hashimoto, S. Usui, S. Nojima, T. Kotani, E. Yamanaka, et al., J Micro Nanolithogr MEMS MOEMS 8, 033007 (2009).

Layout and Technology
[113] L. Liebmann and D. Pietromonaco, M. Graf, Proc. of SPIE 7974-19 (2011).
[114] C. Bencher, H. Dai, L. Miao, Y. Chen, P. Xu, et al., Proc. SPIE 7973-19 (2011).
[115] L. Capodieci, Proc. SPIE 6154-1 (2006).
[116] D. Abercrombie, P. Elakkumanan and L. Liebmann, Proc. of. Electronic Design Process Symposium (EDPS) (2009).
[117] Y. Borodovsky, Proc. SPIE 6153-200 (2006).
[118] W. Fang, S. Arikati, E. Cilingir, M. Hug, P. De Bisschop, et al. Proc. SPIE 9053-9 (2014).
[119] G. Luk-Pat, B. Painter, A. Miloslavsky, P. De Bisschop, A. Beacham, et al., Proc. SPIE 8683-37 (2013).
[120] L. Liebmann, L. Pileggi, J. Hibbeler, V. Rovner, T. Jhaveri, et al., Proc. SPIE 7275-10 (2009).
[121] M. Smayling, R. Socha and M. Dusa, Proc. SPIE 7640-43 (2010).
[122] H. Zhang, M. Wong and K.-Y. Chao, Proc. of the Asia and South Pacific Design Automation Conference (ASP-DAC), (2010) p. 838.
[123] P. De Bisschop, B. Laenens, K. Iwase, T. Yao, M. Dusa, et al., Proc. SPIE 7973-10 (2011).
[124] E. Terzioglu, 2014 Int. Symp. on EUV Lithography (2014).
[125] K. Vaidyanathan, S. H. Ng, D. Morris, N. Lafferty, L. Liebmann, et al., Proc. SPIE 8327-19 (2012).
[126] M. Bohr, Proc. SPIE 7974-1 (2011).
[127] K. Schuegraf, M. Abraham, A. Brand, M. Naik and R. Thakur, IEEE J. Electron Devices Soc. 1, 66 (2013).

Peter De Bisschop
Advanced Lithography Department, IMEC,
Kapeldreef 75, B-3001 Leuven, Belgium,
peter.debisschop@imec.be

Peter De Bisschop obtained a PhD in Atomic Physics in 1984 and has been at IMEC since 1986. Until 1995, he worked in the Materials Analysis group and, since then, he has worked in the Advanced Lithography Department. Among the topics he has worked on are stepper/scanner control, monitoring and qualification, lens aberrations, stray light, laser-bandwidth effects, CDU, 157-nm hard pellicles, 193i air bubble defects, M3D effects, and rigorous lithographic simulations. For the past 10 years, his focus has been on proximity-related topics, OPC modeling and verification, contour metrology, as well as DTCO and logic cell layout, and EUV stochastic effects.