A GPU-Accelerated Physical Model for the New Curvilinear Mask Era

Background and Application of D2S TrueModel®

Executive Summary

Increased complexity introduced by the ongoing march of Moore's Law has increased the need for mask-model accuracy. Many changes are coming to the mask industry – including greater use of inverse-lithography technologies (ILT), the advent of multi-beam mask writing, and the anticipated introduction of extreme ultra-violet (EUV) lithography – that will require more detailed, more accurate models. GPU acceleration opens the door for simulation-based correction of a multitude of complex mask effects based on physical models, affording practical simulation runtimes for these more complex models. D2S TrueModel® is a single, physical model that includes advanced modeling of dose, shape, chemical resist and development to adapt to all these changes. TrueModel has been refined over six generations of test chips with unique structures to enable calibration and validation of these advanced mask effects. TrueModel is a part of the inline pixel-level dose correction (PLDC) capability in the NuFlare Technology MBM-1000 multi-beam mask writer.

Complexity Drives the Need for More Accurate Models

As Moore's Law continues, the need for increased process margin, particularly in depth of focus, is driving optical proximity correction (OPC) to use inverse lithography technology (ILT) to create more complex mask features. Today's leading-edge masks – especially contact and cut layers – will include curvilinear mask shapes, which today are converted to complex orthogonal shapes with small jogs. These desired patterns can result in sub-60nm features on the mask. Linearity, corner rounding, and line-end shortening are among the issues that need mask correction to ensure that the actual reticle matches the expectations of OPC/ILT. Simulation-based mask correction is required; however, simulations are only as accurate as the models they employ.

Masks that include complex shapes require 2D validation. Today's mask writing instruments for precision layers use a variable shaped beam (VSB) tool, which is a Manhattan (1D) writing instrument, so models built using these tools are by definition 1D-centric. Inaccuracies in 1D models are exacerbated when tested against a 2D validation. Physics-based models are far more likely to extrapolate to 2D shapes, and are better for ILT.

EUV, when it is used in production, will also result in smaller, more complex mask shapes. Multi-beam mask writers, which will enable more complex shapes and per-pixel corrections, also add new challenges – and new opportunities – to the modeling process. To meet these challenges, more accurate simulation models that include more – and more-detailed – physical parameters will be required.

Why Physical Models are More Accurate

All mask models are constructed by measuring features on test masks, which are designed to extract parameters of particular effects of the mask-making process from eBeam exposure to etch. Good practice requires a model to be calibrated against one data set and tested against another. Two categories of models are thought to exist: the rigorous "physical" model, which attempts to embed the physics of the physical processes, and the "black-box" empirical model, where the focus is on finding parameters of a mathematical basis. In reality, no model has all the physics, and most rely on empirically useful kernels to make up the difference. The difference in empirical modeling and physical modeling, then, is in the overall philosophy of approach.

The approach of the empirical modeler is much like the big-data scientist using deep-learning techniques: to find correlated patterns in the data, and then to model mathematically the difference between simulated data and measured data. An extreme (incorrect) empirical approach would seek to just fit the data with any mathematical function to reduce the residual error on the training data set without regard to overall predictability. It is extremely easy to overfit these models, and care must be taken to sample the correct mask patterns to ensure coverage over all possible designs. Good empirical modelers take care to use the prediction data set as an overall judge of model quality. Even with this care, however, it can be challenging to find models that predict well over unseen data or extrapolate well to new processes.

The approach of the physical modeler is to base model components on various physical effects driven by principles of physics and chemistry. Physical modeling insists on breaking down the mask-making process into process steps and deducing the equations of motion that are most important for each step. More rigorous physical models seek to include more and more equations, sacrificing runtime for overall model accuracy. Practical runtime considerations necessarily require simplifications to the modeling system, which reduce the overall accuracy of the system on the calibration data set. However, if the crucial physics can be captured, these models almost always predict well versus new data and new processes. That is not to say that one cannot overfit physical models, although this generally occurs by including too many physical parameters.

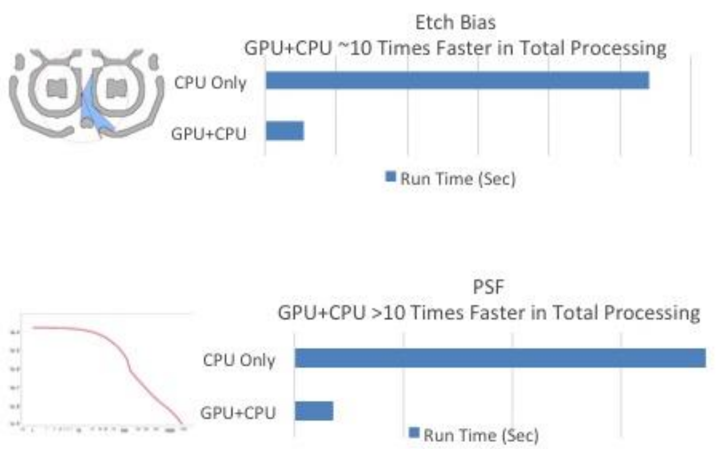
Therefore, in practice, all modeling approaches are a hybrid of both methodologies. Each mask shop needs to balance its approach by being sufficiently accurate for the target node, while being fast enough to meet its turnaround time requirements.

The philosophy behind D2S TrueModel is to be a physical model implemented to work at a full-mask level with practical run-time, and augmented by empirical fine-tuning to provide the best balance for the leading-edge nodes of today's semiconductor manufacturing. Importantly, the test masks used to calibrate and validate TrueModel have been selected to isolate key physics to allow the engineer to calibrate different portions of the model separately through both dose and shape modulation over both Manhattan and complex curvilinear shapes. The current version of TrueModel has been through six test-chip generations, each with dose modulation and overlapping shots as a part of the modeling.

GPU Acceleration Enables Simulation-Based Processing of Physical Models

Historically, simulation-based processing of mask models resulted in unacceptably long simulation runtimes. Thus, the most common approach has been to use model-based or rules-based methodologies that, while providing less accuracy, result in faster runtimes. The advent of GPU-accelerated mask simulation has changed this picture. GPU acceleration is particularly suited to “single operation, multiple data” (SIMD) computing, which makes it a very good fit for simulation of physical phenomena, and enables full-chip mask simulation to be executed within reasonable runtimes.

An additional advantage of GPU acceleration is the ability to employ arbitrary point-spread functions (PSFs), which are a natural choice for the mask-exposure model, including EUV mask mid-range scattering effects, forward-scattering details, and modeling back-scattering by construction. Any dose effect of any type can be exactly modeled during simulation-based processing.



.

Figure 1. The runtimes for etch-bias and PSF examples on a CPU-only computing platform, versus an accelerated GPU + CPU platform. Runtimes with GPU + CPU are 10 times faster.

Modeling Requirements for the New Curvilinear, EUV, and Multi-Beam Era

The resolution limit of 193i lithography has already been reached, so the features projected by light onto the wafer will not shrink for any given mask layer. However, since the wafer pitch must shrink, process engineers are using double-, triple-, or indeed up to hexa-patterning to achieve the small feature sizes needed for the <10nm nodes. The result of this is a need for ultimate precision: a constraint of under 1nm edge-placement error (EPE) is typical for the 7nm node. This means process variability needs to be extremely small to avoid wafer defects.

Because of this, the mask industry is facing several major changes. First, polygon count is expected to grow dramatically due to increased usage of curvilinear ILT for 193i in an effort to improve process windows. Second, the production use of EUV is imminent, which means mask feature sizes will need to shrink. Finally, multi-beam mask writers are on track for production use within the next year, which enable users to have access to advanced eBeam dose profiles. All of these changes will have major impacts on mask-shape complexity and density, and bring fundamental changes to the way we will need to model dose, shape and scattering.

ILT – meaning complex mask shapes – will be used to enhance the lithography process window and reduce variability. These complex features may be Manhattan orthogonal shapes with very small jog sizes as they are today, or, if written with a multi-beam mask writer, they may be designed to be curvilinear in shape. Regardless, the shapes that are actually produced on the physical masks today are curvilinear already due to corner rounding effects. This drives the need to accurately model complex, curvilinear mask shapes.

The production availability of EUV lithography is anticipated eagerly because EUV can avoid costly and complex 3x (or more) patterning with 193i. EUV masks will have more “main” features, each smaller than those on 193i masks, due to smaller EUV wavelengths that provide enhanced resolution. It is likely that EUV will require double-patterning, so we will likely see EUV SRAFS with sizes below 30nm on masks. EUV masks have a more complex scattering profile due to the reflective multilayer, which causes a 1µm “mid-range” scatter. PSF-style models will also likely be needed to provide the accuracy required for EUV masks.

The introduction of multi-beam mask writers will also impact models in a subtle way. VSB mask writers mostly write rectangles sequentially via shaped electron beams. Multi-beam mask writers use several hundred thousand beamlets to write a rasterized field of pixels, with each pixel being assigned its own dose. While the dose profile could be a simple rasterization of the geometry, it is possible to adjust individual beamlets to create a complex dose profile to reduce process variability. This means that the dose profiles for multi-beam mask writers will be much more complex than those of VSB writers.

Because the multi-beam tool has a constant write time, the likely targets for multi-beam mask writing are ILT layers or EUV layers due to their high polygon counts. This enables masks to target features that are non-orthogonal, non-45-degree diagonals, and curvilinear without a write-time impact.

In all cases, a good physical model is much more likely to be accurate across all shapes than an empirical model fit over mostly rectangular test mask structures. All these changes, both separately and taken together, create the need for a fundamental paradigm shift in modeling: treating dose and shape separably.

Historically, features have been assumed to be rectilinear, and typically have been adjusted with a constant bias. Forward scattering, when included, has been assumed to be a set of Gaussians, with length scales under 300nm. Back-scatter ("PEC") and fogging ("FEC") effects of eBeam are ignored, assumed to be taken into account by the exposure tool. The same is true with the density-dependent, etch-loading effect ("LEC") and the charging effect ("CEC") and thermal effect for tools with high exposure current.

The classic eBeam exposure and etch models have treated dose and shape with the same term. To meet the challenges posed by both EUV and multi-beam mask writing – and especially since they are likely to be employed together – a more rigorous treatment of the mask model into specific dose and specific shape effects is necessary to achieve the require results (see Figure 2).

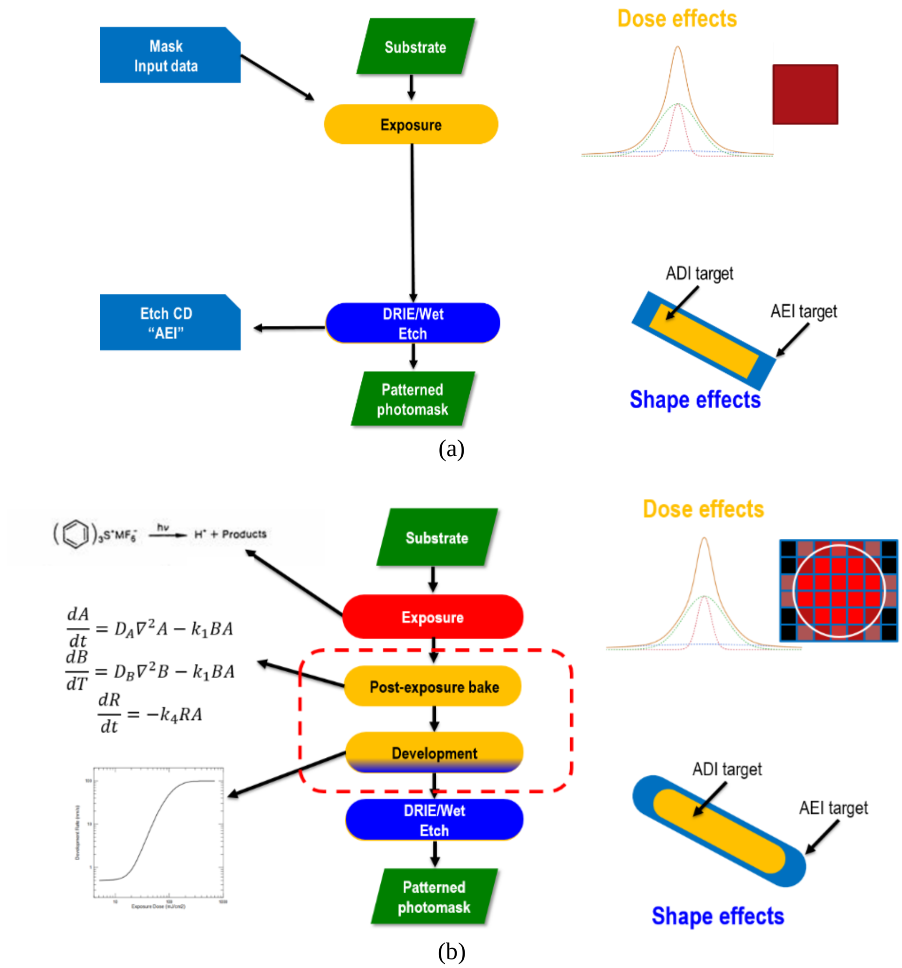


Figure 2. Simple (standard) mask model (a) versus complex (separable) mask model (b). (Images courtesy: Cliff Henderson, Lithoguru.com). Slides presented at Photomask Japan 2017[1].

---

[1] Pearman, Ryan, et al, "EUV modeling in the multi-beam mask writer era," SPIE Photomask Japan, 2017.

In the past when the majority of masks were written at nominal dose, dose-based effects could be "bundled" together with shape-based effects. Any empirical modeling approaches would have been wise to bundle the effects together for computational brevity. With the introduction of overlapping shots by D2S, and the study of the effects of dose changes in the 2X region that followed, it became clear that dose modulation (which includes overlapping shots) requires specific modeling. Some effects are subject only to the resist contour sizes, while other effects cause variation in edge location based on differences in exposure slope near the contour edge. Multi-beam tools add this extra complexity and more – there are significant differences in the dose profile even amongst identical patterns as they align differently with the multibeam pixel grid. TrueModel employs several techniques to model accurately the impacts of these changes.

Since even Manhattan shapes drawn as the target design for masks become corner-rounded on the actual mask, all shape effects are computed more accurately in the curvilinear space. Etching effects computed with Manhattan jogs will be inaccurate, since the physical effects of etching depend on curvilinear distances that do not honor the concept of X and Y axes.

Dose effects, too, need to be better than just a multiple Gaussian model. Many physical effects, including the EUV-mid-range scatter, are not fundamentally Gaussian effects, so any multi-Gaussian approximation will have built-in error. The same holds true for chemically amplified resist effects. For model compatibility, TrueModel supports multiple Gaussian models. However, the new arbitrary-PSF-based dose modeling can provide the ultimate in flexibility for the end-user to efficiently express the dose model, while constraining to known physical effects.

A Single Physical Model: The Most Complete – and Most Practical – Approach

With all of the disparate changes currently underway in the semiconductor manufacturing world, it might be tempting to develop a specialized model for each situation. Why not develop a model for EUV, and another for multi-beam mask writers, and yet another for slow resists? An empirical modeling approach may do that, but a physical modeling approach would not.

A model form that is capable of being adapted to any mask writer, any resist, any dose profile, and any shape is not only the most complete approach, but also the most practical. Additions to the model to accommodate a new target (e.g., the EUV mid-range) also makes the model more accurate for everything else since 193i masks have always had a faint mid-range signal that is ignored for practicality.

TrueModel Application Example: Inline PLDC in the NuFlare MBM-1000

Combining a physical model that includes advanced modeling components, such as dose, with both GPU acceleration and multi-beam mask writing technology creates an opportunity to manipulate mask writing and correction on a per-pixel level. This provides a new level of accuracy for advanced masks. At SPIE Photomask Japan 2017, D2S and NuFlare Technologies jointly presented such an application: inline

PLDC using D2S TrueModel technology in the NuFlare MBM-1000 multi-beam mask writer.[2]

As the paper details, in this application, PLDC provides short-range (effects in the 10nm scale to 3-5µm scale) linearity correction while at the same time improving the overall printability, line-edge roughness (LER) and CDU of the mask. The traditional effect corrections included inline with mask writers (PEC, LEC, FEC), such as 4G PEC modeling, continue to be inline in the MBM-1000. PLDC combines a 10nm-100nm short-range linearity correction with a 1µm-scale mid-range linearity correction, which is especially useful for EUV. This is the first time for any mask writer to include either of these linearity corrections. GPU acceleration enables the PLDC to be performed inline and therefore helps to maintain turnaround time in the mask shop.

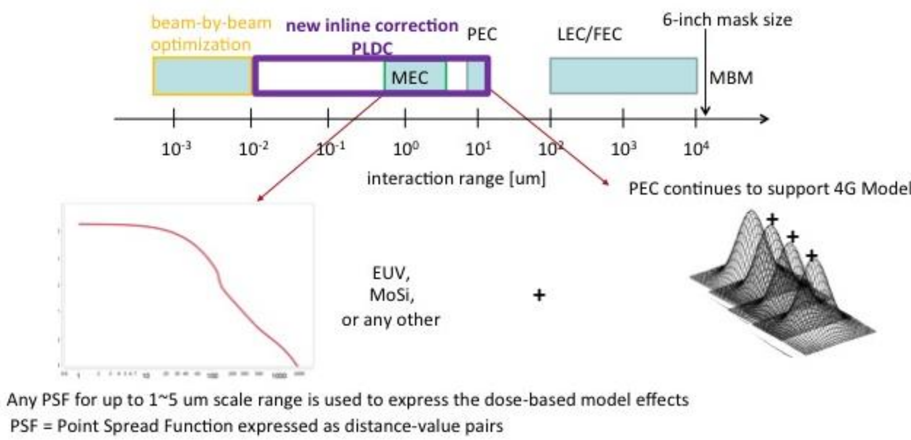


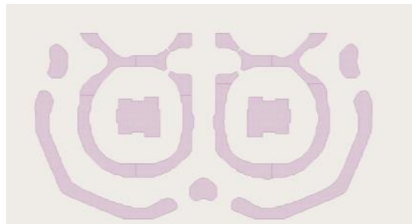
Figure 3. PLDC in the context of other correction mechanisms in the MBM-1000 (As presented at SPIE Photomask Japan, 2017).

In Figure 3, the PLDC portion of the inline corrections is depicted in purple. As the paper states, "Within the purple region, the dose-based effects portion of TrueModel are expressed as an arbitrary point spread function (PSF) for an interaction range up to 3-5µm and with a 4G PEC model for interaction range up to 40-50µm."

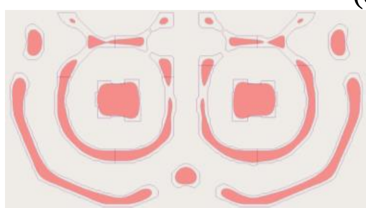
This ability to model physical effects and correct for them inline with mask writing results in more accurate masks, including for smaller EUV shapes and for curvilinear ILT mask shapes. In the example shown in Figure 4, the main features are targeted to be rectilinear (but actually print as curvilinear) and the SRAFs are curvilinear. The MBM-1000 in this example is writing multiple passes at each of (0,0) and (5,5) offset passes.

[2] Zable, Harold, et al, "GPU-accelerated inline linearity correction: pixel-level dose correction (PLDC) for the MBM-1000," SPIE Photomask Japan, 2017.

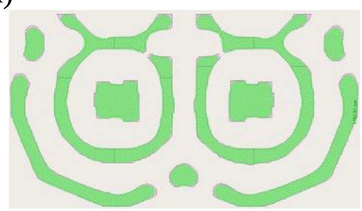
Figure 4(a) shows the target ILT pattern; (b) shows in pink a simulation of the mask pattern that would print without PLDC; (c) shows in green the mask pattern that would print with PLDC. The simulation uses an example PSF model shown in 4(d) to predict an ADI (after development, but before etch) contour. The SRAFs narrow without PLDC (pink, b), but are corrected with PLDC (green, c).



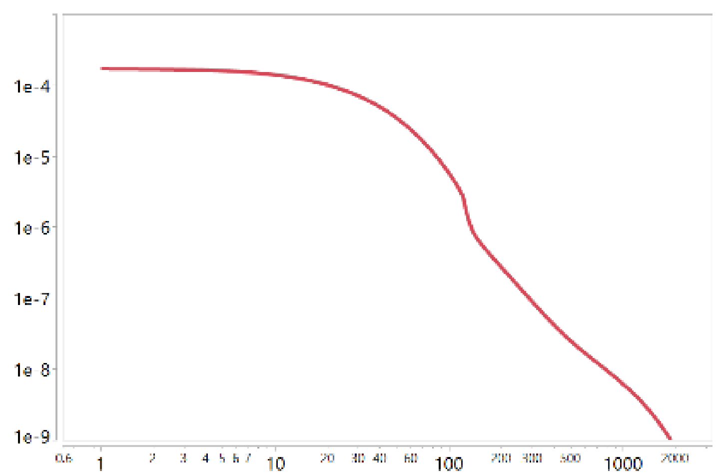
(a)



(b)



(c)



(d)

Figure 4. For ILT pattern (a), using the PSF as shown in (d), (b) is the simulation prediction of ADI pattern without PLDC (only rasterization) and (c) is with PLDC. In (d), the X-axis is in nm, and the Y-axis is the proportion of energy dispersed to that distance.

PLDC is simulation-based, so it has the ability to be very accurate regardless of targeted shape, regardless of mask type (e.g., positive, negative EUV, ArF, NIL master) with the right set of mask modeling parameters. Because TrueModel uses a physical modeling approach, actual physics, chemistry and math are reflected directly in the model, allowing PLDC in the MBM-1000 to model any type of a mask accurately.

Conclusion

TrueModel provides a single, physical model for substrates written by eBeam. It includes complex modeling of dose, shape and exposure to provide the necessary accuracy for today's leading-edge curvilinear designs and anticipated EUV and multi-beam applications. TrueModel has been calibrated and validated by six generations of advanced test chips, an effort that continues to keep pace with advances in process, mask writing and lithography. GPU acceleration enables TrueModel to provide physical-model accuracy within practical runtimes. TrueModel is in production use in several semiconductor-manufacturing applications, including D2S TrueMask MDP mask data preparation and NuFlare Technology MBM-1000 multi-beam mask writer.