Mask Hotspots Are Escaping the Mask Shop; Model-Based Mask Verification Can Stop Them

By Aki Fujimura, D2S, Inc.
Managing Company Sponsor of the eBeam Initiative

Abstract

Although the overwhelming majority of wafer production issues at the 28nm-and-below process nodes are lithography- and OPC-related, the semiconductor industry is starting to see problems caused by mask hotspots: wafer-level production issues that are caused when the shapes specified by optical proximity correction (OPC) are not faithfully reproduced on the mask on a per-instance basis. Mask hotspots will account for an increasingly larger portion of wafer-level issues as we go into 14nm-and-below nodes, regardless of which lithography technique is used. Today's mask inspection methodology reliably flags mask defects only down to a level that is becoming insufficient for the increasingly smaller and more complex shapes on today's advanced masks. For these masks, rules-based solutions are no longer adequate. While some parts of variation in nominal contour can be remedied through mask process correction (MPC), the problem of poor resilience to manufacturing variation must be detected and resolved in the mask shop.

The best solution is a simulation-based mask data preparation flow, such as model-based mask data preparation (MB-MDP), in which potential hotspots are eliminated by construction. For masks that are conventionally fractured, a simulation-based full-chip verification is the only way to catch shape-dependent mask hotspots across the entire mask before the mask leaves the mask shop. Model-based mask verification (MB-MV) is a methodology that can be deployed today to perform full-chip mask verification in parallel to mask writing, without displacing any portion of the current mask making process. This new methodology can keep costly mask hotspots from escaping the mask shop without risk to established flows.

As with the Wafer, So with the Mask

For many years, the semiconductor industry has been working to overcome the physics-based issues impacting the semiconductor lithography process that have accompanied the adoption of advanced process nodes. Simply put, as linewidths shrink below a certain size, the physics of light cause "what you see" to *not* be "what you get."

Rules-based OPC and, more recently, model-based OPC have been adopted to correct for the diffraction of light by making small "corrective" additions to mask shapes in order to ensure that the desired shape is produced on the wafer. OPC has become an indispensable part of creating semiconductors in the "nanometer" era. As process nodes have continued to shrink, the mask shapes created by OPC have become more complex as well as smaller.

At the 28nm process node, the same types of physics-based issues that have haunted lithography for decades have started to impact mask writing as well. The increasingly small and complex OPC-specified mask shapes required for faithful wafer lithography at 28nm-and-below nodes have given rise to an increase in mask hotspots where the shapes produced on the mask deviate enough on a per-instance basis from the shapes designed by OPC to impact negatively on wafer quality and yield.

Both mask hotspots and the more familiar wafer or lithography hotspots can cause wafer quality and yield issues. The difference between these two kinds of hotspots is that mask hotspots are errors located on the mask itself. If the specified OPC shapes are written faithfully to the mask, but then turn out to be insufficient to reliably produce the desired shapes on the wafer, this is a lithography hotspot. However, if the specified OPC shapes are not produced faithfully on the mask, and this causes the wafer shapes to be insufficient across process margin, this is a mask hotspot.

Neither type of hotspot is avoidable in all situations. For a lithography hotspot, it may be that better OPC would have avoided the hotspot. Or, it may be that no process – OPC, ILT, or even pixilated mask – could have avoided the hotspot. For mask hotspots, it may be that better mask data preparation, mask writing, or correction could avoid the hotspot. Or, it may be that there isn't any process that could write the incoming shape faithfully enough.

The need for overlay accuracy in multiple-patterning increases the need for precision on the wafer across a sufficiently wide process window. The need to avoid lithography hotspots in turn drives the need for smaller, more complex shapes on the mask. Smaller shapes, more complex shapes, and particularly non-orthogonal edges, on the mask produce an increasing likelihood of mask hotspots.

The Shape-Dependent Nature of Mask Hotspots and Dose Margin

Lithographic process windows – a proxy for the allowable deviation in likely sources of manufacturing variation to achieve desired device performance – are getting smaller with shrinking process nodes. Smaller process windows not only make manufacturing more difficult and expensive, but also put limitations on designers through stricter design rules, which define the boundaries of the process window and limit the shapes and sizes of design features.

The mask-making equivalent of lithography process windows is dose margin (DM), or normalized image log (NIL). DM is the inverse of the feature-size sensitivity to the amount of energy provided in the eBeam shot used to produce the feature. DM is a good proxy for many sources of manufacturing variation on the mask, such as reticle-film uniformity, across-reticle process uniformity, absorber-edge roughness and sidewall angle control, in the same way exposure latitude (EL) is a good proxy for manufacturing variation on the wafer. There are, of course, other sources of

mask manufacturing variation not correlated with DM, such as CD split[1]. We focus on DM, and specifically shape-dependent variation in DM, because it is a newly emerging problem in leading-edge masks. Depth of focus (DOF), which is the other important dimension for wafer lithography, isn't a factor for mask making because the 50KeV eBeam projection used to write the mask has better DOF.

In the old days of mask making, when only orthogonal edges were specified to be written on the mask, and when the minimum size of any "jogs" on the mask were 100nm, the innate accuracy of eBeam made "what you specify is what you get" a reasonable assumption. The combined process blur in mask making is roughly 25nm to 35nm in range, so 100nm was more than three sigma away from any danger zone, making it possible to write the edges precisely enough every time.

Now that mask shapes are smaller, and more complex shapes with small jogs and non-orthogonal edges need to be written on the mask with precision, "what you get" is often no longer what you specified. This is true for two different reasons. First, the nominal contour might not be where the contour was drawn, depending on shape and context of the shape. Second, the dose margin could be worse depending on the shape and the context of the shape, making the variation from the nominal contour too large. As a result, resilience to manufacturing variation is poor for those shapes, leading to poor line-edge roughness (LER) and poor critical dimension uniformity (CDU). Displacements in mask contours caused by either of these two sources of error can result in mask hotspots.

Mask hotspots – whether due to the predictable displacement of the nominal mask contour away from the mask design designated by OPC, or due to the displacement of the manufactured mask contour away from the mask design due to manufacturing variation – are shape-dependent. Accordingly, both kinds of mask hotspots must be monitored shape by shape, over the entire mask. While some parts of variation in nominal contour can be remedied through MPC – either embedded in the OPC process or in the mask shop – the problem of poor resilience to manufacturing variation must be detected and resolved in the mask shop.

The old assumption that the mask process adjustments are largely shape-independent, and that only the overall mask process needs to be controlled, biased, and centered is no longer valid. Because of this change, the current rules-based approaches to mask inspection are no longer sufficient. For today's masks with complex shapes at 28nm-and-below nodes, a simulation-based approach that can verify the entire mask with each shape in its surrounding context is required.

---

[1] D. S. S. Bhardwaj, et al (SoftJin Technologies Pvt. Ltd. (India)), "QoR analysis of fractured data solutions using distributed processing," Proc. SPIE 8166-16, Session 6, Photomask Technology, September 2011.

Current Mask Inspection Methodology Allows Hotspots to Escape the Mask Shop

The semiconductor industry began to see an increase in mask hotspots starting at the 28nm node. The biggest problem is that because current mask verification methodologies do not flag these hotspots, they are usually identified only after expensive mask production, or even after wafer production, so they are very costly in terms of both production cost and time.

Today's mask-making methodology (see figure 1) relies on mask inspection to catch mask errors other than data-handling errors (which are detected through routine XOR-based verification). Mask inspection machines shine light through the mask under magnification and employ sophisticated image processing to flag potential defects. Other tools such as critical dimension scanning electron microscopes (CD-SEM) and aerial image metrology systems (AIMS) help to categorize defects by taking a more accurate look, but only in the isolated spots flagged by mask inspection. Both CD-SEM and AIMS, while accurate, are too expensive and time-consuming to be used to examine the entire mask.

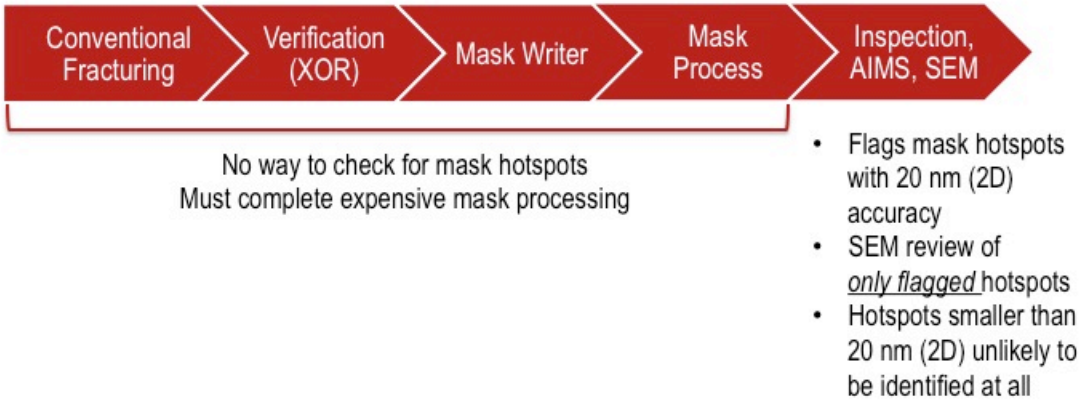

Figure 1: In today's mask inspection flow, there is no way to check for mask hotspots until expensive and time-consuming mask processing is finished. Hotspots of less than 10 nm (1D) are not flagged at all.

With today's inspection machines, 1D feature differences down to 10nm (4X dimensions) can be observed with reasonable reliability, depending on machine settings. Particularly for two-dimensional features, shape-dependent differences in the actual mask shapes from the intended mask shapes of less than 10nm are not flagged by current mask inspection methodologies[2].

As masks with very small, complex shapes are becoming the norm, mask shops need a way to insure that mask hotspots are caught before they can impact wafer

[2] W. Broadbent, et al (KLA Tencor Corporation),"EUV Reticle Inspection with a 193nm Reticle Inspector," Proc. SPIE 8701, Photomask and Next-Generation Lithography Mask Technology XX, 87010W (June 28, 2013).

production. This problem will only get worse as minimum mask-shape size moves well below 60nm, as edge-placement accuracy requirements for all shapes on the mask become increasingly stringent, as mask shapes continue to increase in complexity for immersion, and as line-end-to-line-end precision becomes critical for EUV and other techniques.

Model-Based Mask Verification: The New Requirement at 28nm and Below

Simulation-based verification for the entire mask is the only way to keep hotspots from escaping the mask shop. Until recently, such full-chip verification was a practical impossibility because of the tremendous computational power required to execute such a simulation. However, with the recent advent of general-purpose graphics-processing unit (GPGPU) acceleration, more extensive simulation can be performed to account for manufacturing variation, rather than just checking for the nominal (perfect) manufacturing conditions.

MB-MV offers full-chip simulation that can find mask hotspots with 1-2nm precision. GPGPU-accelerated simulation enables full mask simulation within the time it takes to write the mask. MB-MV checks the nominal contours of every shape on the mask, each in its unique context.

A unique and critical part of the MB-MV mask simulation technology is DM simulation. To be truly accurate, a simulation-based methodology must not only check for nominal contour errors, but also for random manufacturing variation in mask contours. Using DM detection enables MB-MV to predict a range of mask contours. This process is analogous to OPC process window calculation. This DM detection technology flags suspected mask hotspots (see figure 2).

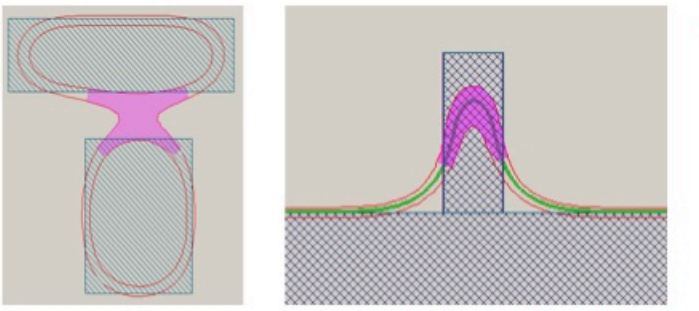

Figure 2: Dose margin (DM) simulation is a critical part of MB-MV technology, as it enables detection of manufacturing variation in mask contours.

Very soon, full-chip MB-MV will be a requirement in the mask shop. For those mask shops that employ MB-MDP, MB-MV is part of that larger process, and thus is completed before mask writing, which maximizes time and resource savings (see figure 3).

Figure 3. For those mask shops that employ MB-MDP, MB-MV can be completed before mask writing to maximize time and resource savings.

However, the majority of mask data today is prepared using conventional fracturing methodologies. As with any verification step, the number of correctly identified hotspots found by MB-MV and the rate of false positives (identified hotspots that are found not to be hotspots) will dictate the viability of adoption of the methodology. If the current mask-making methodologies for conventional fracturing largely produce successful wafers and chips, insertion of any new step that adds to the overall turnaround time of the mask writing process may be unacceptable to a production mask shop.

Fortunately, when conventional fracturing is employed, MB-MV can be added to the current mask inspection flow as a parallel process during mask writing, without introducing any new risk into the existing mask making process. This parallel insertion, which provides a fail-safe mechanism that enables mask makers to "stop the presses" when hotspots are identified, may be a good option for providing full-chip verification in situations where a change to the main flow is considered unacceptable (figure 4).

Figure 4: MB-MV can be added to the current mask-inspection flow as a parallel process without introducing any new risk to the existing flow where a change to the flow is considered unacceptable.

In this parallel process, even if a mask must be re-written, the mask hotspot will be caught before mask processing and before the mask leaves the mask shop – far less expensive than if the hotspots cause wafer defects downstream.

In other situations, such as for production of masks that are known to contain complex shapes, non-orthogonal shapes, or small features such as narrow sub-resolution assist features (SRAFs), a mask shop may choose to insert MB-MV prior to the mask writing step (figure 5).

Figure 5. In some situations, mask shops may choose to insert MB-MV as a new step in the mask-making flow prior to mask writing.

Clearly, it is better to automatically fix all potential hotspots, rather than merely to identify them. And it is even better to have all potential hotspots eliminated by a correct-by-construction approach. However, the first step is automatic full-chip detection, and MB-MV represents that step.

Conclusion

Starting at the 28nm node, mask hotspots are an increasing problem for the semiconductor industry. Current mask inspection methodologies are only accurate for mask shapes to 10nm at best, and allow mask hotspots to escape the mask shop and have a negative impact on downstream wafer processing. MPC can remedy some parts of variation in nominal contour, but the problem of poor resilience to manufacturing variation must be detected and resolved in the mask shop. Full-chip mask verification is the only way to prevent mask hotspots from leaving the mask shop undetected.

MB-MV is an existing technology that can be applied today to this emerging problem. GPGPU-accelerated MB-MV finds many of the potential hotspots, and can simulate each shape in its unique context. Critically, MB-MV includes DM detection, so it can check not only for nominal contour but also for random manufacturing variation.

MB-MV can be deployed initially as a parallel process during mask writing, without any disruption to existing flow. This low-risk methodology enables mask makers to find hotspots before mask processing, saving both time and production costs. In addition, it eases the need to create new, stricter design rules to address the hotspot problem, and enables designers to use the shapes they need to use without fear of creating an unidentifiable hotspot on the mask. Adding automated repair of the hotspots is a natural next step. But ultimately, the best flow is a correct-by-construction approach using MB-MDP.