The Need for Multiple Alternatives for sub-20 nm Lithography

Moshe Preil
GLOBALFOUNDRIES

Introduction

Over the past 30 years, minimum feature sizes for advanced integrated circuits have been reduced by over 100X, from several microns in the early 1980s to just 32 nm today, with plans for sub-20 nm production firmly in place at most major manufacturers. While predictions abound that the relentless pace of geometry shrinks will have to end at some point, it seems clear the industry is well on target to continue shrinking geometries to less than 10 nm before the end of the decade. What is less clear is which lithographic technology will be used in producing these leading-edge devices. There is currently no single solution that is proven to meet both the technical and economic requirements for volume manufacturing below 20 nm.[1] Roadmaps exist for multiple competing solutions, but the process of selecting a viable manufacturing solution takes far more than a roadmap.

In the past, key technology inflection points have been decided by an industry-wide consensus, with one clear winner emerging from among a group of competing alternatives to gain widespread adoption. In optical lithography, prime examples of this have been the adoption of 248 nm excimer laser tools, followed by 193 nm and 193 nm immersion (193i). Competing alternatives such as 157 nm, despite years of development, failed to gain even limited adoption. In the technology arena of next-generation lithography (NGL), only extreme ultraviolet lithography (EUV) has survived as a valid candidate for HVM insertion. e-beam projection (EPL), optical maskless (OML), 1x X-ray and ion beam lithography (IBL), to name a few, have fallen off the roadmap over the past decade.

It would be tempting to conclude that history shows there can be only one winner at each inflection point, and that single chosen technology must therefore be adopted by all segments of the industry that wish to stay on the path of Moore's Law. As we look to the future below 20 nm, however, many in the industry feel that "one size fits all" is no longer a viable model for all use cases. Today we see numerous competing options that are favored by one segment or another, and a growing number of lithography strategists have begun to consider the real possibility that different choices will emerge as being best for different applications. Instead of

one size fits all, we are seeing the beginning of custom-tailored solutions. In this paper, we will explore the economic and technical reasons for this change and propose that the industry can and should be more open to multiple complementary solutions rather than the winner-take-all approach of previous technology cycles.

Historical Survey of Past Inflection Points

To gain some perspective on the evolution of lithography options, it is worth taking a brief survey of past industrywide transitions (Figure 1). While the exact dates when a specific technology was first used for high-volume manufacturing (HVM) and when that technology became the dominant choice are always subject to debate, the general trend is clear. At each major changeover point, one and only one new technology successfully transitioned from lab to fab to become the industry-wide technology of choice for leading-edge production.

For every roadmap winner, however, there have always been multiple technologies that were not adopted by any sizable segment of the industry. Between the introduction of 193 nm dry and immersion lithographies, extensive development resources were invested in 157 nm. Scanners were designed and constructed, entire factories were built to produce lithographic grade optical materials, and new fluorine-based resist chemistries were developed before

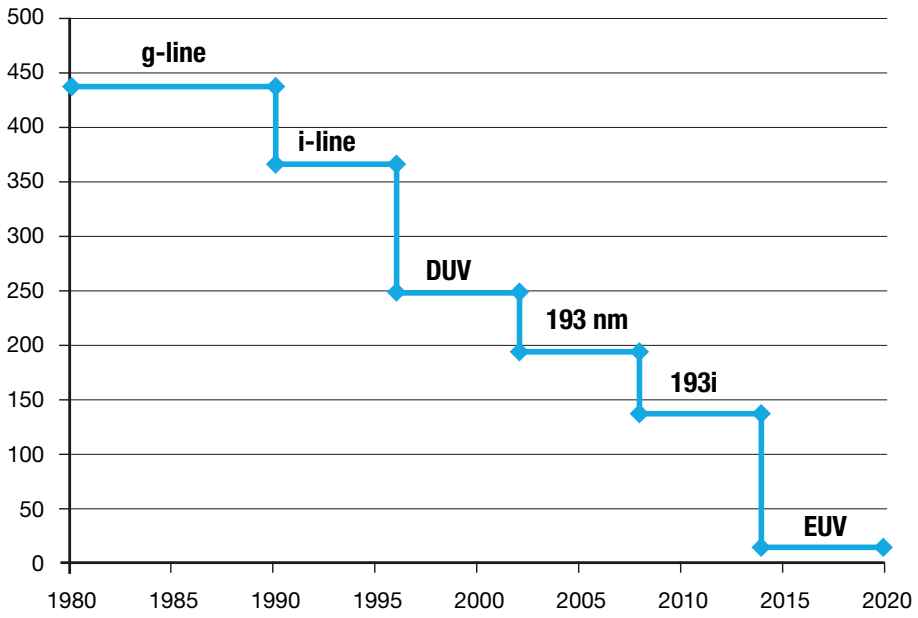

Figure 1. Trend in Leading-Edge Lithography Wavelength Vs. Time

the technology was abandoned in 2003. Several years later, high-index 193 nm immersion was under serious consideration before being shelved in late 2008. Finally, multiple alternatives to EUV were considered and actively debated in a series of industrywide NGL forums and their successors from 1997 to the present. Despite strong advocacy by the champions of each technology and years of impressive presentations and data, only EUV has survived to stand today at the threshold of insertion into manufacturing.

This trail of technology roadkill would seem to suggest there are fundamental reasons the industry can only support one new lithography option every five to seven years. If this "rule" were applied to the 2013-2020 time frame, it would seem to imply that only EUV can succeed as a new HVM technology for the rest of this decade, given how far along EUV is on the road to commercialization and how far any alternative is from production readiness. The only other option would be to stay with 193 nm immersion and move to three or more masks per layer (pitch splitting), which would drastically increase the cost of manufacturing due to the proliferation of masking steps, critical masks, and reduced overlay and CD control budgets per exposure. It would also vastly complicate the task of circuit layout and design, making it inaccessible to many end-users and sharply reducing the available market for new designs.

Before accepting this as an inevitable conclusion, it would be instructive to look beyond simple extrapolation and ask, "What are the underlying reasons for this past experience, and do these reasons still hold true for today's competing technology options"? The key points that doomed past technology also-rans were the lack of adequate infrastructure and the reluctance of the industry to fund that infrastructure due to limited extendibility of the technology. Without a viable ecosystem to deliver every needed element of the new technology, building an exposure tool alone proved to be insufficient, and funding for the tools was eventually terminated. Table 1 lists the most critical missing links for some of the abandoned options of the recent past.

There are, to be sure, those who would argue that EUV still faces daunting infrastructure challenges of its own; sources with adequate power to enable high-throughput operation, defect-free mask blanks, a proven vacuum-compatible mask handling and storage system to keep the mask clean over time without a pellicle, and affordable actinic wavelength mask inspection capability. In fact, the huge investment being made to develop a viable EUV infrastructure is often cited as a major reason for the lack of funding for other potential candidates. The frequently heard comment, "If we only had 10 percent of the money that was invested in EUV we could have solved all of our problems by now," has been heard from more than one competing technology champion. Nonetheless, the fact remains that EUV is firmly on the path to commercialization. It has literally become too big to fail.

But this does not mean that EUV will be the only choice for the next several nodes. It only means that any other technology must not require any major new infrastructure investment. In this respect, we must consider the EDA world and design constraints as part of the technology ecosystem that would need to be reused. In the following sections, we will consider the cost/benefit trade-offs of several competing options and demonstrate their viability not as direct competitors to displace EUV,

but rather as complementary tools to create cost-effective solutions for different segments of the industry.

Alternatives for sub-20 nm Lithography

The 2010 ITRS roadmap[1] predicts the adoption of sub-20 nm technology in 2016, with minimum half-pitches of 23 nm for DRAM, 19 nm for MPU and ASIC metal layers and 16 nm for flash memory. The options still listed as being under consideration are 193 nm immersion with multiple patterning (DPT), EUV, imprint, direct write (often referred to as maskless lithography, or ML2) and 193 nm immersion with directed self-assembly (DSA). The leading option for ML2 remains e-beam direct write (EBDW) although optical and hybrid opto-electronic tools have also been proposed. As usual, major IC manufacturers are already predicting they will begin production of this node significantly ahead of the roadmap and/or at even smaller geometries.

The design constraints imposed by different industry segments play a critical

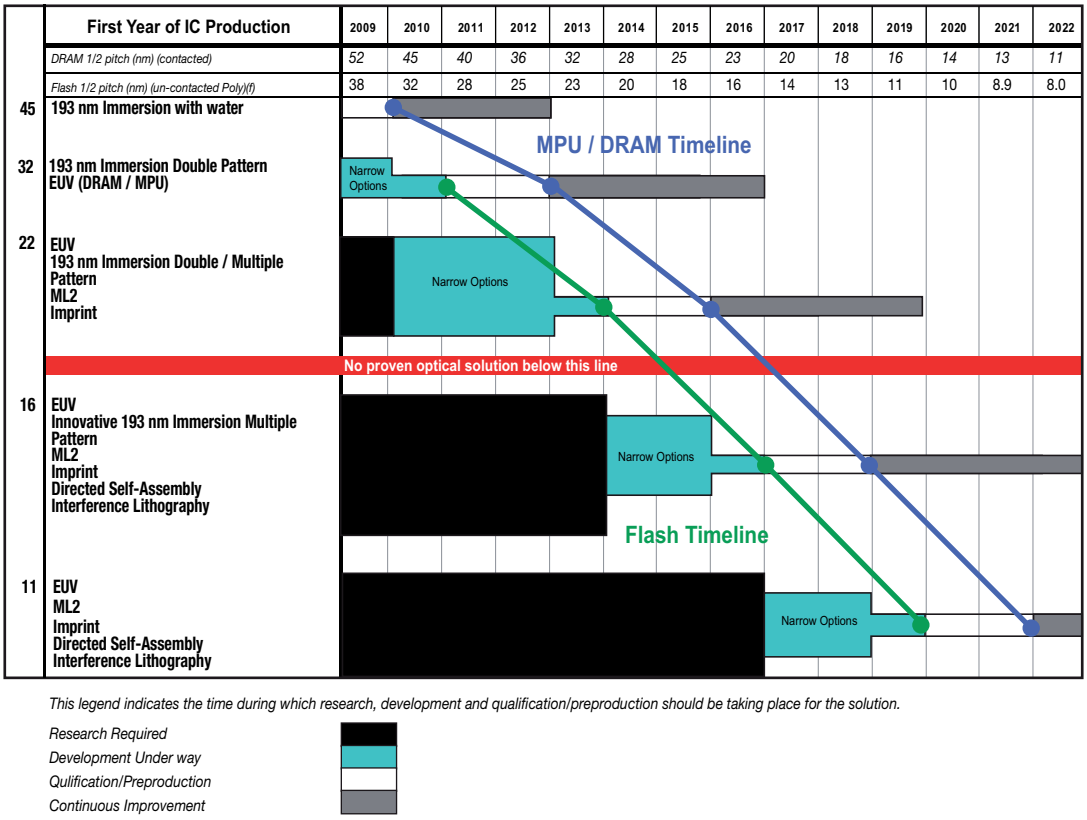

Figure 2. ITRS Roadmap for Semiconductor Lithography, 2010 Update, Lithography Exposure Tool Potential Solutions

role in developing and implementing a cost-effective lithography strategy. Memory can push to the tightest possible pitch and take advantage of novel techniques that print perfectly regular arrays of lines and spaces. Integrated device manufacturers (IDM) with a small number of high-volume, high-value products such as microprocessors have the ability to impose very restrictive design rules and work closely with the designers to co-optimize the process and design. In the foundry arena, the huge number of designs and customers requires the most robust processes that will yield working die for thousands of different products running many different flavors of the base process.

In addition to adopting different levels of layout constraints and design-process co-optimization, these industry segments are also under different levels of economic pressure in dealing with the soaring cost of masks. A multimillion-dollar mask set that produces millions of high-value CPU or PLD chips adds only a small percentage of the cost per die; that same mask cost adds an unsupportable burden to the cost per die of a low-volume ASIC. On the other hand, reducing throughput by moving to maskless lithography or multiple patterning may be cost-effective for low-volume parts but would be unable to manufacture memory or CPU chips in sufficient volume. Memory and logic also diverge in terms of defect tolerance. The built-in redundancy in memory chips allows more compromise in terms of mask and wafer defectivity than complex logic devices.

The issue of mask costs is further complicated by the low percentage of masks that actually result in profitable, high-volume device production. Actual data on wafers produced per mask is notoriously well protected for commercial reasons. One study,[2] published in 2002, is shown in Figure 3. This data showed the mask usage dilemma with striking clarity:

- Over 50 percent of the masks produced only 1 percent of all wafers in the foundry environment.
- The mean number of wafers per mask was roughly 700. However, the median number was even lower, at barely 100 wafers per mask. It is the median number that truly reflects the severity of the problem since the mean is inflated by a very small number of highly successful devices.
- Over 80 percent of the wafers were printed by just 10 percent of the masks; 65 percent of the wafers were printed

157 nm	High-Index 193i	EPL	1X X-ray
Pellicle issues (hard pellicle required)	High-index fluids	Fragile membrane masks	1X membrane masks • Defect, overlay and CD requirements at 1X • Membrane mask distortion • Fragile • Seemed viable when first proposed for 1 micron, could not meet tighter specs < 250 nm
CaF_2 supply issues	High-index glass material • Availability in sufficient quality and quantity	Alternate: stencil masks • Also fragile • Cannot print isolated solid structures	
No immersion medium (water doesn't work) • Limited extendibility	High-index resist materials	Mask heating and distortion issues at required throughput	

Table 1. Key Infrastructure Issues in Previous Unadopted Technologies

by only 5 percent of the total masks. While these are raw numbers with no weighting for the sales price per chip, it is clear that most of the value of IC production comes from only a small handful of high-volume "winners." The vast majority of masks never pay for themselves.

While this specific data set represents only one foundry 10 years ago, the general trends still hold true in the foundry world today. Even for IDMs, limited data that can be inferred from worldwide mask usage shows that a significant fraction of initial designs fail to reach volume production. Re-spins of entire mask sets are not uncommon, and certain layers may undergo numerous revisions. Given that the cost of defect-free EUV masks is still unknown, and that the alternative requires multiple masks per process layer, it is readily apparent that the current mask usage scenario is not economically viable in the sub-20 nm world.

The mask usage plot in Figure 3 clearly suggests that we should focus on three distinct regimes: high-volume production of a few leading designs, where mask cost is averaged out over many wafers; low-volume prototyping, where ML2 can produce a few dozen wafers in a cost-effective manner; and an intermediate regime, where a cost-effective solution will need to balance the relative expense of EUV and ML2 with the availability of new solutions such as directed self-assembly (DSA) and ultra-regular layouts. For each of these three basic mask usage regimes, the winning option is not simply the one that costs less. If all options are too expensive, there is a third – albeit unpalatable – choice: Customers will simply stop designing as many devices at the leading-edge node, and large segments of the IC indus-

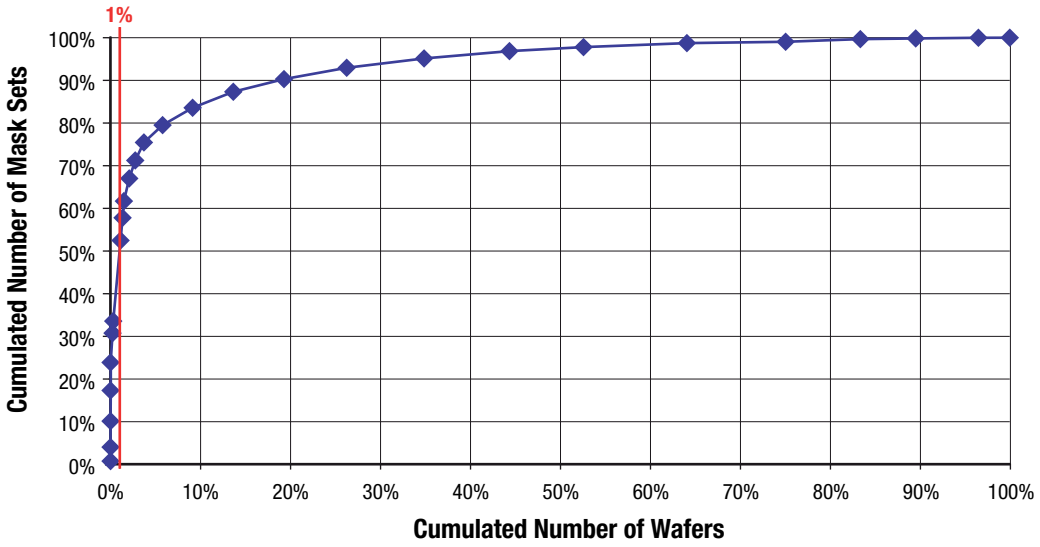

Figure 3. Cumulative Number of Wafers Printed Vs. Number of Mask Sets
Courtesy of Kevin Cummings, ASML[2]

try will transition from high growth to mature, steady-state businesses. To prevent economics from bringing Moore's Law to a premature end, we must consider how different strategies can be optimized for each of the three mask usage regimes and not force one solution onto three use cases with very different cost models.

The Case for Complementary Technologies: ML2 + EUV

Maskless lithography has often been proposed as a replacement technology for expensive mask-based approaches,[3,4] but the low throughput of any proposed ML2 exposure tool makes this impractical for high-volume production. As long as we regard ML2 and EUV as competitors in a winner-take-all battle, there can be no clear winner. An obvious solution to this dilemma is to develop a moderate-throughput, low-cost EBDW tool that is a complement, not a competitor, to the more expensive, high-volume optical and EUV exposure tools.

While the concept of using ML2 for prototyping has been proposed numerous times in the past, the maskless exposure tools available to date have not met the resolution and throughput requirements even for this specialized use case. Single-column direct-write e-beam tools currently produce one to two wafers per day, not the several wafers per hour required. Several tools now under development show promise for delivering five-20 wafers per hour (WPH), which would be adequate for a low-cost prototyping tool. Even at just five WPH, a $20 million ML2 tool would be more cost-effective than a larger,

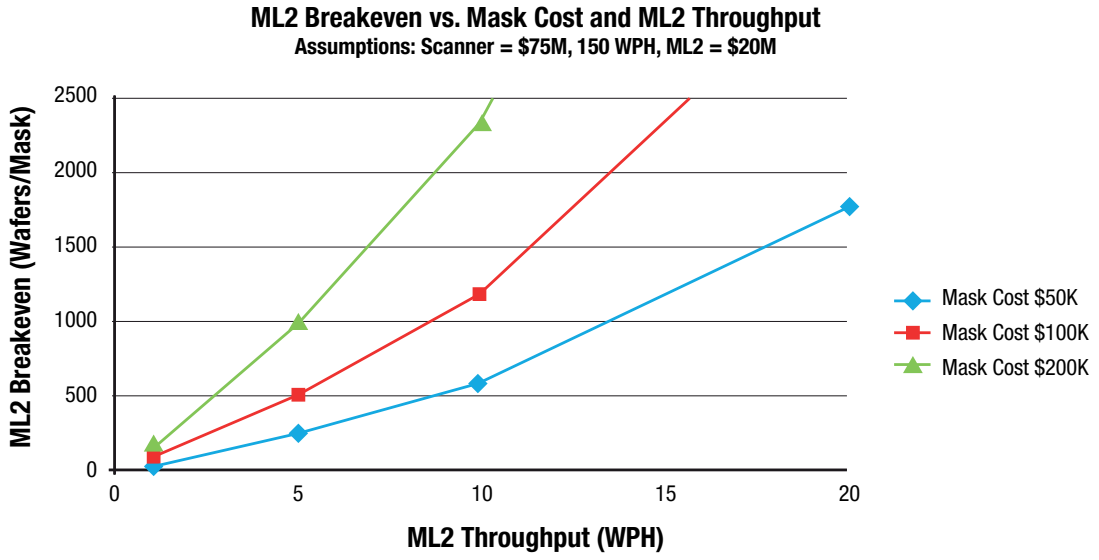

Figure 4. Maskless vs. mask-based breakeven as a function of EUV mask cost and ML2 throughput. Assumed tools costs and EUV throughput are fixed in this example.

more expensive EUV tool running 10-20X the throughput with a mask cost measured on the order of $100,000.

Figure 4 shows a sample trade-off between ML2 and EUV over a range of assumptions for ML2 throughput and EUV mask cost; other parameters that were fixed in this example are listed in the figure. The key performance indicator plotted here is the breakeven number of wafers per mask. If the mask is used for more than this number of wafers, mask-based lithography is cheaper; less than this value and ML2 is more cost-effective. This simple model shows that a 10 WPH, $20 million maskless tool would reach breakeven with a $75 million 150 WPH EUV tool and a mask cost of $200,000 at about 2,300 wafers per mask. Clearly we would not bother to incur the expense of maskless prototyping if the actual number of wafers/mask was even close to the breakeven point. However, as shown in Figure 4, a very large number of masks are used less than one-tenth as many times, making maskless prototyping a significant cost savings strategy. The cost of equipping a fab with several small, low-cost ML2 exposure tools will be offset many times over by the huge cost savings of not producing hundreds of expensive EUV (or DPT) masks that are used a limited number of times and discarded.

A key factor in this complementary scheme is to recognize that the maskless exposure tool must mimic the patterning performance of the high-volume tool. This is a non-trivial constraint that has not been widely applied to maskless tools in the past. Rather than trying to push the ML2 tool to print the squarest line ends and contact holes possible, care must be taken to emulate the actual rounding and proximity effects that will occur in the optical process once the design passes the prototype phase and moves to high-volume production. This may seem to be imposing an unfair burden on the ML2 tool that would not be necessary in a purely maskless application, but it is a critical requirement for a successful complementary strategy. Early discussions with several candidate suppliers have shown that this constraint can be met through extensive simulation, data handling and beam control computation currently under development.

The Case for Complementary Technologies: Ultra-Regular Layouts

Another emerging opportunity for multiple technologies to work together is to use a mask-based tool to print an ultra-regular 1D array of lines and spaces at a single pitch, then follow up with a maskless tool to cut the lines into a usable device layout (see Figure 5).[5] This concept has been discussed for close to 10 years but, as with many novel ideas, has not been widely adopted yet due to the relentless progress of optical lithography. With optical now running up against the physical limits of numerical aperture and wavelength, as well as the economic limits of mask and tool costs, this hybrid approach has gained significant traction, especially since it was promoted by Yan Borodovsky of Intel at SPIE in early 2010.

The initial grating pattern can be formed by several different technologies. 193i with some type of frequency multiplication – either multiple patterning technology or directed self-assembly – would enable us to print gratings as small as 15 nm half-pitch. The first generation of EUV production tools could lower this to just 14 nm, and EUV plus multiple patterning or DSA might well take us under 10 nm. Bear in mind that these are actual grating pitches, not node

names, and could therefore meet the requirements of logic device manufacturing as far as the 6 nm node in 2021.

The clear advantage of such a complementary approach is the dramatic reduction in the number of masks needed. In the mask usage plot (Figure 2), the cut layer approach would help in all three regimes:

- **Low-volume prototyping** – this approach would eliminate a large percentage of the masks wasted due to design re-spins. Even if the initial device does not work, the redesign could be done entirely by modifying the direct-write cut layers without requiring any new masks.
- **High-volume regime** – this approach would require fewer masks for layers requiring multiple patterning.
- **Mid-volume region** – it would be preferable to use 193i and extensive cut masks rather than absorbing the cost of building EUV masks for devices that are not expected to be manufactured by the millions. It is even conceivable – though admittedly a remote option - that multiple designs could be close enough in terms of die size to share a single grating mask and simply trim the die differently.

On the other hand, implementing ultra-regular layouts shifts the burden from mask costs to design constraints. While IDMs may still have the luxury of large design and process teams working closely together to customize each design and cell, in the fables-foundry model, this is not a supportable way of working. The ability of the electronic design automation (EDA) community to implement software tools that can lay out any required cell on a purely one-dimensional array without an excessive die size penalty will be a key enabling technology for the adoption of cost effective complementary lithography.

e-beam Direct Write Considerations for Complementary Lithography

One of the critical limitations of EBDW is the finite number of electrons that can be generated and moved through the electron optics without excessive beam blur due to Coulomb repulsion.[7-9] At the larger geometries used for mask making, this limitation has previously been met by using more sensitive resists with target doses in the range of 5-20 μC/cm^2. For wafer-level geometries, however (and even for leading-edge masks), it is no longer possible to use such fast resists

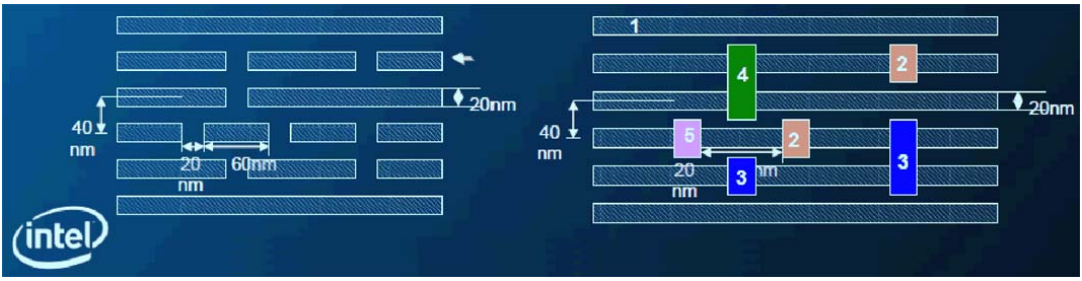

Figure 5. Complementary lithography using ultra-regular arrays and cut masks.[5] Four separate masks would be required if the cuts were all done with 193 nm lithography. This can be reduced to one mask using EUV or none with direct write.
Courtesy Yan Borodovsky, LithoVision 2010, February 21, 2010 San Jose, CA, USA

due to shot noise. For a resolution element, or pixel, of decreasing size, the number of electrons that physically strikes the resist is disturbingly low and gets lower as 1/(pixel size) squared (see Figure 6). In the simplest model, ignoring resist effects and beam blur, the shot noise is proportional to 1 over the square root of the number of electrons. For example, for a 20 nm pixel, fewer than 500 electrons are required to expose one pixel at 20 μC/cm^2; the shot noise is over 4 percent. A 10 nm pixel receives just over 100 electrons with a shot noise of 9 percent. It would clearly be impossible to maintain any reasonable CD control at these noise levels.

The only way to mitigate this shot noise limitation is to move to slower resists, which would result in a substantial throughput penalty. This can be offset by using more beams to write the pattern, but only if an electron source is available to provide sufficient current to all of the beams. Without brighter sources, the benefits of parallelism would not be achieved. But since electrons are charged particles, putting more exposing particles and more beams through the electron optics would result in more blur and an unacceptable reduction in resolution.

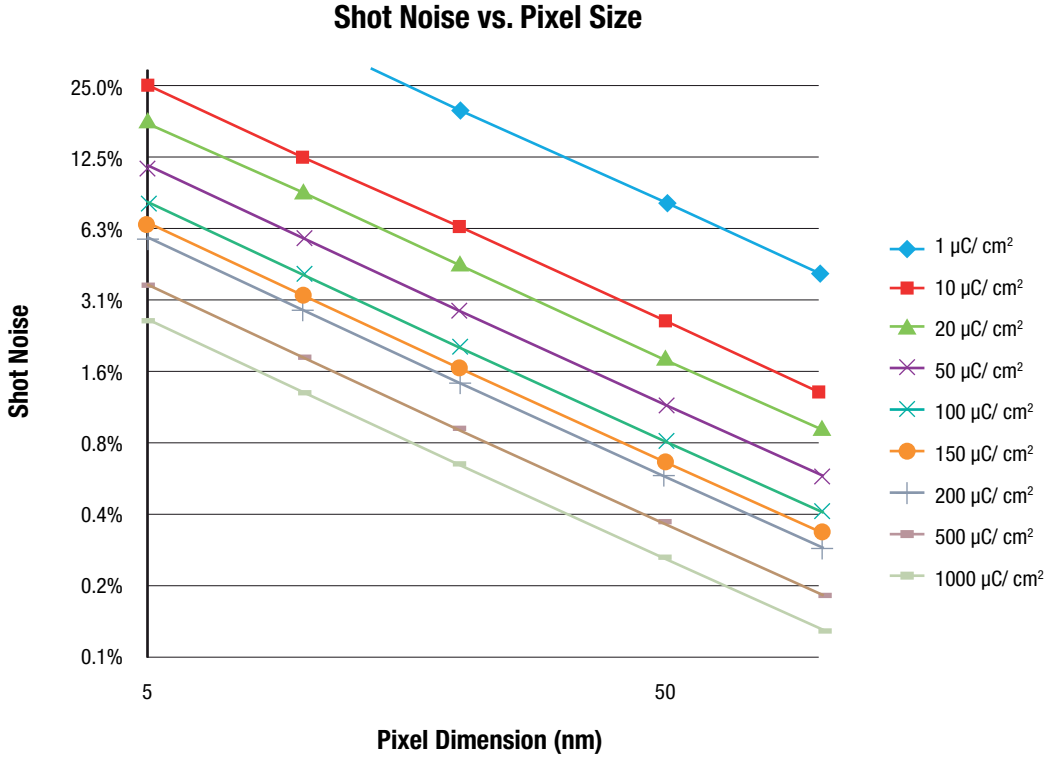

Figure 6. Simplified shot noise model vs. pixel dimension as a function of exposure dose. The model does not include resist and beam blur effects.

The cut mask approach is particularly well suited to a single-column, multiple-beam design. Since only a few pixels would be written simultaneously, the maximum beam current allowed would be divided by relatively few beams, enabling higher dose writing while maintaining adequate throughput. This particular design advantage has been recognized by KLA-Tencor in their proposed REBL system.[10] The throughput of this tool for writing a sparse pattern such as a cut mask could be up to 5X higher than trying to write a dense array of line/space patterns. A competing concept embodied by the MAPPER system[11] uses one beam source but many micromachined columns. The current per column is thus kept at a manageable level for any pattern density. The raster scanning approach required for a multicolumn array provides the same throughput regardless of pattern density. Such trade-offs between system design, cost and intended use case will become increasingly critical as these technologies get closer to production. It is not unreasonable to think they will each play a role in different applications.

Conclusions

While EUV will be the mainstream technology of choice for the highest-volume products below the 20 nm node, other technologies will be necessary to complement EUV in a cost-effective mode. Direct-write technologies and ultra-regular 1X arrays patterned by DSA, multiple patterning or EUV will all be part of the mix. Rather than engaging in fratricidal competition to determine a single winner, it would benefit the industry to put more effort into making these different technologies work together to deliver cost-effective solutions for all segments of the industry. Developing complementary lithography technologies will require the development of new exposure tools, new EDA tools and new strategies to mix exposures between these different tools. While this new way of working may not be as appealing to investors hoping for a winner-take-all windfall, the alternative could well be the end of affordable scaling and a lose-lose situation for suppliers and IC manufacturers alike.

Endnotes

1. International Technology Roadmap for Semiconductors
2. K. Cummings, “A business perspective for maskless lithography,” BACUS Panel Session, September 2003.
3. R. Pease et al. “Distributed Axis Fixed Aperture (DIFA) Electron Beam Technology,” International SEMATECH Lithography Forum, January 2004.
4. T. Chang et al. “Electron Beam Microcolumns for lithography and related applications,” J. Vac. Sci. Technol. B14(6), 3774 (1996).
5. Y. Borodovsky, “ArF Lithography Extension for Critical Layer Patterning,” Nikon LithoVision seminar, 2010
6. This value is used for illustration only and is not intended to represent any specific exposure system.
7. A. Neureuther et al. “Shot noise models for sequential processes and the role of lateral mixing,” J. Vac. Sci. Technol. B 24, 1902 (2006).
8. P. Kruit et al. “Optimum dose for shot noise limited CD uniformity in electron-beam lithography,” J. Vac. Sci. Technol. B 22(6), 2948 (2004).
9. N. Rau et al. “Shot-noise and edge roughness effects in resists patterned at 10 nm exposure,” J. Vac. Sci. Technol. B 16 (6), 3784-3788, (1998).
10. P. Petric et al. “New advances with REBL for maskless high throughput

EBDW lithography," Proc. SPIE 7970-43 (2011).
11. M. Wieland et al. "MAPPER: High throughput maskless lithography," Proc. SPIE 7637-14 (2010).

Acknowledgments

The themes captured in this paper have evolved over time and benefited greatly from frequent discussions with the members of the Strategic Lithography Technology group at GLOBALFOUNDRIES, especially Harry Levinson and Keith Standiford. The author would also like to thank Yan Borodovsky of Intel, Kevin Cummings of ASML, and SEMATECH for permission to reproduce their published results.

About the Author

Moshe Preil

Moshe Preil is the manager of Emerging Lithography and Tools, part of the Strategic Lithography Technology department at GLOBALFOUNDRIES. He has over 25 years of experience in both the fab and supplier sides of the industry, working in various lithography, process control and yield management positions. Previously at AMD, Dr. Preil was involved in the early development of deep-UV technology. He earned his Ph.D. in physics from the University of Pennsylvania, working on optical and electron spectroscopies of novel graphite-based compounds. Dr. Preil has published numerous journal papers and has been issued nine U.S. patents. ■