E-beam direct write is free

Lance A. Glasser[1]
Saratoga, CA 95070

ABSTRACT

In this paper we discuss four business concepts that will impact the adoption of e-beam direct write (EbDW). They are: (1) The economically advantageous region for EbDW. At what costs and volumes EbDW is economically advantageous is controlled by a two-sided constraint involving the cost of reticles on one hand and the cost of design on the other. (2) The important role of product derivatives and other markets that can be satisfied by designs with heavy IP reuse. The natural long tail in demand for differentiated products is today chopped off by the high costs of reticles. We show data on the elasticity of the product derivative market with respect to certain costs. (3) That because reticle prices typically decline at a 30% per year for the first few years after a new node is introduced, delaying the fabrication of that first reticle set for a new product can save millions, more than paying for EbDW. The applicability of this technique is, however, limited by the need for product requalifaction. (4) Finally, we introduce the business concept of the virtual reticle as a possible component in EbDW pricing.

Keywords: e-beam direct write, maskless, lithography, prototyping, virtual reticle

1. INTRODUCTION

At a lithography tool capacity of a few critical wafer-levels per hour, e-beam direct write (EbDW) can have a significant place in the lithography tool kit. Despite the fact that EbDW has been a "technology of the future" for 40 years—it was mentioned as such in Gordon Moore's first paper[1] on Moore's Law—the challenges and costs of multi-exposure optical and EUV lithography have given new impetus to EbDW efforts[2]. E-Beam Direct Write has a long history in the development of semiconductor research and manufacturing. The clear leader in this area during the 1970's, 80's and into the 90's was IBM, who championed the development of Variable Shaped Beam (VSB) machines[3]. In their QTAT lines, the EL-1 reached a throughput of 22 wafers per hour. Later, IBM manufactured, installed, and ran over 30 EL-3 machines 24/7 to produce the many different part numbers needed for their bipolar based mainframes. During this time, the effective pixel writing rate was increased by over two orders of magnitude. However, the decrease in feature size with the increase in wafer size combined to drive up the effective pixel count for directly writing wafers by over 1000. Moore's Law outpaced the ability of EbDW to maintain throughput. A number of throughput improvement ideas are being implemented for production use within the decade. It is hoped that 45nm 300mm write times better than one wafer layer per hour are within reach. In the twilight of single exposure 193 nm optical lithography, EbDW techniques are again a topic of huge interest. This paper deals not with the technical advances in the field, but rather with some of the business questions assuming that one or more of the EbDW technologies that are now being pursued can be realized.

In this paper we discuss four business concepts that will impact the adoption of EbDW. They are

- The economically advantageous region for EbDW. At what costs and volumes EbDW is economically advantageous is controlled by a two-sided constraint involving the cost of reticles on one hand and the cost of design on the other.
- The important role of product derivatives and other markets that can be satisfied by designs with heavy IP reuse. The natural long-tail in demand for differentiated products is today chopped off by the high costs of reticles. We show data on the elasticity of the product derivative market with respect to certain costs.
- That because reticle prices typically decline at a 30% per year for the first few years after a new node is introduced, delaying the fabrication of that first reticle set for a new product can save millions, more than paying for EbDW. The applicability of this technique is, however, limited by the need for product requalifaction.
- Finally, we introduce the business concept of the virtual reticle as a possible component in EbDW pricing.

---

[1] www.lanceglasser.com

2. THE ECONOMICALLY ADVANTAGEOUS REGION

The cost of producing an IC is generally split between the fixed cost of design; which includes the design team, the cost of EDA tools and the computers needed to run them and any third party IP, the fixed cost of the reticle set, and the variable cost of chip manufacturing. Because of its size, the cost of design is one of the more important parameters in any analysis of the return on investment for an integrated circuit. Typically, a business will need on the order of a factor of ten return on design cost to have a sustainable business. If we call the cost of design D and the ROI factor R, then the integrated circuit revenue must be RD. If the selling price of the integrated circuit is p, then RD/p good chips must be sold. Let Y be the total yield, a be the area of the chips, and let's approximate the useable area on a 300 mm wafer as 700 cm^2. In this case, the number of wafers needed for a design to be economically feasible is aRD/700pY, where a is measured in sq cm. Since design costs D tend to go up superlinearly with chip area, it is actually often smaller die that need fewer wafers to be economically feasible, even though the chip price also tends to go down with area. The mega-design projects such as microprocessors or graphics processors are the least applicable to EbDW production since so many wafers are needed to make the project economically advantageous. (Prototyping for systems testing is, on the other hand, applicable to even these areas.)

The major manufacturing cost components are broken up into depreciation cost, which includes the depreciation of the exposure tool and the tracks; operating cost, which includes the cost of power, maintenance, and a fab cost calculated on equipment footprint; and finally mask, resist and developer cost. Also, machine utilization rates must be included in any realistic cost analysis. Here, they are projected with a high of 80% at 10000 wafers per design, but that is brought down to a 55% utilization rate at 10 wafers per design. This utilization rate is much closer to that seen in mask manufacturing[4], where every mask is a design with a volume of one and thus has the highest penalty for setup costs.

Table 1 shows the major parameters for different lithography options. The capital costs for current tools have been taken from conversations on the street price. Some estimates are made for next generation EbDW capital costs. Machine throughputs for optical systems are from the manufacturer's literature, while throughputs for the EbDW machines are estimated from simulations of emerging techniques. For mask costs, we estimate the cost of the most complex layers at 65nm and 45nm to be $250K and $333K respectively. More details on the modeling itself can be found at[5].

By factoring in these costs on a per wafer per layer basis, we can look at the relative effectiveness of the techniques. Figure 1 shows the results of the cost model on a log(cost) vs. a log(wafer volume per design). The motivation for using a log-log plot is that power laws are shown as straight lines. In the low volume per design regime, the total cost is dominated by the fixed cost of a reticle set, i.e., (constant total cost) = (wafer cost)* (number of wafers), or in another way (wafer cost) = constant / (num wafers) and so is a power law with an exponent of minus one. As the volume is increased, we see that the reticle costs become amortized over an increasing number of wafers. Ultimately, this reticle cost per wafer will become very small and the wafer cost will be dominated by machine amortization, operating and consumables costs, which are all at a constant per wafer cost basis and thus would show as a horizontal line. We can start to see the transition to this regime at the 100,000 wafers per design volume. Conversely to the optical systems, EbDW moves all of the cost into a volume related constant cost per wafer. There is a slight slope downward as the number of wafers per design increases and the utilization moves up, as mentioned previously.

The most interesting information occurs at the cost crossovers between the different lithography options and these are highlighted by the circles enclosing the relevant crossings. The crossover for 65 nm between ArF and EbDW happens at approximately 217 wafers. Given that a 300mm wafer can have over 1000 8mm by 8mm die, this means that production runs of over 200,000 chips can be made more cost effective by utilizing EbDW for the most complex layers than by the optical techniques. One caveat here is that these calculations assume basis-cost only, and does not give profitability to the supply chain. The dynamics will vary depending on supply and demand in the supply chain.

Fundamental Lithography Tool Parameters				
Node (nm)	Exposure System	Field Size (mm)	Throughput wafers / hour	Tool Cost Million $
65	ArF	26 x 33	85	25
45	ArF / I	26 x 33	130	40
65	VSB	N/A	0.05	20
65	EbDW	N/A	1	20
45	EbDW	N/A	1	25
45	NextGen EbDW	N/A	5	35

Table 1. Lithography Tool Parameters

The crossover for 45 nm between ArF Immersion and EbDW is at approximately 305 wafers, or over 300,000 chips. If a next generation of EbDW can attain a throughput of 5 wafers per hour, the breakeven point moves out to over 1200 wafers. This is a very reasonable lifetime production volume of 1.2 million 8mm by 8mm chips.

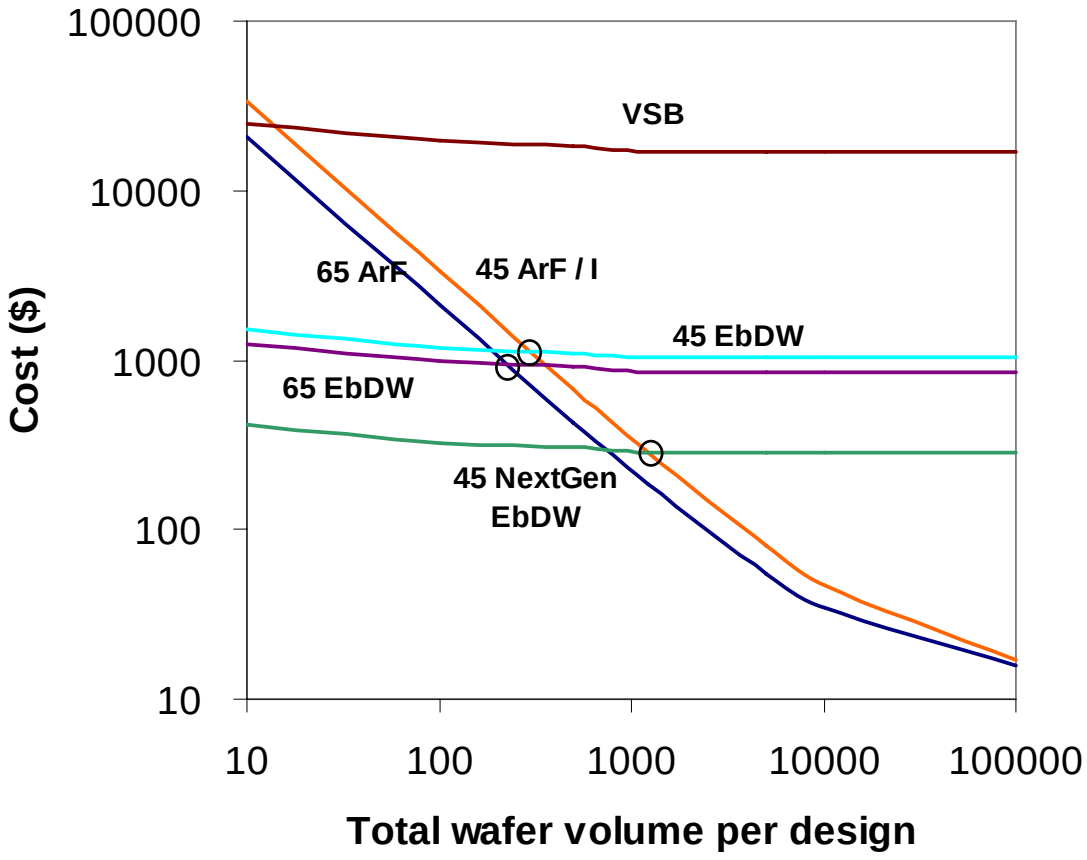
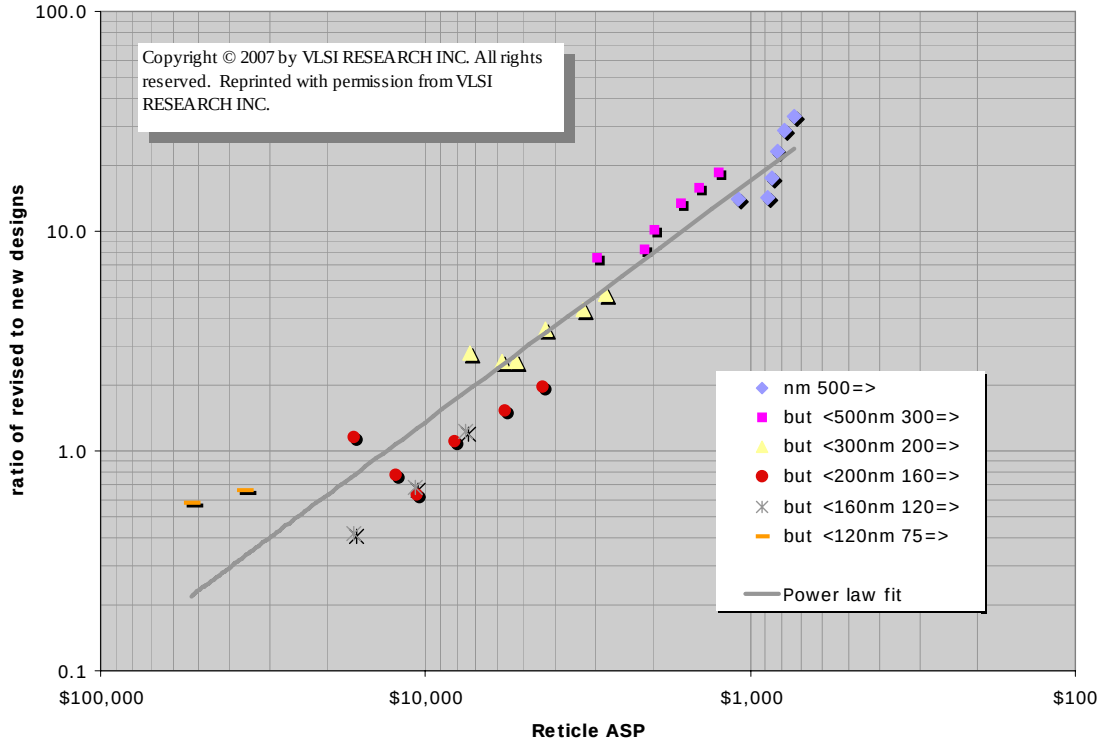


Fig. 1. Crossover between EbDW and conventional lithography

3. DERIVATIVE PRODUCTS

Since minimizing design cost D is key to making EbDW economically advantageous, we can expect that heavy IP reuse characterizes such products. One of the more important special cases is product derivatives. That is, the systems customer of the semiconductor supplier wants something "exactly like that chip except. . ." Since, from the customer perspective, differentiation can lead to competitiveness, the demand for product diversity is quite high. In addition, in only six months, standards and competitive landscape change in the rapidly moving high-tech world, making such demands frequent and opportunity great for leveraging further revenues from the initial platform design effort. If the initial people and tool cost for a platform design (say, a digital TV chip) was $30M, the cost of a derivative can often be $3M or even less. In the first year of 45nm, the mask cost could exceed the people and tool cost for a derivative design. Derivative designs are exactly how the Long Tail of SoCs will be enabled, but only if there is not an expensive mask cost getting in the way.

Fig. 2 Correlation of derived design activity to reticle cost

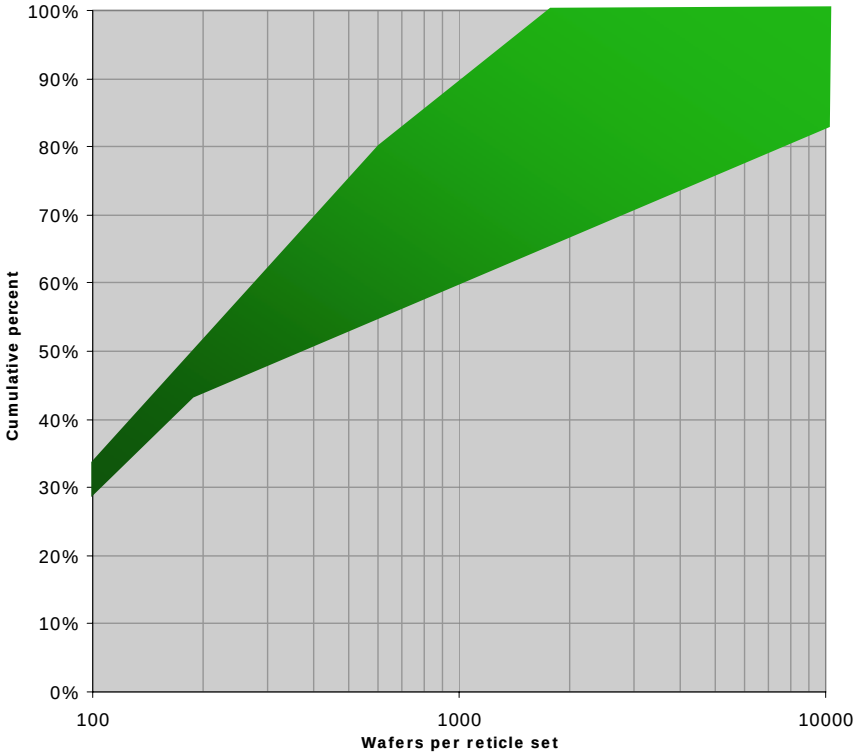


This need for a similar but different product is sometimes anticipated in the original platform design and software/firmware changes alone may be able to accommodate the customization. There are still large opportunities for semiconductor businesses that are being disabled currently by high mask costs. Higher bit bandwidth, a higher quality A/D or D/A and other analog or RF circuits, and an increase in the amount of cache to address a higher-end niche are examples of such demand. For this reason, product derivatives, where the incremental design cost is low, are a rich opportunity for EbDW. Other candidates for EbDW are anywhere that IP reuse is high and design costs are low.

That such latent demand exists is highly credible, but difficult to quantify. Examples such as multi-project chips and wafers[6] give hints and the Long Tail theory[7] supports the notion that there is a significant opportunity in the large variety of low-volume applications enabled by the low design cost of derivatives and the low manufacturing cost of EbDW. An alternative to small wafer runs is supposedly multi-project shuttles, but the reticle write time is basically the same for the design (it has the same number of e-beam reticle writer shots whether it is on a reticle alone or with other designs). Another problem with multi-project shuttle runs is the same as the reason so few people carpool. The timing is inconvenient. The most convincing data of latent demand we have seen comes from VLSI Research's data comparing new designs to revised designs, what we are calling derivative designs. In Fig. 2, we plot the ratio of revised (derived) designs to new designs as a function of reticle price. The inverse correlation could not be clearer. We interpret these results to mean that if the effective NRE costs for derived designs could be lowered, which EbDW has the potential to do, the number of designs would dramatically increase.

In Fig. 3 we see the composite results for several wafer foundries, including NEC's data[8] from 2003, on the number of wafers run per reticle set. As one can see, despite differing foundry business models, there are numerous designs where even a one wafer per hour EbDW system could be advantageous.

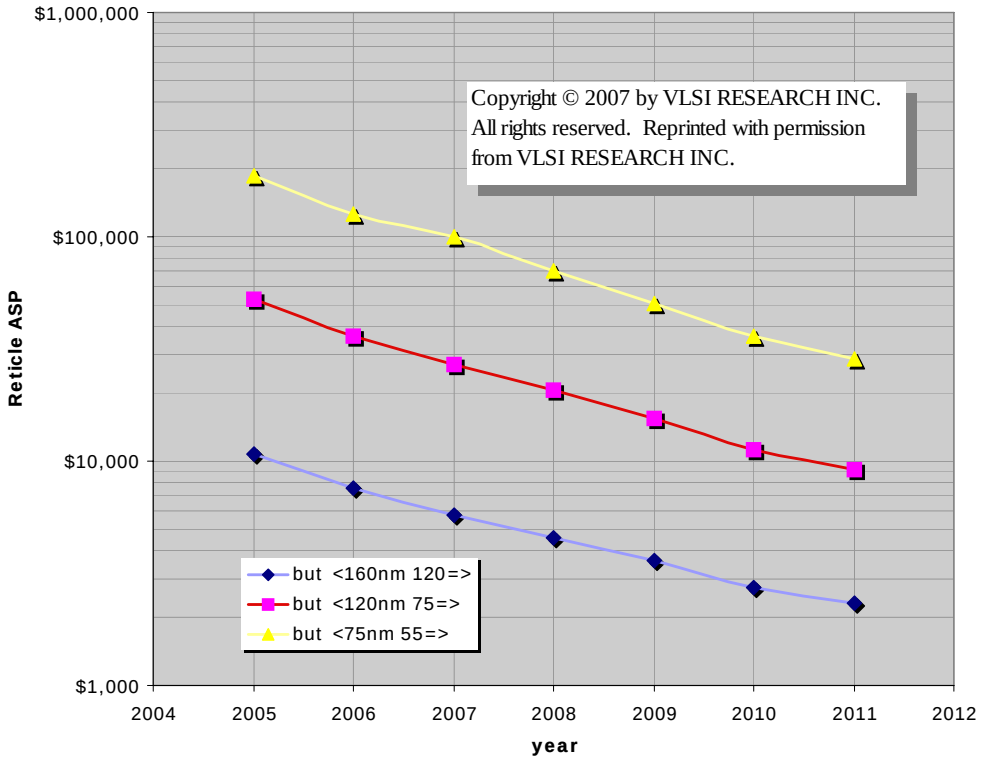
Fig. 3 Foundry wafers per reticle set



One of the more interesting, but unanswered questions relates to intent. Did the folks who eventually requested fewer than 200 wafers actually plan to make that many, or did they discover a technical or market failure that caused them to prematurely truncate production? This question of intent is important in that people might choose a lithography solution based on how many wafers they *intend* to produce. The analysis of the economically feasible region actually suggests that many of these modest production runs were in fact planned, either because that was all the production that was needed for a high-value chip or because the purpose of the device run was prototyping.

4. PROTOTYPING

Fig. 4 Reticle ASP decline with year



The purpose of a prototype is debugging. It may be debugging electrical characteristics, optical RET, logic at the chip level, logic at the system level, or even debugging the marketing. Depending on the purpose, EbDW can be an extremely cost-effective approach. One can see from the increased interest in FPGA based simulators, that getting the logic right at today's high levels of integration is increasingly challenging. An ASIC can take 3-6 months to debug and a more complex SoC can take 6 months to a year. The most common sort of problem is in fact logic and a prototype that can actually fit in the socket and run at speed can be enormously helpful. One might only need parts of a few wafers to get a sense of the product performance distributions. While today's EbDW designs are not generally designed to emulate the optical effects on the pattern, such an approach might be used to get as close to the electrical performance as possible in order to debug electrical problems. This will become easier as people move to more restrictive design rules. By the way, if etch loading and other such issues need to be contemplated, one can always direct write a few of the die on a wafer and print any other similarly dense design around the rest of the wafer just as a dummy pattern. In any case, the biggest value of EbDW for prototyping is lower cost and faster time to system checkout.

5. RETICLE COSTS

In some sense, the fundamental reason to entertain the concept of EbDW is to avoid reticle costs. There is a variation on this that is worth discussing. That is, avoiding reticle costs for a period of time. The concept here is similar to that employed by Dell in the early days of their competition with IBM[9]. Since the price of microprocessors and memory were falling so fast, by cutting down the time from component purchase to customer delivery, Dell was able to effectively get lower prices even though they had less buying power. Today, the price for high-end reticles declines roughly 30% per year as shown in Fig. 4. Thus, delaying the purchase of a $2M reticle set by six months can save over $300K. More generally, one can see a cost effective scenario where if you are doing ASICs or SOCs, it makes sense to let the flash,

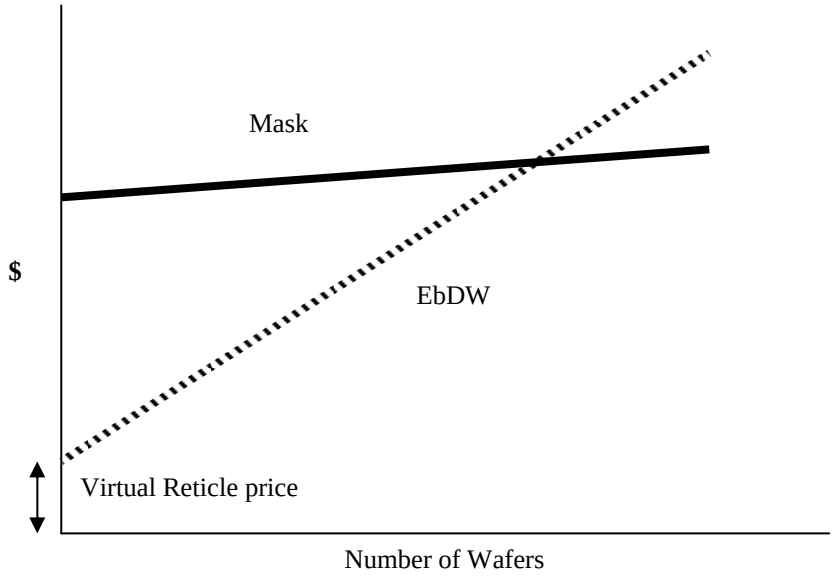
DRAM, and microprocessor businesses pay for the early development of a new lithography while you do your early prototyping, product sampling, and even early production with EbDW.

We need to point out that there is a potential challenge here. Since one of the high risk areas in production is in fact the lithography, switching from EbDW to optical lithography will probably require the product to be at least partially re-qualified, which is expensive and time consuming. One area of future study is the best way to do this re-qualification, minimizing cost and time. Qualification can take six months from tape-out to final qual and cannot be expedited. It requires multiple independent lots. Most Fabless customers are getting, if anything, more conservative about qual, even for relatively small lots. Re-qual is the biggest impediment to starting with EbDW and switching to masks for production even though there are potential cost benefits of doing so. Thus we expect that prototyping and smaller-lot production runs that can be done completely with EbDW (e.g., derivatives) will be the majority of the applications. Note that mix and matching layers is perfectly fine from a qual standpoint, just as long as that mix doesn't change during the product run.

6. VIRTUAL RETICLES

Finally, we introduce here a business concept that may be useful in EbDW business because it allows customers to buy in exactly the way they always have while providing cost benefits to the customer and pricing flexibility to the EbDW vendor. Today, customers buy ASICs by paying for a reticle set and then paying additional charges per wafer. We suggest that for EbDW, the customer should purchase a "virtual reticle set" plus an additional charge per wafer. The two options might look as shown in Fig. 4 as a function of wafer or chip volume. The crossover is clearly shown and the economically infeasible area is shaded. It seems likely that virtual reticle pricing will need to be 20-60% of the cost of

Fig. 5 Optical lithography vs EbDW costs including reticle and virtual reticle offsets



the physical reticles. The virtual reticle price might include setup costs, data prep costs, software and IP costs, etc.

One of the advantages of this methodology is that it enables EbDW pricing to follow the competitive pressures of the falling prices of physical reticles as shown in Fig. 4 and the diversity of layer costs. Of course, when the physical reticle cost gets low Fig. 5 should be looked at layer-by-layer, year-by-year. We can expect, for instance, that contacts and vias will be particularly advantageous for EbDW since these are the most difficult layers to expose with optics (with depth of focus for EbDW being substantially better than for light, optical energy requirements are very high because the light falls off in two dimensions around the via, and MEEF being particularly problematic) and the easiest to do with EbDW (be-

cause the exposed area is only about 5% of the total, lowering space charge effects and allowing higher current for the same distortion and blurs).

As we have seen, there is a good business case for EbDW for prototyping and for smaller run production volumes, particularly with respect to product derivatives where design costs are more contained. Should the numerous groups around the world succeed in technical efforts to make productive EbDW systems, perhaps the number of ASICs could even return to the larger numbers we saw in previous decades.

7. ACKNOWLEDGMENTS

The author would like to acknowledge helpful discussions with Dan Dobberpuhl, Aki Fujimura, Franklin Kalk, Burn Lin, Don MacMillen, Scott Paull, VLSI Research, and many others.

REFERENCES

1. G. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, 38(8), April 19, 1965. Available at the Intel corporate web site: http://download.intel.com/research/silicon/moorespaper.pdf
2. B.J. Lin, "The ending of optical lithography and the prospects of its successors," *Microelectronic Engineering*, 83, 604-613 2006.
3. H.C. Pfeiffer, "The History and Potential of Maskless E-Beam Lithography," *Microlithography World*, 14(1), 2005.
4. P. Buck, C. Biechler, and F. Kalk, "A Mask Manufacturer's Perspective on Maskless Lithography," *Bacus Symposium* 25; October 2005.
5. Device Sub Working Group, EVUA, "The Estimation of Cost of Ownership of EUV and ArF Immersion Exposure Systems," May 2006. http://www.euva.or.jp/images/technical_info/pdf/CoOMay2006/exposure.pdf
6. See http://www.mosis.com/ and also http://en.wikipedia.org/wiki/MOSIS
7. C. Anderson, *The Long Tail*, Random House Books, London, 2006.
8. Y. Nagahiro, "Application scope of EB exposure writing expands from R&D of advanced devices to prototyping and small production," *Nikkei Microdevices*, p. 31, June 2003.
9. S. Paull (private communication)