Design for e-beam: design insights for direct-write maskless lithography

Aki Fujimura*
D2S Inc. 4040 Moorpark Ave, Suite 250, San Jose, CA, 95117, USA

ABSTRACT

Designers always want maximum freedom in design, but they understand that chips have to yield and at a reasonable cost. The strong ecosystem support of restricted design rules to make 193i workable for sub-30nm nodes is evidence of this. In direct write e-beam, there are design insights that lead to a tangible improvement in throughout while minimizing the restrictions on the designer. It turns out that a smaller number of primitive cells in a standard cell methodology can enable data compression for multi-beam systems, and enable faster write times for character projection in VSB-based multiple column machines. This requires a co-design of the standard cell library with the stencil mask (either virtual or real) that goes into the machine. This co-design step is required only once per library and not on a design-by-design basis, thus minimizing the impact on designers. 10-20X speedups in e-beam throughput depending on layer are seen in typical layout examples for character projection machines.

Keywords: ML2, Maskless, EbDW, Direct Write, DFM, DFEB, Character Projection

1. INTRODUCTION

The interest in maskless e-beam direct write technology is on the rise again. Multi-patterning allows 193i lithography to continue to scale down. Additional investments in EUV continue to break down barriers to its implementation. And Nano-Imprint Lithography (NIL) continues to improve the throughput. But being maskless has unique benefits and being e-beam has long-term scalability that is attractive. Compared to 193i or EUV, e-beam is very well behaved. Furthermore, 30 years of both direct write and mask write history and technology scaling provides ample supply chain competence and know-how for a fast ramp-up required for adoption. The issue with e-beam direct write is the time it takes to write a wafer.

By clustering of multiple machines, particularly if each column unit can be made small enough, a factor of ten in speed is forgiven in recent discussions, making high volume production use of these machines once again within reach. 10 WPH per machine would be sufficient for high volume production then.

In addition, the ever-rising mask costs make low-volume production and initial entry cost of a new semiconductor design implemented in the leading edge technology node extremely difficult. The difficulty is mostly economic. It is not practically possible to configure a win-win proposition between a silicon entrepreneur and a venture capitalist with the non-recurring engineering cost of the mask set at $4-6M to get through Round A funding for a proof of concept. Similar equation prevents existing larger companies from investing in new ideas at the leading edge nodes.

The mask cost also becomes a prohibitive factor for derivative designs. Even where the design cost is relatively low because of the large degree of design re-use in derivative designs, a high mask cost would prevent the win-win between the semiconductor supplier and its customer. The many talented semiconductor designers are prohibited from creating differentiation for their companies and their customers. Increasingly, differentiation is derived from embedded software and re-programmability, even though there are many valuable features that can only be enabled on silicon.

Because innovation is inherently risky, being able to collectively try a large number of ideas is critical to fueling growth in finding the next big hits. The declining ability to do that at the leading edge node is a crisis for the semiconductor industry.

Maskless lithography helps with this, even when applied only for lower-volume production, prototypes, test chips, and research. Even though the per-wafer cost is higher, by eliminating the expensive critical dimension masks, both turnaround time and the per-chip cost is lower for these types of designs with maskless lithography.

*aki@design2silicon.com; www.design2silicon.com

There are a handful of massively parallel beam projects for direct write around the globe. Some of these machines [1], [2] draw the patterns on the wafers by pixels, and therefore can draw any shape in the same amount of time. They, in essence, provide a virtual programmable mask. The multiple shaped beam (MSB) approach [3] shoots a large number of multiple beams of the traditional VSB shots and extends it with character projection (CP) shots. The multiple column cell (MCC) approach [4] miniaturizes a character projection column and shoots 8 shaped beams in parallel on the wafer.

This paper addresses mostly character projection. The benefit of character projection is that there is maximum transfer of energy onto the wafer for each shot. Since it is a natural extension of the existing shaped-beam production writers, the effort required to productize it is significantly less. The drawback is that the productivity of the characters is important in achieving the faster write times. Traditionally, even though theoretical exercises indicated 3-5X improvement in write times using CP, practically, 2-3X improvements have been more realistic for real designs. It is easy to lose a factor of 2 in shot count because it only takes one VSB shot difference between what is available on the list of available characters and the required shot shape. A more careful collaboration between design and manufacturing is required to maximally leverage the CP capability. We have demonstrated over multiple designs an additional 10X reduction in write times from the Design For E-Beam (DFEB) techniques that are described below.

The same reduction is achievable for data compression of the data path in all massively parallel beam machines using DFEB. There are likely other ways to coordinate and collaborate between design and manufacturing for the massively parallel systems as well to highlight the strengths of the machines and to minimize the effects of the weaknesses of the machines through design optimization.

2. CHARACTER PROJECTION

Maskless character projection (CP) e-beam direct write (EbDW) technology eliminates mask costs for the critical layers. Both cost and time-to-market considerations are attractive for the lower-volume applications, or for applications with uncertain near-term volumes.

Traditional Variable-Shaped Beam (VSB) method fractures the desired shape into constituent rectangles and sometimes 45-degree triangles. CP deploys more complex characters on the 2nd aperture of the e-beam machine. These characters on the stencils project complex shapes in one shot.

Since the CP-based EbDW write times are nearly linearly correlated with the shot count, shot count reduction achievable for a wafer is a key measure that determines the feasibility of maskless SoCs. The shot count reduction achieved from CP depends on how many VSB shots a shape with a given character would have required. With SRAM core cells for example, over a 100:1 reduction in shot count is possible with CP without loss of accuracy. But you do not always get such effective reductions. Some characters have less reduction. And other shapes on the wafer are written with VSB shots, because there isn’t a character available to shoot that shape. The number of characters that can be made available on the stencil is a critical success factor for CP-based EbDW.

Each CP-based EbDW machine has a different number of CP characters available. The Advantest F3000 writer has the most number of characters available of any production writer today. Its specifications provide that there are 100 character slots per block area and that there are 21 user-specifiable block areas per stencil mask. Four additional block areas are reserved for calibration.

In the next-generation 8MCC machine from Advantest for the 22nm node and below [4], each e-beam column contains 8 50-keV electron guns, each equipped with a stencil mask with 21 user-specifiable block areas with each block area containing 1000 character slots.

The stencil mask is physically moved to position a certain character block area in the 2nd aperture area at any given time. Typically, a set of characters for a given layer of wafer processing is provided in one character block area. This is because the physical movement of the stencil requires re-calibration which takes orders of minutes. Any of the characters within a character block area can be chosen by electrostatic deflection. Switching between characters is accomplished during the blanking time in between shots. To maintain accuracy, only a certain radius of e-beam deflection is allowed, limiting the number of characters that can be put in a character block area. Since the critical layers that are usually written with EbDW are diffusion, active, contact, metal 1, via 1, etc. where the mask costs are expensive and time-consuming, 21 character block areas is more than sufficient for a given standard cell library. The Advantest machines also contain a stencil storage cassette, allowing up to 5 stencils to be in a machine, ready to use for multiple cell libraries without breaking vacuum.

3. DFEB PACKED STENCILS

Increasing the number of characters available is a key to further shot count and write time reduction. The DFEB Packed Stencil technology more than doubles the available characters by co-designing a standard cell library overlay with the stencil mask layout. The Packed Stencil represents an achievement that is only possible through a deep collaboration between design and manufacturing.

In the traditional division of labor between design and manufacturing, the machine knows nothing about the design or the cell library that the machine is intended to project, and the designer is not at all aware what is easier or faster for the EbDW machine to shoot in the fab.

The 100 character stencil mask of the F3000 (Figure 1) is laid out in a grid pattern because of this division. The machine needs to be built to handle any character. So there are 100 pre-designated spots for characters, each one being of any size or shape up to the maximum allowed size.

Each of the characters on the stencil is designed to shoot the desired shape as accurately as possible. In the DFEB methodology, characters on the stencil are shape-corrected for short-range e-beam and resist effects including Coulomb effect, forward scattering, resist diffusion, and micro-loading. Mid- and long-range effects including backscattering are adjusted through dose correction.

The maximum allowed size is dictated by the 1st aperture. In case of the F3000 for a 65nm node design, the maximum size is a 4μm by 4μm square. In case of the 8MCC for 22nm node and below, the maximum size is between 1μm by 1μm square and 2μm by 2μm square. The selection of the appropriate size is also a subject of design to manufacturing collaboration. The target node, its standard cell heights, the SRAM core dimensions, and minimum pitch are critical for determining the appropriate size. Bigger characters mean fewer characters per block area. Layers such as contact and cut layers have less than 25% of the character size open for electrons to pass through. These characters can be larger because the Coulomb blur is less. For VSB shots, where 100% of the character size is open, the maximum size would need to be smaller for shots that require accurate edges. These design and manufacturing considerations need to be balanced carefully for optimal effect.

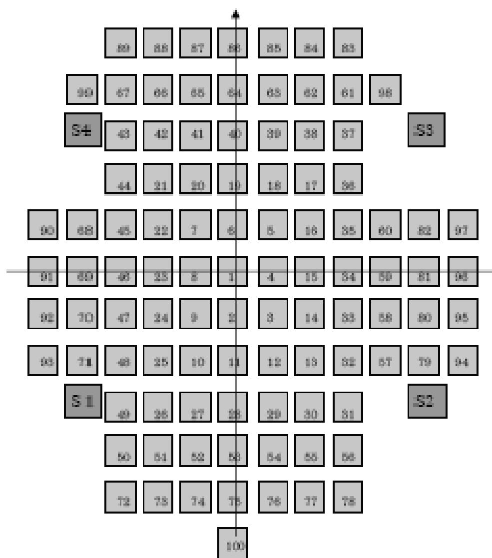

Figure 1. Previous specification of the F3000 character block with 100 characters.

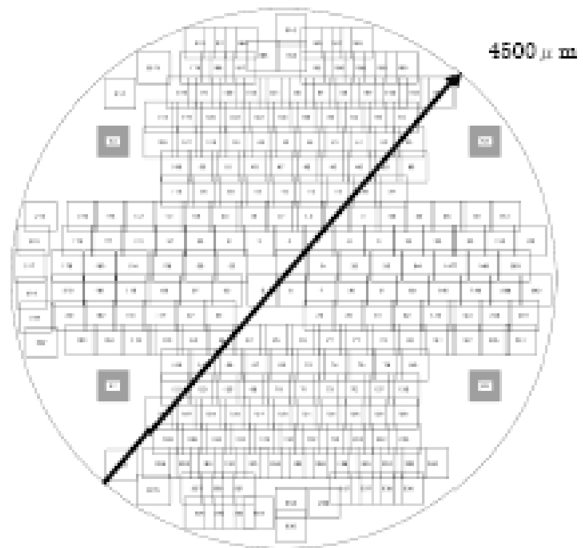

Figure 2. The Packed Stencil allows, for example, this packed layout of 220-280 characters.

Even after the machine's character size is determined, there is significant opportunity for further optimization. Specifically, the stencil mask layout is co-designed with the DFEB overlay cell library that the designers would use in physical design implementation in a standard cell methodology. This allows the characters on a character block to be packed, because there is specific knowledge of the exact dimensions for each character. The layout of the characters is

optimized to fit as many characters on the stencil mask as possible. This allows more patterns on the wafer to be shot as CP shots, thereby reducing the write times. An example layout for a 45nm standard cell library case for the F3000 machine is depicted in Figure 2. In this case, between 220 and 280 characters (depending on the layer) are packed in the same block area. This significantly improves the number of standard cells and SRAM cell types that can be written in one or two CP shots. While any design of any shape can be written using VSB shots, the more of the design can be shot with CP, the faster the write time will be.

The four VSB apertures in the corners (shaded darker) have blanking areas below and to the left of each. These areas are identical in both Figure 1 and Figure 2. The large blanking area is required because the VSB shot maybe a very small shot. In order to shoot a very small shot, the electrons projected through the first aperture onto the stencil mask will need to be mostly blanked out. In this machine, the blanking is always to the bottom lower corner. So the total area required for a VSB shot is four times the size of the VSB opening, plus some margin.

CP apertures where partial character projection is not used (see below) do not require such blanking areas. There must be enough space between the character and any adjacent character for the square image of the e-beam from the first aperture plus some margin. But no additional blanking area is required.

The DFEB Packed Stencil technology reduces the area required for the characters by sharing blanking areas between two adjacent characters when they require less than the size of the first aperture opening.

Because the total area used by characters on any given character block area is not changed by the Packed Stencil, the characters are projected just as accurately as was the case with the 100 character stencil. Although the machine needs to interpolate the calibration results, since calibration is done only with the 100 character positions, the added benefit of the increase in character count far outweighs the cost.

Extrapolating these results on the next-generation 8MCC machine, the 1000 character positions will be leveraged to pack over 2000 characters for each layer of the design.

4. DFEB OVERLAY LIBRARY

In a standard cell methodology, a system-on-chip is designed by floor-planning I/O cells, standard cells, SRAMs, and other macro cells together. In a leading edge design, the chip is dominated in area, and even more dominated in shot count, by the standard cell and SRAM areas. In VSB shot count, SRAM cores are the densest. Standard cell and SRAM periphery sections are of average density. I/O and macro cells, such as analog and other components, are typically low in shot count density. Therefore, the majority of shot count reduction from CP and DFEB comes from the standard cell and SRAM sections.

A typical standard cell library for a particular technology node may have 600 to over a thousand standard cells, and may have available 4 to 20 SRAM types. In the DFEB methodology, a standard cell library subset is chosen to have a DFEB Overlay Library counterpart for which every cell in it will have a corresponding character in the stencil mask in some orientations.

At first, the various SRAM cores and the “top 40” standard cells may seem to be sufficient. But upon closer inspection, the order of magnitude improvement through CP and DFEB can only be achieved if there are more than 100 characters available. This is why the DFEB Packed Stencil technology was essential.

Of the eight orthogonal orientations possible for any standard cell or SRAM cell, the DFEB methodology needs to limit the design methodology to deploy a limited subset. Typically, a north-south direction with the “normal” and the flipped south (or flipped about the X axis) orientations will be provided for each standard cell in a DFEB methodology. Because the e-beam machine cannot “mirror” characters, every desired orientation needs a different character, unless there is symmetry that can be exploited. For SRAM cells, both the core and the periphery need four orientations, even if the transistor direction is restricted to only up-down or left-right. SOC floor planning is sensitive to macro cell orientation because of interconnect pin access.

SRAM cores are the densest by far in shot count. And SRAM cores are typically axis-symmetric. But upon closer inspection, it turns out that shooting the periphery (sense amps, etc.) of the SRAMs with CPs are important to the overall shot count reduction number. As much as 20-30% of the shots of an SRAM macro may be from outside the core areas. This gets worse for smaller SRAM macros, of which today’s SOC designs have many. So even if we magically made SRAM cores shoot with zero shots, the best improvement that can be made on the overall shot count reduction for an

SRAM macro without CP characters for outside the core area is 3-5X, far less than the order of magnitude we seek. Having SRAM periphery on the character block area is therefore critical to shot count reduction. So SRAM periphery ends up needing four orientations each, even though the core typically requires only one character. This requires a large number of the available characters.

For the F3000, with only 250 characters in a block area, about 50 different standard cell types with its north-south flipped versions can be turned into a CP shot. The rest are shot with the conventional VSB shots.

For any one DFEB design, the design can be synthesized so that the top 30 or 40 standard cells represent an overwhelming majority of the shapes in the standard cell areas. Accounting for the flipped-south orientation, 60 to 80 characters would be sufficient to represent them. We should note that a design that is not DFEB (i.e., designed for the traditional mask-based process) will also have the top 30 or 40 standard cells be the majority, but not an overwhelming majority. The difference is that DFEB explicitly synthesizes the design in the logic synthesis step to prefer the use of logic cells that can be shot with CP over those that cannot be. Since logic synthesis typically has many arbitrary choices that are all equally good for area, power, and timing considerations, teaching synthesis tools to prefer one-shot CP cells in those situations does not hurt the quality of the design.

However, the DFEB methodology is that the design-independent stencil mask is waiting in the EbDW machine for all designs that use a particular standard cell library. To experience the turn-around time advantage of going direct to silicon, having more than 100 characters is necessary. Indeed, having only 250 characters per layer even with the DFEB Packed Stencil technology represented a compromise. The over 2000 characters available in the next-generation 8MCC system is a welcome relief for maximum impact from DFEB and CP.

5. DFEB CHARACTER SHARING

The co-design of the stencil mask and the standard cell library goes further. In order to represent as many standard cells as possible on a stencil mask, in addition to increasing the number of characters per block area, many of the characters are designed to be used by multiple standard cells in the DFEB Overlay Library by partial projection of characters.

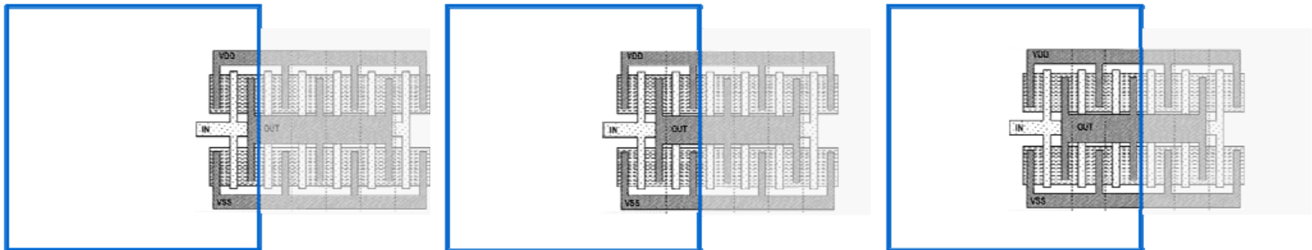

Figure 3. By "cutting" first aperture deflection at various positions, different drives of the cell can be shot from the same character.

One example of character sharing in the DFEB mapping of standard cells and characters is in how the buffering of the output drive capability is handled using the Partial Character Projection (PCP) capability on the EbDW machines.

Just as the overlapping of a square 1st aperture with a square 2nd aperture produces rectangular VSB shots on the EbDW machines, an overlapping of the square 1st aperture with a character in the 2nd aperture can produce partially projected characters, or PCPs.

In a standard cell library, many of the basic cells that are used often have multiple drive capabilities. For example, a buffer cell may have a 1/4X, 1/2X, 1X, 2X, 4X, 8X, and a 16X drive variation in the output. In a typical standard cell library, these are different standard cells. The layout may or may not have shared components in them. Standard cell layout designed for regular mask-based manufacturing has no incentive to have these different buffer components share components. But in DFEB, there is a large incentive to have a single character represent as many standard cells as possible.

Figure 3 is a layout of a sample cell whose output buffer section has been designed for e-beam using PCP. The diffusion, poly, metal, and contact layers each would have a character that is used by the 1X, 2X, 3X…6X buffer versions of the same logic function. By "cutting" the character by the 1st aperture deflection at various positions

indicated by the squares in the figure, different sized buffer cells in the cell library can be shot from the same characters. In this case, one character shoots 6 different standard cells.

There is a high correlation between cells in the cell library that are most often used in designs, and cells in the cell library that have multiple buffer size variations.

As a result, even after taking flipped south orientations into account, with only about 250 characters in the character block area, over 130 standard cells in the library in addition to SRAM and other characters can be shot by CP (or PCP).

Having over 130 standard cells in the DFEB Overlay Library means that most designs can have the standard cell portions shot with CP. The cells from the original library would be shot using the VSB method. If the DFEB-driven synthesis methodology insisted on using VSB shots to shoot a cell that is not in the DFEB Overlay Library, it must have been needed for performance, power, or area optimization. Having 130 cells instead of 30 or 40 makes a large difference in this trade-off of shot count as a consideration vs. timing power and area. This means that more often, a design can be optimized for shot count without sacrificing any other quality of result metrics.

The increase in the number of cells that can be in the DFEB Overlay Library as "one shot" CP cells also means that there are more variety of design types that can experience the order-of-magnitude shot count reduction that is needed to make EbDW practical. In the 8MCC system, the use of partial projection increases the effective number of characters available for a layer in a given block area to over 2500 virtual characters.

6. DFEB METHODOLOGY

Prior to DFEB, the use of character projection was envisioned as depicted in Figure 4.

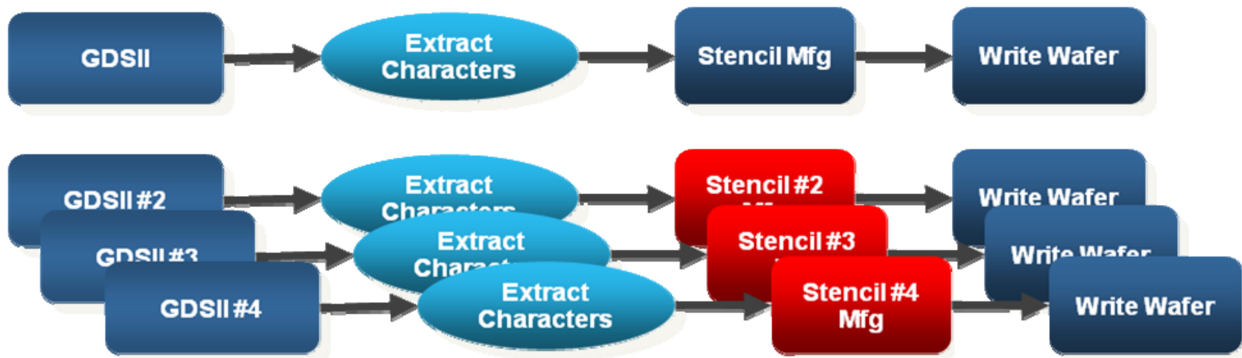

Figure 4. Conventional Character Projection (CP) flow prior to using Design For E-Beam (DFEB).

Each design that comes into the fab would be inspected to extract the 100 most commonly occurring characters. These most commonly occurring characters are then put on a stencil mask. When the stencil mask is available, wafer writing starts for that design. For each design, there would be a different stencil mask.

The DFEB flow is different. As depicted in Figure 5, a stencil mask is co-designed with the DFEB Overlay Standard Cell Library. The stencil mask is prepared per library, not per design, and is waiting for all designs done with that library in one of the stencil cassettes in the machine at the fab. There is no waiting time for the stencil mask to be manufactured. Since there is also no OPC, Mask Data Prep, mask writing, inspection or repair, when the layout is completed, direct write data prep pipelined to the e-beam writer can start (if available). The savings in the turnaround time, particularly for test chips, prototypes, and lower-volume production designs are substantial.

For each design, starting at the Register-Transfer Language (RTL) description of the design with logic synthesis, leading into place and route and design completion and verification, each design is explicitly optimized to use the available characters maximally. This overlay design flow leverages only existing SP&R tools and does not rely on any proprietary algorithms.

In addition, the DFEB design flow is an overlay on the existing Synthesis, Place and Route methodology including DFM steps and rules that are in place for optical processing (Figure 6). This allows all designs created using this flow to be compatible for later processing with light. More importantly, most engineers already trained and tooled for optical

processing want to minimize complexity to minimize mistakes. What they already know is least complex from their perspective.

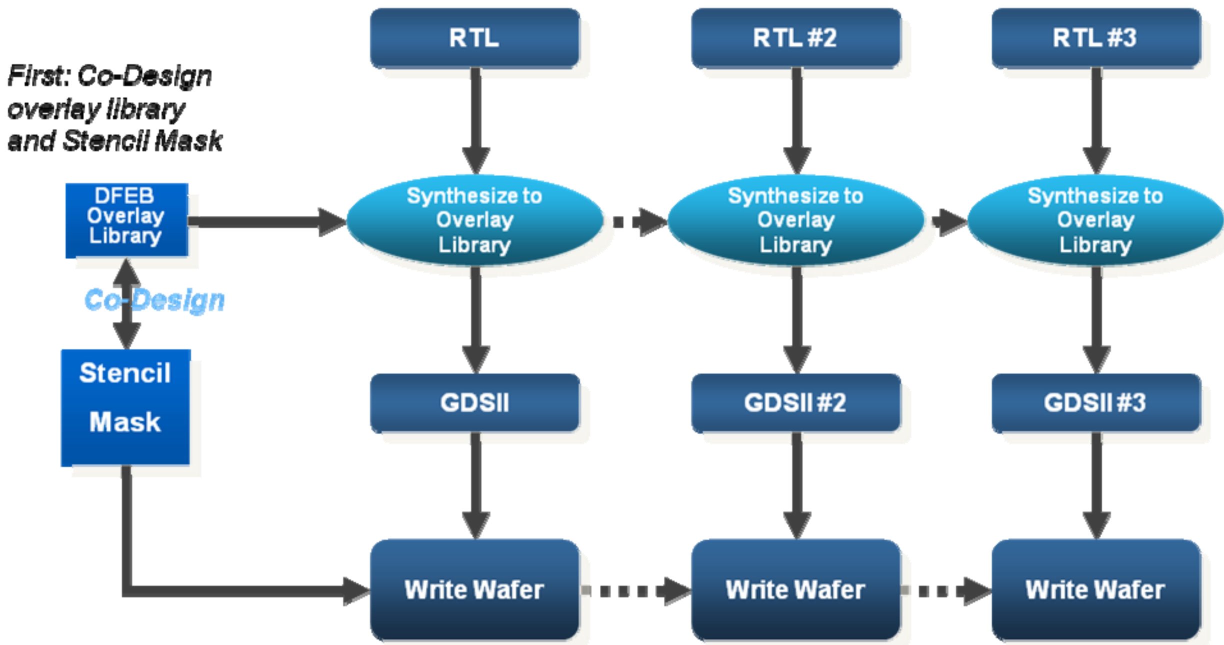

Figure 5. In the DFEB methodology a stencil mask is co-designed with the DFEB Overlay Standard Cell Library. The stencil mask is prepared per library, not per design, and is waiting for all designs done with that library in one of the stencil cassettes in the machine at the wafer fab.

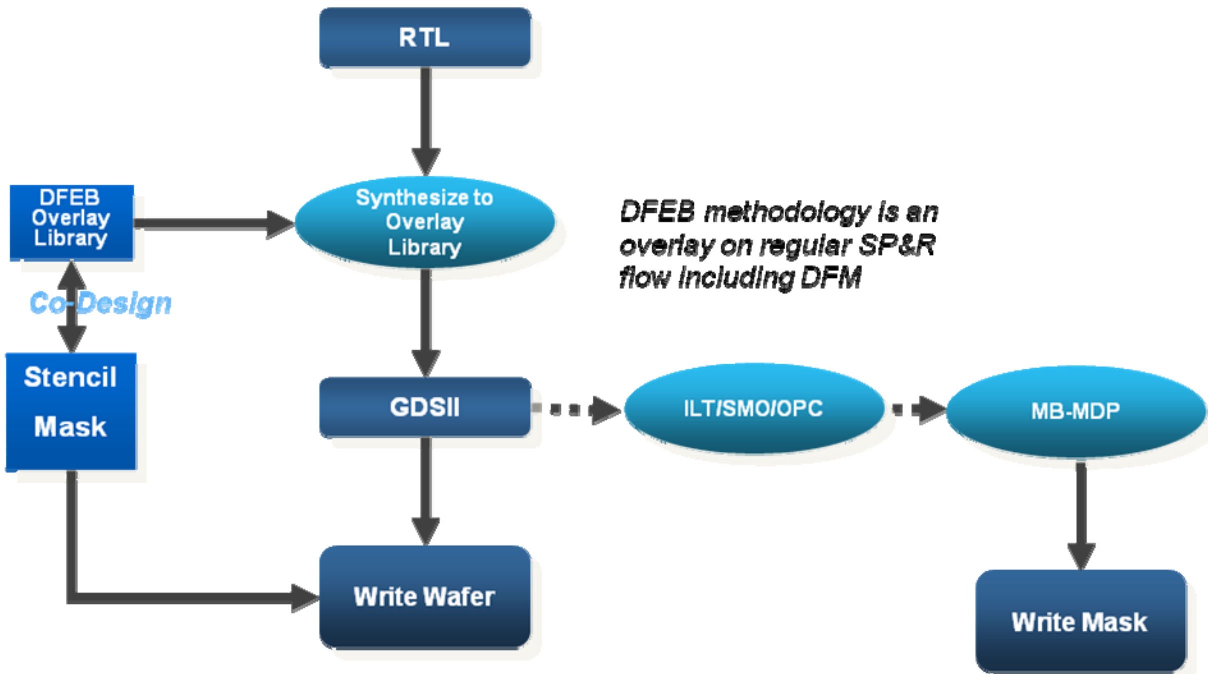

Figure 6. The DFEB design flow is an overlay on the existing Synthesis, Place and Route methodology including DFM steps and rules that are in place for optical processing.

Combined together, at the 45nm node, a 250 character stencil leveraged by DFEB is demonstrated to yield 10-20X improvement in write times over VSB shots depending on layer and depending on design for the standard cell layers. At

the 22nm node, a 2500 character stencil leveraged by DFEB is projected to yield 10-25X improvement in write times over VSB shots depending on layer and design for the standard cell layers.

In the interconnect areas of metal 2 and above metal layers, there are less shots required to start with. The reduction in shot count achievable with CP even with DFEB is limited there, particularly with gridless routing. A further restriction in design rules projected at the lower technology nodes that return us to gridded routing will improve the shot count. Via layers for Via 2 and above layers have better savings of shot count from CP due to the large number of via arrays.

7. CONCLUSIONS

We have demonstrated that a production EbDW machine can write over 130 standard cells as single shot CP characters. This is done by the DFEB concepts of packing characters on the stencil mask, and sharing characters across multiple standard cells. By extension the next-generation 8MCC machine will be able to write virtually all standard cells and SRAMs with character projection. Although it is the most important feature of CP-based EbDW to be able to write any shape on the wafer using VSB shots, maximizing the percentage that is shot with CP helps to minimize write times.

This is a result of a deep collaboration of design and manufacturing to enable EbDW for low-volume designs. An order of magnitude improvement in EbDW throughput enables maskless SoCs for the leading edge nodes, eliminating mask cost as a barrier. The resulting increase in design starts at the leading nodes is good for every segment of the semiconductor supply chain.

ACKNOWLEDGEMENTS

The author would like to thank Kiichi Sakamoto, Akio Yamada, and Hideaki Komami and the many people of Advantest Corporation for their deep collaboration that enabled these features. This article reflects the work of the many people of D2S, Inc. and D2S KK, including Larry Chau, Tam Nguyen, Shone Lee, Shohei Matsushita, Kazusumi Kuwano, Haruyuki Tago, Takashi Mitsuhashi, and Kenji Yoshida. Thanks to Ingo Bork for his help in the preparation of the manuscript.

REFERENCES

[1] Kampherbeek, B. J., “High-throughput maskless lithography,” to be published at BACUS 2010, paper 7823-140, (2010).
[2] McCord, M.A., Petric, P., Bevis, C., Carroll, A., Brodie, A.D., Ummenthala, U., Grella, L., Freed, R., “REBL : reflective electron-beam lithography for high-throughput direct-write patterning”, to be published at BACUS 2010, paper 7823-138, (2010).
[3] Slodowski, M, Doering, H-J, Stolberg, I., Dorl, W.H., “Multishaped-beam (MSB): an evolutionary approach for high-throughput e-beam lithography”, to be published at BACUS 2010, paper 7823-139, (2010).
[4] Yamada, A., “Cell projection use in maskless lithography for sub-32-nm logic nodes,” to be published at BACUS 2010, paper 7823-137, (2010).