ENABLING THE LONG TAIL OF SOCs

AKI FUJIMURA, CHIEF EXECUTIVE OFFICER, D2S, INC.

The ever-rising cost of semiconductor masks is making low-volume production of systems-on-chip (SOCs) economically infeasible. The "long tail" of the SOC business is in a large variety of low-volume SOCs. In aggregate, this segment can represent as much volume as the high-volume segment. Maskless e-beam direct write (EbDW) technology enabled by character projection (CP) capability in today's production equipment, coupled with design for e-beam (DFEB) software and design technologies, enable this long tail of SOCs by eliminating mask costs. By not requiring the development of any revolutionary new hardware technologies, this design and software approach represents a low-risk, low-cost path to a new production paradigm. Enabling greater direct-write productivity rates and removing the need for optical proximity correction (OPC) changes the cost equation around leading-edge manufacturing at the 45-nanometer node and below. Through the collaboration of the semiconductor design and manufacturing supply chain, DFEB will enable the long tail of SOCs.

THE LONG TAIL DEFINED

In "The Long Tail: Why the Future of Business is Selling Less of More," the author, Chris Anderson, makes a strong case for the Internet enabling low-cost access to niche markets, thereby enabling the long tail of a large variety of products, each with low volume, but in aggregate forming half of the overall market both in terms of volume and in dollars. The argument is not that less of the usual high-volume products are sold, but that the overall market doubles by enabling the long tail in addition to the traditional high-volume products.

Key to enabling the long tail is easy, low-cost access to producing low-volume products. PC-based authoring, on-demand book publishing, PC-based home studios and MPEG-based access to music are examples of successful implementations of this concept.

In the SOC business today, the long tail is difficult to imagine because of high design costs and high mask costs. The way to address high design costs is with derivative designs. Platform design with derivatives is the highest form of intellectual property (IP) reuse. There are a relatively small number of product types (digital TVs, video cameras, network routers, cell phones, global positioning systems (GPS), navigation, etc.), but there is a large variation of creative differentiation for each type. Designers try to make as much variation available as possible through firmware changes, but there are still many variations that require a new mask set. The variations

might include a wider bandwidth version, a version with twice the memory or a different precision in the digital-to-analog converter (DAC). A customer of a fabless semiconductor company saying, "I'd like a version of that chip with this slight improvement" is a frequent occurrence.

So while design costs can be minimal for these derivative designs, ever-rising mask costs prevent the wide adoption of design variants at leading-edge nodes. Since mask costs, in general, fall by 30% annually after initial introduction, derivative designs do not start to be exploited until three years later when products from the next technology node are already hitting the market.

What if the mask costs associated with leading-edge nodes were able to be eliminated?
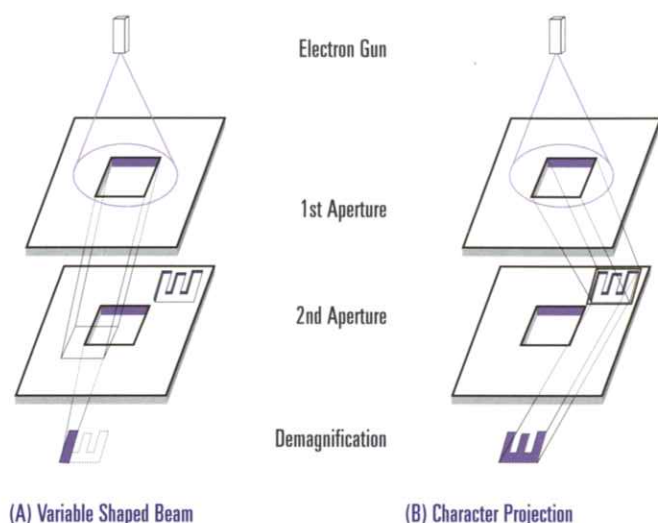
CP E-BEAM TECHNOLOGY

E-beam projection of masks and wafers has evolved over time from Gaussian spot beam to variable shaped beam (VSB) and CP. Figure 1 illustrates the difference in feature writing capabilities between VSB and CP. All VSB-based e-beam projection machines are fundamentally capable of CP. Today's mask making machines do not have that feature built in, but EbDW machines currently in production do possess CP capability. Regardless of technique (i.e., VSB- or CP-based), the cost of e-beam projection is dominated by the e-beam shot count. The ability to shoot all features at once, rather than in the many tens of shots, enables a linear reduction in the cost of lithography. In character projection e-beam (CpEb), complex shapes, such as the poly layer of a D flip-flop with scan, are projected in one shot instead of the many tens of VSB shots that are the result of fracturing complex shapes into rectangles.

Even though this CP capability has existed in EbDW machines for many years, it is only at the 65- and 45-nanometer nodes that standard cells have become small enough to fit within the allowed dimensions of a CP character. Being able to project a D flip-flop with scan — usually about 10% of the cell count and as much as 25% of the standard cell areas (and VSB shot count) — in one shot is a critical difference in the effectiveness of CP technology for shooting SOC designs.

EbDW has long been accurate, but too slow to be practical for all but research and exotic applications. Now that SOCs are dominated by RAM and standard cell areas, it would seem plausible that a large percentage of a SOC could be shot using the CP technique with an order of magnitude difference in shot count,

Figure 1. CP Projects Complicated Figures All at Once

Electron Gun

1st Aperture

2nd Aperture

Demagnification

(A) Variable Shaped Beam   (B) Character Projection

Source: Hitachi High-Technologies Corporation; Provided by Genya Matsuoka, SEMICON Japan 2003

writing speed and lithography cost. Recently published research from Fujino[1] and Kazama[2] indicates that DFEB design and software technology adds a large incremental advantage in shot count, enabling the total shot count within the threshold of all computer-aided design (CAD) layers of a wafer to be written cumulatively in less than one day.

DESIGN FOR E-BEAM

Figure 2 illustrates the DFEB difference. In this example, even though the use of CP without DFEB yields only a 5X shot count reduction, the use of DFEB with CP yields a 25X difference in shot

Figure 2. Computer Simulation of E-Beam Write Time on a Particular Test Case (Speed up is Dependent on Aperture Size and Utilization %)
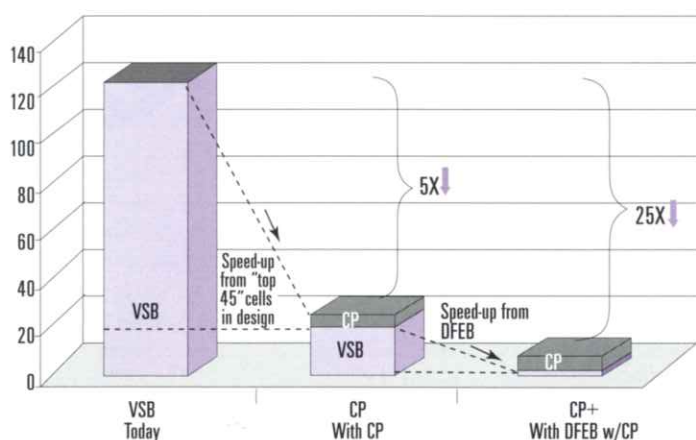
count. The typical division of responsibility between design and manufacturing results in the "with CP" bar without DFEB. Here, the foundry's job is to write on silicon whatever the designer may have specified in Graphic Data System II (GDSII) (so long as it meets the fab's design rules), and the design is analyzed to see which patterns appear most frequently. If there are 100 characters that can be made available on the CP machine, then the top 100 characters are chosen. Since characters need to instantiate orientations, assuming that standard cells need a north and a flipped south

orientation made available, and RAM and other features require some of the 100 characters, in this example, 100 characters equated to the top 45 most frequently occurring standard cells chosen.

The "with CP" bar indicates that many VSB shots were converted to CP shots. For those converted shots, the reduction in shot count is quite significant. But there were also a large number of VSB shots that could not be converted to CP shots. Therefore, those remaining VSB shots end up being the bottleneck. The overall reduction is only 5X.

With DFEB, there is a closer collaboration between design and manufacturing. Instead of having a new stencil mask made of a different set of characters for each design, a stencil mask of characters is pre-designed for a particular standard cell (and RAM) library (e.g., a 45-nanometer, high-performance library). Taking advantage of the specific knowledge of the layout characteristics of a particular library enables a vast increase in the number of characters that are available on a stencil mask. Further minor modifications to a standard cell library and the careful layout of cells (including RAM cells) yields a set of CP characters that can optimally and accurately shoot designs of the particular library. The increased number of available standard cells that can be exposed in one CP shot is then provided for synthesis, place and route (SP&R), where a new methodology explicitly optimizes the DFEB shot count simultaneously with area, timing, power and yield. The additional 5X shot count reduction is achieved by optimizing the design for shot count from the register transfer level (RTL), thus vastly reducing the number of shots that cannot be converted to CP characters. Since the DFEB methodology only affects post-synthesis design, the RTL is completely unaffected for systems designers.

Although the clear goal of DFEB is to reduce the number of remaining VSB shots as much as possible, a critical feature of this methodology is that those VSB shots are always available and possible for any special designs such as analog, radio frequency (RF) or high-speed input/output (I/O) IP. The VSB shot density is extremely low for these types of cells due to their relatively large feature sizes. In addition, the shot count percentage in a chip is typically very low due to these types of cells.

An extremely critical feature of SOC design is that any transistor-level innovation is possible. However, due to design complexity, RAM and standard cells dominate in area and even more in shot count. This is exactly the kind of condition that DFEB-based CpEb technology exploits to create designs that are optimized for EbDW projection cost, yet equivalent in performance, power and area. The resulting design can be projected at a rate of one wafer per day for all CAD layers of the design. Furthermore, DFEB designs are carefully created so the same design can go through the required design-for-manufacturing (DFM), reticle enhancement technique (RET) and OPC steps when using masks in high-volume production.

DFEB ECONOMICS FOR HIGH-VALUE, LOW-VOLUME CHIPS

Let's take a look at what this means from an economic standpoint. With mask costs escalating at each technology node, the first-year cost of a 45-nanometer design required a mask budget of $3 to $4 million. Mask costs do decline about 30% per year for any given node, yet the next node is at least twice as expensive as the last, making the projected first-year mask budget for a 32-nanometer design $6 to $8 million.

Although a lower mask price for high-volume projects can be negotiated at a major foundry, for high-value, low-volume chips, there is no such option. Since the depreciation cost of EbDW machines is roughly $20,000 per day, and all other wafer manufacturing costs are roughly equivalent to mask making, 300 to 400 wafers would be the break-even unit volume where the increased per wafer cost of maskless lithography equals the mask budget. However, this cost equation only applies to raw depreciation costs, so taking various profits and inefficiencies in the supply chain into account, 100 wafers or roughly 100,000 8mm x 8mm chips is a safe estimate for a break-even unit volume. For high-value designs and derivatives of high-volume designs that may produce 100 wafers or fewer, maskless production provides the same quality of all-layer SOC designs with faster turnaround time (no OPC and no mask making required) and easier engineering change order (ECO) capability (no mask to remake and no need for metal-only ECO).

PROTOTYPES

For designs of all volumes, DFEB and CpEb enable at-speed, all-layer custom SOC prototypes for early system verification, including software testing. Unlike gate array prototypes, field-programmable gate array (FPGA) prototypes or via-only prototypes, these all-layer custom SOC prototypes have the same density, speed, power consumption, SRAM and analog IP of volume production SOCs. For Consumer Electronics Show or MacWorld demonstrations of battery-operated devices without interface cables protruding unattractively, this is the only economically feasible way to produce the prototype demonstration vehicle. Maskless SOCs are also ideal for providing system test engineering samples to systems customers before a competitor. Finally, for the innovative, first-of-its-kind designs typical of start-ups, the volume of the first version is very much at risk even if the business plan calls for high-volume adoption. In these situations, starting with maskless SOC production and moving to mask-based volume production over time is the lower risk approach with a lower capital requirement for Series A funding.

With an economically feasible way to produce engineering samples and prototypes, silicon-level innovation will continue to be viable at leading-edge technology nodes. This, in turn, will spur growth in the semiconductor industry. Having the ability to try new ideas will generate more successful products earlier.

SUPPLY CHAIN COLLABORATION

The supply chain that collaborates to create this environment spans from DFEB design kits for application-specific ICs (ASICs) or customer-owned tooling (COT), electronic design automation (EDA) tools for SP&R and design services; to EbDW equipment manufacturing, resist making, mask making for stencil masks, process development and enhancement; to SOC wafer fabrication. One benefit of DFEB is that the current state of development in the supply chain is already sufficient for the use of this technology at both the 65-nanometer and 45-nanometer nodes. But to improve write times and to increase accuracy without the loss of speed for 32-nanometer and 22-nanometer applications, the supply chain needs to engage in close collaboration.

Ever-increasing mask costs are a large threat to all parts of the supply chain that scale by number of designs. In addition, every part of the supply chain is motivated by increasing the total unit volume and the total dollar volume of semiconductor consumption. The semiconductor design market has demonstrated that it is wise in its use of platform designs and their derivatives, supplying a large variety of differentiated offerings with its existing engineering resources. The availability of outsourced design services is an example of this. The one problem that remains, therefore, is mask cost.

The emergence of DFEB and CP e-beam techniques for EbDW technology will enable the long tail of SOCs for high-value, low-volume designs, derivatives and prototypes. ∎

About the Author

Aki Fujimura is chief executive officer of D2S, Inc., which provides VirtualMask™ solutions that enable maskless e-beam direct write for prototyping and low-volume production of SOCs. Prior to founding D2S, Aki was chief technology officer of new business incubations at Cadence Design Systems following Cadence's acquisition of Simplex Solutions, where he served as president and chief operating officer. Before Simplex, Aki was vice president and an inside board member of Pure Software. He previously held various executive management positions at Cadence after its acquisition of Tangent Design Systems, which he co-founded and where he helped to define Tangate, the source of such standards as Library Exchange Format (LEF), Design Exchange Format (DEF) and ECO. You can reach Aki Fujimura at aki@direct2silicon.com.

Resources

[1] T. Fujino et al., "Character-Build Standard-Cell Layout Technique for High-Throughput Character-Projection EB Lithography," BACUS Photomask Japan 2005, Proc. SPIE Vol. 5853, pp. 161-167, (April 2005). Note: This is from Ritsumeikan University.

[2] T. Kazama et al., "Shot reduction technique for character projection lithography using combined cell stencil," 25th Annual BACUS Symposium on Photomask Technology, Proc. SPIE Vol. 5992, 5992V, (Oct. 2005). Note: This is from Tokyo University.

Of course, compliance goes beyond environmental compliance. Financial compliance under Sarbanes-Oxley as well as Global Trade Compliance should be part of the holistic design of a compliance strategy and integrated within the operational system landscape. This approach makes compliance part of day-to-day business, allowing companies to focus on serving customers and introducing new products to market. ∎

About the Author

Greg has over 20 years experience in the high-tech industry, holding management positions in supply chain management, project management, systems implementation and international trade. Over the last eight years, Greg has consulted and worked closely with leading semiconductor companies in the U.S., Europe and Japan. Greg has been instrumental in the development of many SAP products, including SAP GTS, CRM Design Win, CRM Channel Management and, most recently, SAP Outsourced Manufacturing. Greg holds a master's degree in software engineering from Regis University and is currently responsible for SAP's semiconductor industry segment.

Resources

Aberdeen Group, "Building a Green Supply Chain: Social Responsibility for Fun and Profit" March 2008.

SAP AG, "Environmental Compliance for High Tech Manufacturers," 2008.

Semiconductor Industry Association, "Environment, Safety, and Health" – 2008.

Kubin, Richard "Navigating the New World of Green Technology" – 2007.