Semiconductor scaling with high-productivity multicolumn electron-beam lithography

By David K. Lam, Kenneth P. MacWilliams [Multibeam Corporation]

Businesses and consumers are adopting artificial intelligence (AI) at a rapid pace, placing enormous new demands on computing and communications. Next-generation data centers supporting upcoming AI applications will require massive new scale in high-performance computing (HPC) and high- bandwidth memory (HBM), as well as higher energy efficiency.

The industry faces major challenges in meeting this surge in demand. While the number of transistors has grown to 100 billion or more in today's state-of-the-art semiconductor chips, new AI applications have a near insatiable demand for additional computing performance. As the cost and complexity of continued scaling are increasing rapidly, leading chipmakers are turning to advanced integration of multiple chips to deliver necessary performance improvements.

Advanced chip integration is gaining adoption, but is also facing challenges: chip-to-chip communication bandwidth, I/O density, energy per bit transfer, and thermal effects. These must be addressed to reach performance levels required by AI, high-performance and edge computing, 5G/6G infrastructure, high-speed communications, and other emerging applications.

This paper shows how our multicolumn e-beam lithography (MEBL) system, working with advanced integration technologies, can help address these challenges and achieve the next-level of performance—in computing, bandwidth, and power efficiency.

Challenges in continued scaling

Major challenges facing chipmakers in continued scaling are not easily resolved. Some of these challenges are:

- Silicon area scaling limit of one "classic" reticle (26mm X 33mm), being further reduced to a half-size reticle in high-numerical aperture (NA) extreme ultraviolet (EUV) lithography systems;
- High cost and slowing progress in feature size scaling;
- Energy required to move bits from one chip to another increasing rapidly, consuming more than 70% of total power of the chip in some cases; and
- Cost and time of new full designs exceeding half a billion dollars [1] with long timelines.

Advanced chip designs have reached the maximum size of a single scanner field, currently at 33mm × 26mm. This is a challenge for chipmakers as the upcoming generation of high-NA EUV scanners are expected to further reduce this field size by half, to 16.5mm × 26mm. This shrinking field size constrains the area of each chip and limits the number of transistors in a given chip. Conventional scaling of transistors in accordance with Moore's Law has also slowed, while the cost to design and manufacture leading-edge chips continues to rise (**Figure 1**) [2-3]. It is increasingly difficult for chipmakers to deliver required performance and yield at reasonable costs (**Figure 2**) [1]. At the same time, more chip-chip interconnects are needed to support the growing demand for data processing [4]. Continued scaling will now require tighter integration of multiple chips with increasing numbers of chip-to-chip interconnects. Fortunately, high-resolution adaptable patterning with MEBL combined with new integration technologies enables such scaling [5].

Chiplet-based designs enable continued scaling and performance improvements, and provide space and power reduction benefits by employing denser interconnects. This means breaking complex chips into smaller chips called "chiplets" to avoid limitations imposed by today's chip

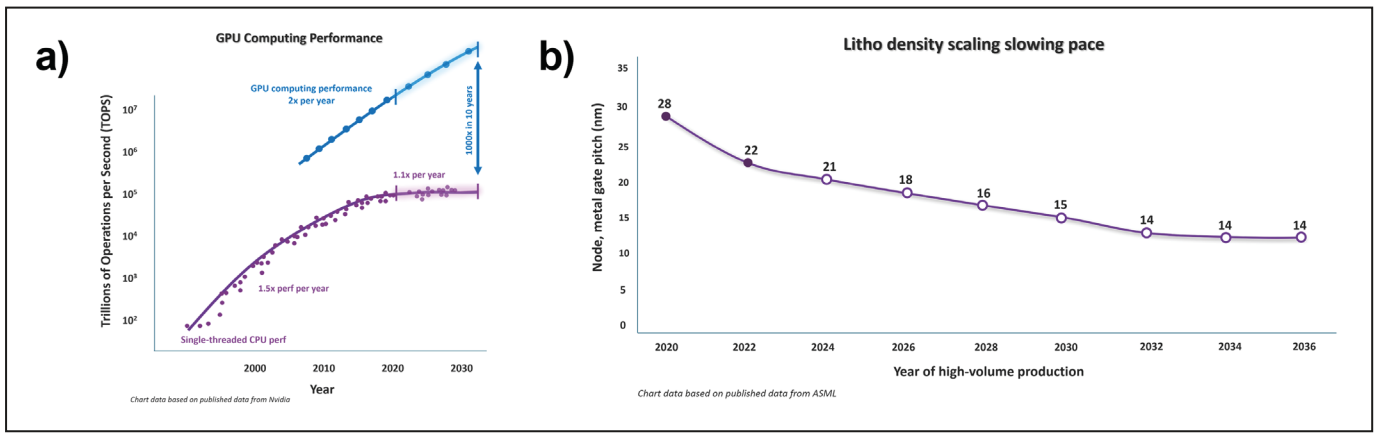

Figure 1: Slowdown of silicon scaling as highlighted in presentations by: a) NVIDA [2]; and b) ASML [3]. SOURCES: NVIDIA, ASML, respectively

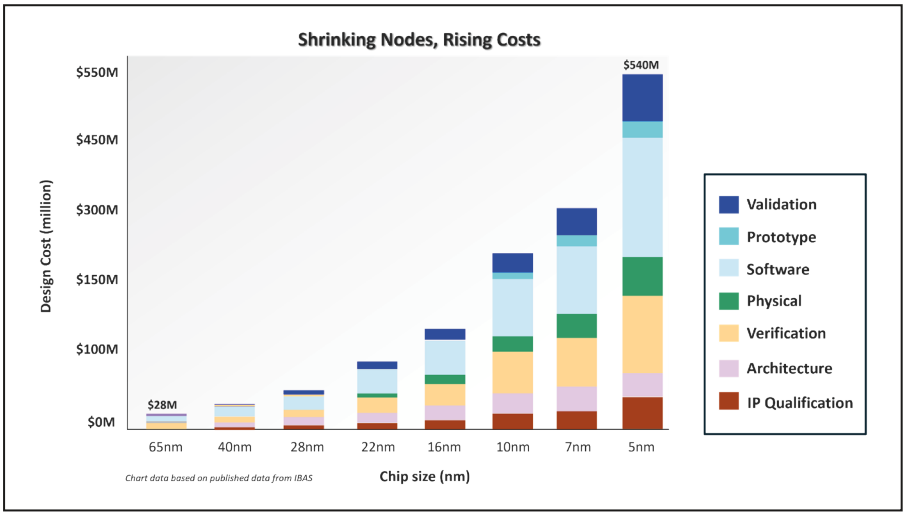

Figure 2: Physical design and verification costs are increasing exponentially with shrinking transistor sizes. SOURCE: Chart generated by Multibeam, based on McKinsey and IDC published data [1]

size constraints. By integrating these chiplets into multi-chiplet packages, manufacturers gain improved flexibility, scalability, and efficiency, allowing them to customize chiplet combinations for specific functions or applications. To meet new demands for high-density chip-to-chip bandwidth and power efficiency, advanced multi-chiplet packaging will turn to high-resolution adaptive patterning offered by the MEBL system.

Leading chipmakers have been shifting towards heterogeneous multi-chip modules (MCMs)—away from the monolithic systems-on-chip (SoC). In MCMs, multiple known-good dies are integrated, interconnected to mimic the functionality and performance of the larger SoCs. These MCMs can also combine various functionalities from different process technologies. The MEBL direct-write system is emerging as a cost-effective lithographic option for patterning substrates for these large MCM assemblies. The technology ensures the necessary resolutions to maintain performance comparable to SoCs and also enables interconnect density and bandwidth approaching metal interconnect layers in monolithic chips. This is a significant advancement in semiconductor manufacturing because it allows the industry to overcome challenges in scaling and opens new avenues for innovations.

Table 1 summarizes the challenges in traditional scaling and solutions enabled by MEBL technology. It also highlights a new era of Moore's Law, in which multicolumn e-beam patterning, when used for "advanced integration" of chiplets, enables a large scale-up in processing capability within each package, as well as high-performance (high-bandwidth, low-power) chip-to-chip interconnects.

Challenges of Traditional Scaling	Multicolumn E-Beam Patterning Solutions
Silicon scaling challenges • Silicon Area Scaling limit of one "classic" reticle (26x33mm) being further reduced to ½ size reticle for High NA EUV • Slowdown in lithographic feature size and overlay scaling	• Heterogeneous integration of singulated dies with interconnect speed and power performance approaching those of on-chip interconnects • Stitching of pre-singulated dies, to create larger effective integrated die • Much-larger, up to full-wafer size, interposers with deep sub-micron resolutions
Power challenges • Large proportion of power consumed by energy required to move bits from one chip to another	• Multibeam's high-resolution writing of radial lines at arbitrary angles, combined with adaptive patterning, resulting in higher interconnect density and 1/50th the chip-to-chip dynamic power (equivalent to a 98% reduction of energy per bit-transfer)
Cost challenges • Rising cost and time of new full designs – exceeding half a billion dollars, coupled with extended timelines	• Rapid cycles of learning, and cost-effective short-run manufacturing for maskless low-cost product development and accelerated time to market

Table 1: Challenges in traditional scaling and solutions enabled by MEBL technology.

Driving advanced integration using multicolumn e-beam lithography

Advanced packaging and MCM assembly are key applications for multicolumn e-beam patterning. Compared to conventional (single-column) electron-beam lithography systems, our technology enables a 100× increase in productivity because of its multiple miniaturized electron beam columns and a write-on-the-fly motion system [6]. In addition to throughput, these advancements enable large interposers of arbitrary shape and size, up to a full wafer—80x larger than today's optical reticle. Other significant benefits include adaptable patterning (>10× higher bandwidth), interconnect density (5× improvement), on-wafer die-die integration (stitched die transcending reticle limit), and reduced energy per bit-transfer by >98% [7]. These capabilities, unique to our technology, enable a new generation of advanced packaging—sometimes called "advanced integration"—that delivers new levels of interconnect performance for today's fastest growing applications.

Multibeam's direct-write patterning systems are capable of full-wafer, maskless lithography with deep sub-100nm resolution and high productivity for production (Figure 3). When paired with Deca Technologies' Adaptive Patterning®, Multibeam's full-wafer field-of-view (FoV), fine resolution, and large depth of focus drive advancements in chip-to-chip communication, die-shift adaptability, energy efficiency, and rapid development in three crucial areas that warrant further elucidation: a) Wafer-scale interposers; b) Adaptable patterning—heterogeneous integration; and c) On-wafer die-die stitching. These areas are discussed in the sections below.

Wafer-scale interposers. By utilizing our high-resolution and wide FoV technologies, it is possible to expand the traditional interposer size by more than 20 times to cover an entire wafer. This full-wafer scaling capability extends the interposer roadmap significantly beyond the current state-of-the-art of 3 to 4 reticle fields, enabling integration of trillions of tightly-interconnected transistors in a single device.

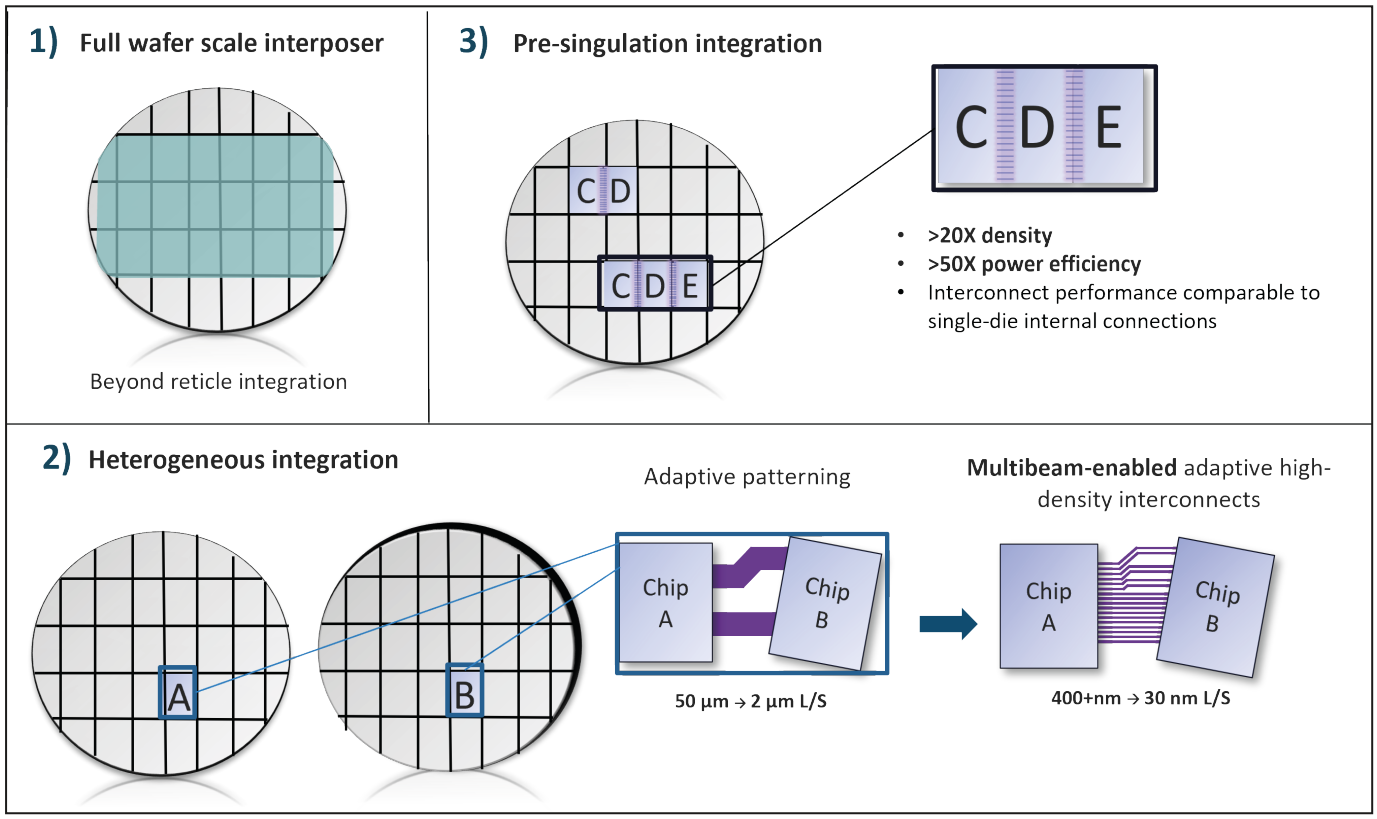

Figure 3: Multibeam's capabilities include: 1) fabrication of wafer-scale interposers; 2) high-resolution adaptive patterning for chip interconnects; and 3) on-wafer die-die stitching, enabling seamless integration of multiple dies on a single wafer.

Adaptable patterning—heterogeneous integration. Our technology can extend adaptable patterning to finer resolutions. This allows for the direct creation of distinctive die-die interconnect patterns to accommodate variations in die positions in chip-first packaging, reducing effective pad pitch and parasitic capacitances on MCM assemblies. Ultimately, this enhancement can lead to an increase in die-die interconnect bandwidth by more than 10 times while simultaneously reducing energy per bit transferred.

In **Figure 4**, we examine one implementation case of Multibeam technology for advanced packaging, whereby Deca's Adaptive Patterning® capability can be extended for further chip-chip communication improvements. Additionally, Deca Technologies uses Gen 2 adaptive patterning, creating 2µm lines/spaces (L/S) with 405nm maskless laser direct imaging photolithography. Similarly, our MEBL system writes a comparable chip-to-chip interconnect pattern with 400nm lines/spaces on a silicon wafer (a 5x shrink in this example—resulting in a 5x improvement in the linear feature size and beachfront interconnect density). Our high-resolution e-beam can scale even further, to higher-density sub-100nm line and space redistribution layers, yielding a 10x or greater improvement in interconnect density, lending significant performance improvements without needing to leverage complex embedded bridge chip structures.

Figures 4c-d show scanning electron microscope (SEM) images of 200nm lines/spaces and 30nm wide lines, respectively—demonstrating the adaptive patterning capability at finer features. **Figure 5** shows that Multibeam's direct-write methodology also allows for greater flexibility in chip design and interconnect routing angles, including patterning of radial lines at various angles—a capability that can be used to further improve interconnect impedance and capacitance.

On-wafer die-die stitching. Integration of two or more die is possible on-wafer during back-end-of-line (BEOL) processing prior to the singulation process. With this integration method, the optical reticle size is no longer a scaling inhibitor. Furthermore, die-die interconnect pitch can match chip metal layer pitch, virtually eliminating previous chip-chip communication or integration bottlenecks. Results of initial modeling show interconnect bandwidth between dies comparable to the performance of interconnects within a single die, which represents a significant improvement when compared to current chip-chip interconnects. This merging of dies blurs traditional advanced packaging boundaries, allowing the industry to overcome on-wafer reticle limitations, resulting in larger effective dies with unparalleled performance.

Advantages of direct-write e-beam lithography

Traditional photolithography, while effective for high-volume production, has become extremely capital intensive at advanced technology nodes. Also, the optical photomasks used to print patterns on wafers have become extraordinarily complex, leading to higher development costs and longer cycle times. In contrast, maskless e-beam direct write facilitates

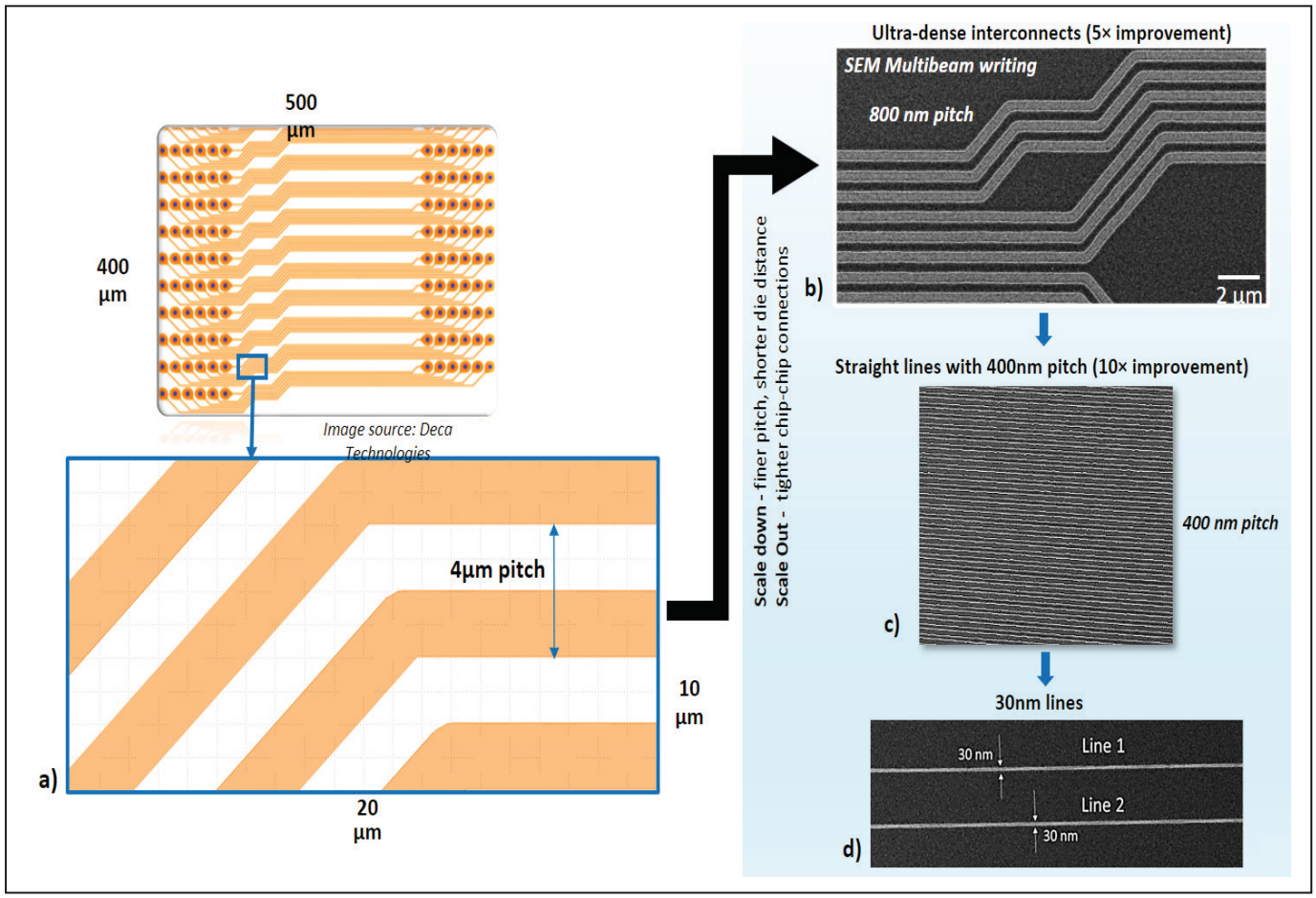

Figure 4: Extending adaptive patterning capability to nanometer scale: a) Deca Technologies' Gen 2 adaptive patterning with 4µm pitch; b) Multibeam's direct-write interconnects with 800nm pitch; c) Multibeam dense interconnects with 400nm pitch; and d) 30nm linewidth for further extendibility.

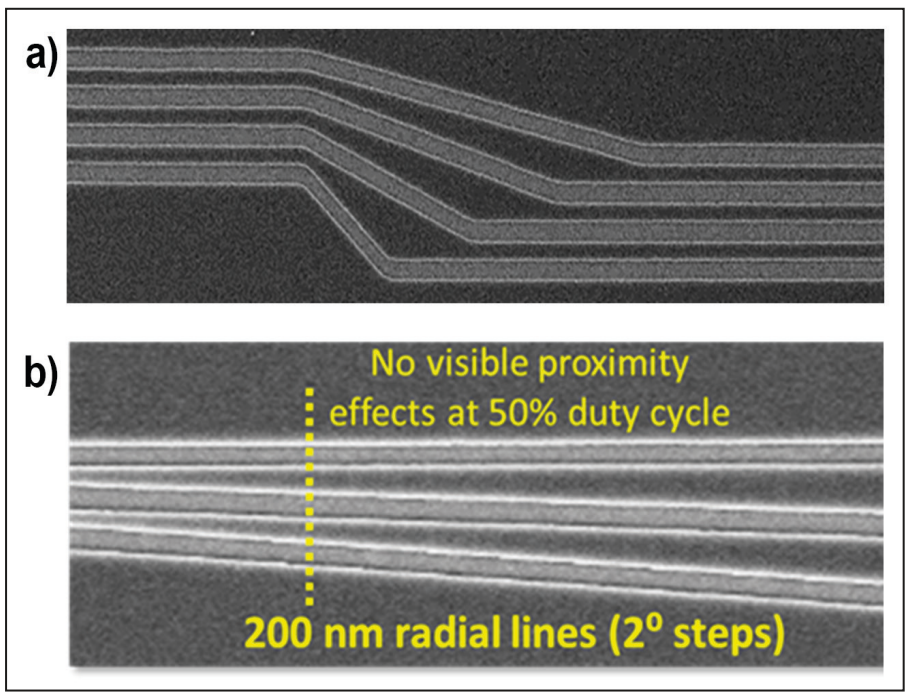

Figure 5: Multibeam write results: a) Direct-write interconnects with flexibility for any angle (800nm pitch example); b) 200nm radial lines with no visible proximity effects at 50% duty cycle.

rapid prototyping, patterning of individually customized chips for design optimization, and quick cycles of learning, significantly reducing the cost associated with the development of advanced packages and speeding time to market of new products.

Furthermore, while traditional R&D e-beam systems are extremely slow, the MEBL system achieves high productivity with a new system architecture that combines multiple miniature e-beam columns with advanced algorithms, parallel vector writing, write-on-the-fly, and interferometric feedback control to meet today's requirements for high speed with high precision.

Modeling results

We modeled the effect of scaling down the line width and pitch on the electrical properties of interconnects and interposers. The results showed an improvement in overall performance [8].

a)

MULTIBEAM

	Gen 1 DIP-DIP on PCB	Gen 2 Flip Chip-Chip on PCB	Gen 3 Flip Chip-Chip on Interposer	Gen 4 "Padless" Chip-Chip on Interposer	Gen 5A E-beam Enabled Chip-Chip on INTERPOSER	Gen 5B E-beam Enabled Chip-Chip on WAFER
Interconnection pitch (µm)	500	300	6	4	0.4	0.3
Chip-to-Chip path length (µm)	20,000	5,000	2,000	1,200	500	200
R_drv (Ohm)	35	35	50	50	250	250
C_IO (pf)	2	1.2	0.62	0.62	0.14	0.14
ESD Capacitance (pF)	1	0.5	0.2	0.1	0	0
Max Bit Rate (GT/s)	1.7	4.5	12	17	10	16
*Bandwidth density (GT/s/mm)	6	14	3000	4000	12,000	25,000
ΔV switched (V)	2.5	1.5	1	0.5	0.25	0.25
Energy per Bit (fJ)	40,000	5,000	1000	200	20	20

*Giga Transfers per second / mm of chip-to-chip beachfront, 1 signal layer

b)

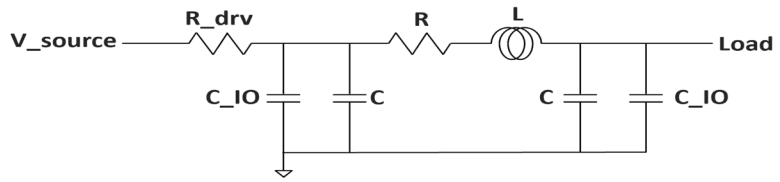

Figure 6: a) Results of five generations of chip-chip communications, modeled using co-planar waveguides to simulate advanced packaging; and b) A simplified RLC circuit of co-planar waveguides for electrical performance modeling of interconnects.

Figure 6 summarizes the results of our modeling analysis of chip-chip communication electrical behavior using co-planar waveguides in LTspice®.

Gen 4 in Figure 6 is enabled by Deca Technologies' adaptive patterning with improved overlay accuracy to eliminate pads. This represents a significant improvement over traditional chip-chip interconnects on interposers. However, much larger improvements in bandwidth density and energy efficiency are enabled with Gen 5 advanced integration. Gen 5A and 5B in Figure 6 captures MEBL-enabled advanced chip-chip communications via interposers and on-wafer die-die stitching. As the pitch and chip-chip path lengths decrease, the bandwidth density increases while energy per bit decreases.

Impact of very high-density interconnects

Bandwidth density—measured in giga transfers per second per mm of chip-chip beachfront—increases by nearly 10X when compared to today's Gen 3 (flip-chip chip-on-interposer). Furthermore, the energy required to transmit a bit decreases by 98%, resulting in significant improvements in overall energy efficiency and thermal control. These advancements in performance are crucial for coming generations of high-performance computing, edge computing, AI-at-the-edge, and other applications. Computing levels required to handle these large and growing language models consume an astonishing amount of energy [9], 70% of which is from data bit transit alone.

In addition to enabling low-power, high-performance communications between chiplets, high-productivity e-beam direct write can also be leveraged for additional emerging methods to improve power efficiency and heat control within a package. These include integrated voltage regulators, closer packing of discrete components, and direct cooling [10]. Direct-write e-beam lithography will also play an integral role in supporting the next generations of HBM stacking and integration. The high precision of e-beam allows for intricate routing of dense interconnects within and between the stacked silicon dies, which will be needed in next-generation HBM devices.

An inflection point for e-beam lithography

The semiconductor industry has reached an inflection point, where challenges in lithographic scaling, chip size limitations, and production expenses combine to hinder quick-turn chip innovation. Advanced direct-write e-beam lithography technology presents a readily deployable solution with the capability to integrate multiple high-performance central processing units (CPUs), graphics processing units

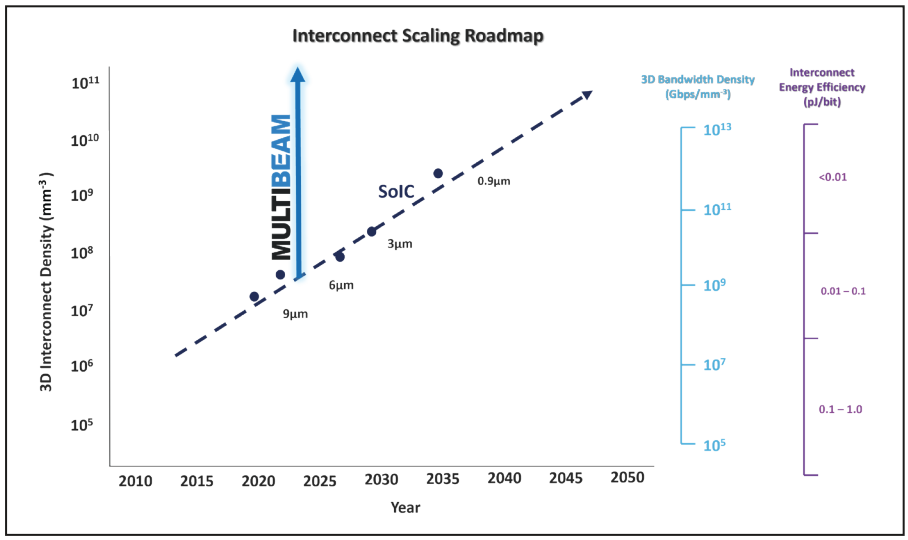

Figure 7: Moore's Law will continue in higher-level chip integration (next-generation advanced packaging enabled by lateral substrate scaling and vertical 3D chip stacking). SOURCE: Chart generated by Multibeam, based on TSMC published data [11]

(GPUs), HBMs, antennas, sensors, and other devices with high-bandwidth, low-power interconnects into an optimized system (**Figure 7**) [11].

For chip makers, multicolumn e-beam lithography is an indispensable tool in the litho toolbox. It complements conventional optical lithography to improve flexibility, scalability, and efficiency in the development and production of chiplets and advanced integrated packages.

Summary

The precision, adaptability, and productivity afforded by our multicolumn e-beam patterning is driving the next generation of advanced packaging, i.e., advanced integration, and pushing the boundaries of leading-edge applications. MEBL is becoming a critical technology that enables continued rapid scaling of HPC and HBM applications at lower power and cost for the AI-era data center.

References

1. S. Göke, K. Staight, R. Vrijen, "Scaling AI in the sector that enables it: Lessons for semiconductor-device makers," McKinsey & Company (April 2, 2021); https://www.mckinsey.com/industries/semiconductors/our-insights/scaling-ai-in-the-sector-that-enables-it-lessons-for-semiconductor-device-makers
2. J. Y. Chen, "Accelerate lithography improvement for high performance computing," SPIE BACUS workshop. Monterey, CA, Oct 1-5, 2023.
3. M. V. D. Brink, "Technology strategy to drive Moore's Law into next decade," ASML ITPC 2023.
4. K. Heyman, "Interconnects essential to heterogeneous integration," *Semiconductor Engineering* (Mar. 28, 2024) https://semiengineering.com/interconnect-essential-to-heterogenous-integration/
5. D. K. Lam, K. MacWilliams, "A new generation of e-beam lithography to enable packaging at the leading edge," *Chip Scale Review*, Vol. 27, No. 4, pp. 7-9 (Jul/Aug 2023); chipscalereview.com/wp-content/uploads/flipbook/34/book.html
6. A. C. Ceballos, K. P. MacWilliams, T. A. Prescop, R. J. Loewen, "Full-wafer, maskless patterning with sub-50nm resolution and large depth-of-focus enabled by multicolumn electron beam lithography," 2024 SPIE Advanced Lithography + Patterning Conf., San Jose, CA, USA, Feb. 25-29, 2024.
7. T. Byambadorj, A. C. Ceballos, K. P. MacWilliams, T. L. Michalka, T. A. Prescop, D. K. Lam, et al., "Transcending the reticle limit in on-wafer die integration and advanced packaging: Full-wafer patterning with high-productivity electron beam lithography," 2024 ECTC, Denver, CO, May 28-31, 2024.
8. P. Triverio, S. Grivet-Talocia, M. S. Nakhla, F. G. Canavero, R. Achar, "Stability, causality, and passivity in electrical interconnect models," IEEE Trans. on Advanced Packaging, vol. 30, no. 4, pp. 795-808, 2007.
9. Z. Ball, "Deploying high-performance, energy-efficient AI," Insights, MIT Technology Review, (Jan. 8, 2024) https://www.technologyreview.com/2024/01/10/1086259/deploying-high-performance-energy-efficient-ai/
10. G. Haley, "Navigating heat in advanced packaging," *Semiconductor Engineering* (Jan. 18, 2024); https://semiengineering.com/navigating-heat-in-advanced-packaging/
11. D. Yu, "TSMC packaging technologies for chiplets and 3D," Hot Chips 33 (Virtual); Aug. 22, 2021.

Biographies

David K. Lam is Chairman & CEO of Multibeam Corporation, Santa Clara, CA. He leads the development of Multicolumn Electron-Beam Lithography (MEBL) systems for semiconductor manufacturing. He is the founder and former CEO of Lam Research and in 2013 was inducted into the Silicon Valley Engineer Hall of Fame for his contributions to the semiconductor industry. Lam holds ScD and MS degrees in Chemical Engineering from the Massachusetts Institute of Technology. Email: dlam@MultibeamCorp.com

Kenneth P. MacWilliams is President and board member of Multibeam Corporation, Santa Clara, CA. He has 10 years of experience in semiconductor device research and 20 years in process and equipment development and management at companies like Applied Materials, Novellus Systems (acquired by Lam Research), Veeco, and Yield Engineering Systems (YES). He holds PhD and MS degrees in Electrical Engineering from Stanford U.