CIRP Annals - Manufacturing Technology 75 (2026) 1027–1053

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

Innovations in advanced processes and systems for semiconductor manufacturing

Albert J. Shih (1)[a,f,*], Wei Gao (1)[b], Gregory W. Vogl (2)[c], Dragan Djurdjanovic (2)[d], Takafumi Fukushima[b], Ajay Malshe (1)[e]

[a] University of Michigan at Ann Arbor, USA
[b] Tohoku University, Japan
[c] National Institute of Standards and Technology, USA
[d] University of Texas at Austin, USA
[e] Purdue University, USA
[f] National Tsing Hua University, Taiwan

ARTICLE INFO

Article history:
Available online 30 June 2026

Keywords:
Semiconductor manufacturing
Integrated circuits
In-line metrology
Process control
Data analytics
Advanced packaging

ABSTRACT

The manufacturing of integrated circuits (ICs) is an enabler of technological innovation and is critical to the resilience of industry and national security. Four levels of excellence are necessary as the foundation for semiconductor manufacturing: ecosystem; fab profitability; IC design for manufacturing and research & development; culture and customer trust. In this paper, the culture, semiconductor manufacturing processes, innovations of equipment, removal processes, in-line metrology for process control, and data analytics are discussed. Future topics, including advanced packaging, sustainability, fab operation optimization, and in-space semiconductor manufacturing, as well as workforce policy and ethics, are elaborated.

1. Introduction

Manufacturing is driven by societal and market needs. Table 1 lists the world's top 10 market capitalization companies, which have changed from oil and automobile industries in 1980, to personal computers in 2000, to mobile devices and the Internet in 2020, and to artificial intelligence (AI) in 2026. Many current top companies need high-performance integrated circuit (IC) chips to maintain their lead in AI. This market pull has driven the growth of semiconductor manufacturing, which is the production of IC or semiconductor chips.

The IC chip is omnipresent as a key enabling part of electronics, the Internet, and cloud-based services in our digital society. IC chips are the foundation for technological innovations, such as AI, high-performance computing, 5G/6G communication, Internet of Things, and autonomous systems (e.g., vehicles, robots, etc.).

IC chips have advanced steadily with the number of transistors doubling every two years (Moore's Law) and an exponential growth in computing performance (Huang's Law) [141]. In the coming decade, to enable AI innovations and capacities, major investments, such as Stargate [122], Colossus [62], and TeraFab in the United States (US) and Nscale [181] in the United Kingdom, have been initiated to build large-scale AI computing centers to process and analyze big data for rapid creation and execution of AI models. Such AI centers require large buildings, consume a lot of energy, and need high-performance and energy-efficient IC chips. New IC designs, such as the gate-all-around (GAA) transistor, backside power delivery, silicon photonics, high-density hybrid bonding, and panel-level packaging, are designed to enhance computing and data transmission speeds and energy efficiency. Advanced IC designs have created challenges for semiconductor manufacturing [110], a critical aspect of AI computing.

The global manufacturing supply chain was disrupted due to the IC shortage during the COVID-19 pandemic. Several countries have recognized the importance of semiconductor manufacturing and established national policies such as the CHIPS and Science Act of the US [106] and the European Chips Act [24], aiming for supply chain resilience, technology leadership, and national security.

Table 1
Top 10 market capitalization over the years and the change in manufacturing driven by society and market needs.

	1980 Oil+Auto	2000 Personal computer	2020 Internet+mobile device	June 2026 AI
1	Exxon	Microsoft	Apple	Nvidia
2	General Motors	General Electric	Nvidia	Alphabet (Google)
3	Mobil	Cisco	Microsoft	Apple
4	Ford Motor	Walmart	Alphabet (Google)	Microsoft
5	Texaco	Exxon Mobil	Amazon	Amazon
6	ChevronTexaco	Intel	Meta (Facebook)	SpaceX
7	Gulf Oil	NTT Docomo	Tesla/Tesla	TSMC
8	IBM	Shell	Broadcom	Broadcom
9	General Electric	Pfizer	TSMC	Saudi Aramco
10	Amoco	Nokia	Berkshire Hathaway	Samsung/Samsung

Blue: Manufacturing companies; Purple: companies that use advanced IC chips

* Corresponding author.
E-mail address: shiha@umich.edu (A.J. Shih).

https://doi.org/10.1016/j.cirp.2026.04.115

A profitable semiconductor fab operation is technically and financially challenging to build and maintain. Every 3–4 years, semiconductor manufacturing advances to a new technology node [8], which was previously defined as the transistor's gate length. As the gate length reached 45 nm, the technology node definition was changed to a representative number, not the size of transistors or a single geometrical feature. In 2026, the 2 nm technology node (with pitch between two metal lines of about 21 nm) is the state of the art in volume production [205]. Lithography is the key technology for miniaturizing feature sizes. As lithography becomes more complex and expensive with increasing miniaturization, advanced packaging becomes critical to sustaining the exponential growth in computing performance by integrating several IC chips into a single package.

Fig. 1 summarizes four levels of excellence in 1) ecosystem, 2) profitable fab, 3) IC design for manufacturing and research & development (R&D), and 4) culture and customer trust in semiconductor manufacturing. The core foundation values are creating a culture and gaining customer trust. The importance and impact of multi-culture in semiconductor manufacturing excellence have been elaborated using Taiwan as an example [153]. Customers are IC designers. Gaining their trust needs to be built on decades of ethical partnerships that benefit both design and manufacturing. The key is to establish a culture of manufacturing service excellence among tens of thousands of dedicated and coherent supply chain partners.

Outside the core is the manufacturing technology for IC design and R&D. Manufacturing activities start 4–5 years before the full-scale production in the innovation phase of IC design using the electronic design automation (EDA) software, which has an IC manufacturer's product design kit (PDK) with intellectual property (IP) for IC designers. Such PDK is proprietary for IC manufacturers and essential for a long-term partnership. About two years before an IC product launch, the IP verification—confirming all process technologies are legally ready—is required to finalize the detailed design (known as tape-out) before mass production. Such proprietary IC manufacturing IP library is created by two parallel manufacturing R&D teams to create the pilot line and mother fab, respectively, for the next two technology nodes, 4–5 years before the mass production.

A profitable fab operation with positive cash flow is the goal for semiconductor manufacturing. Achieving this goal needs: 1) *demand*: stable orders of high-value advanced technology node IC chips, 2) *capacity*: adequate production volume and state of the art equipment, e.g., the extreme ultraviolet (EUV) machines for lithography, to meet customer demands, 3) *workforce*: a team of thousands of well-trained engineers and operators (2000–4000 per fab) working with a dedicated culture for fab operations, and 4) *knowledge*: in the processes, equipment, and logistics of fab operations with good yield and high profit margin (>50%) to support the R&D and construction of new fabs for the next technology node.

An ecosystem with 1) *capital for investment* in new fab and equipment, 2) *education and workforce development* (EWD) of thousands of engineers, 3) *supply chain* of materials, equipment, and other needs for operations, 4) *energy and water efficiencies*, and 5) *societal respect* for excellence in manufacturing is required to support and sustain a globally competitive and profitable fab operation. Taiwan is an example of such an ecosystem with over 50 years of continuing government support (see Sec. 2).

Top IC designers only select the best IC manufacturer to be successful in the fiercely competitive market. Fig. 2 shows Rock's Law [46] or Moore's Second Law, which exemplifies how the cost of building a new fab doubles every four years. The 7 nm, 5 nm, 3 nm, and 2 nm fabs are estimated to cost US$9–11B, $12–15B, $15–20B, and $25–30B, respectively. This capital expenditure is 10–15 times more costly than that of a new automotive plant and will be higher for new fabs of advanced technology nodes. As a result, the cost per 300 mm wafer is steadily increasing for advanced technology nodes (Fig. 3). Once a technology node is launched, the cost per 300 mm wafer decreases over time (Fig. 4), known as the learning curve pricing, the experience curve effect, or Wright's Law [41]. The learning curve is crucial in IC manufacturing fabs to learn process knowledge, improve yield, and increase demand, which in turn enables a positive spiral to reduce cost, attract more customers, and gain more process

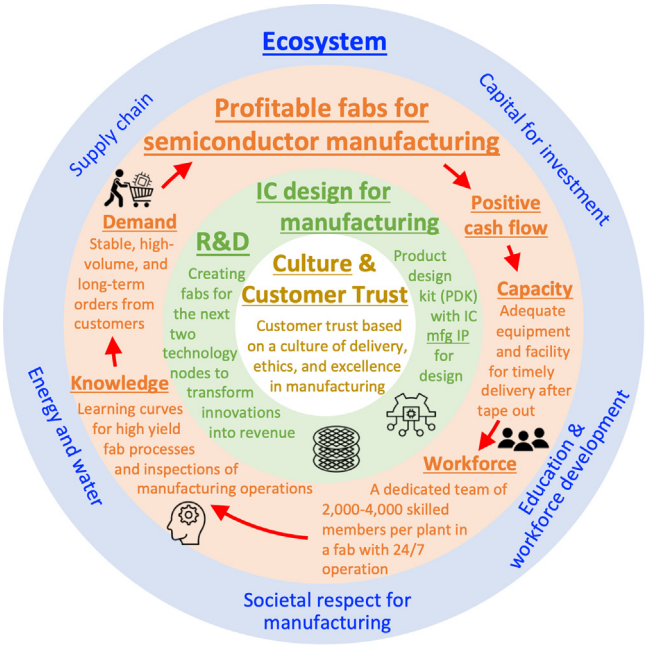

Fig. 1. Four levels of excellence to establish a series of profitable fabs for semiconductor manufacturing.

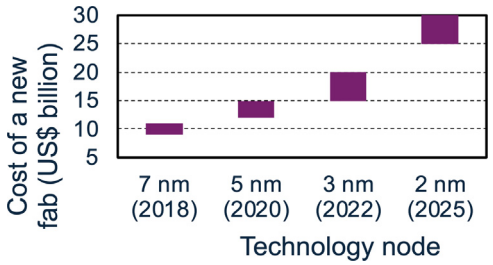

Fig. 2. The increasing cost for a semiconductor fab of advanced technology node.

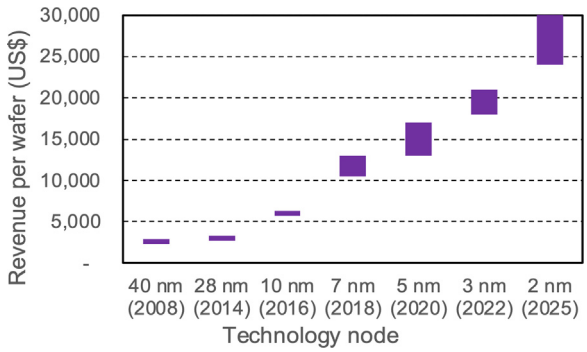

Fig. 3. The cost per 300 mm wafer vs. technology node.

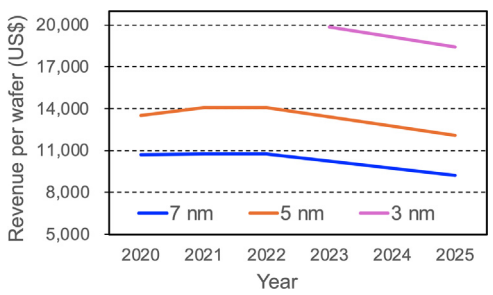

Fig. 4. Wright's law of decreasing cost per wafer over time.

knowledge. Fabs behind in manufacturing technology or newcomers will be challenged to catch up and achieve a positive cash flow in advanced technology node manufacturing due to the high capital investment, long payback period, deep technical barrier, and large workforce.

For the economies of scale, fabs for advanced technology nodes are large with immense production, typically 20,000 to 60,000 (300 mm) wafers per month. Such fabs have a large floor space, ranging from 23,000 m^2 to 93,000 m^2 (250,000 ft^2 to 1,000,000 ft^2) with mostly Class 100 and some Class 0.1 (ISO Class 1) clean rooms for critical (e.g., lithography) operations. Clean rooms in the fab are unique with high ceilings (over 5 m for the tall EUV equipment and overhead hoist transport (OHT) automation), high accuracy in temperature and humidity, heavy load-bearing capabilities (e.g., for the 200-ton EUV machine), and laminar air flow in from the ceiling and out through the perforated raised floor. These fabs have extensive piping made of special corrosion-resistant materials for industrial etchants and gases, most of which are highly toxic, for fab operations. Building a new fab takes 1–2 years. Installing equipment and ramping up production with high yield takes another 1–2 years. The return on investment time is long and may take over a decade. Behind profitable fabs are visionary leaders, dedicated workers, and committed investments.

This paper highlights culture and innovations in semiconductor manufacturing processes and systems. As shown in Fig. 5, after presenting culture (Sec. 2) and semiconductor manufacturing processes (Sec. 3), the following four sections cover the equipment (Sec. 4), removal processes (Sec. 5), inspection (Sec. 6), and data analytics (Sec. 7) of semiconductor manufacturing processes and systems. The future technical (Sec. 8) and societal trends (Sec. 9) are summarized in Secs. 8 and 9.

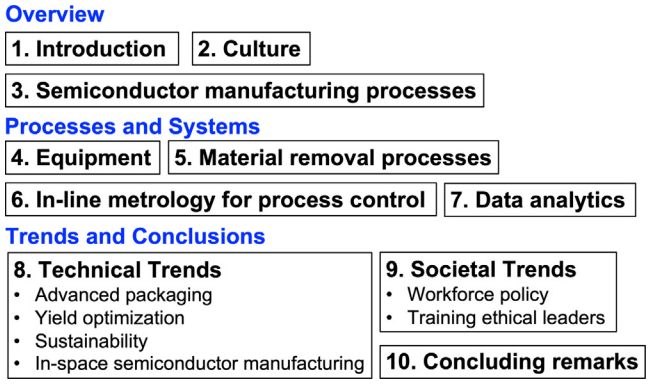

Fig. 5. Organization of sections.

2. Culture in semiconductor manufacturing

Culture has been raised repeatedly in CIRP Collaborative Working Group (CWG) meetings as a key factor in semiconductor manufacturing. As shown in Fig. 1, culture and customer trust are cornerstones for building and sustaining profitable IC fabs following the cadence of technology node advancements every 3 years. Although a one-time significant investment can establish a fab with an advanced technology node, within 3 years, another new, more expensive fab with a higher technology node needs to be built to remain competitive. These fabs also need to remain profitable with sufficient demand and good yield.

Culture is complex and deeply rooted in a region's history. The diversity to respect five cultures—i.e., 1) US culture of innovation, design, software, higher education, and professional training; 2) Japanese culture of pursuit of perfection and meticulous craftsmanship (kodawari); 3) European culture of excellent small and medium sized enterprises (Mittelstand) in the supply chain; 4) traditional Chinese culture of integrity and trustworthiness to grow business with customers; and 5) modern Chinese culture of speed and internal competition—in Taiwan has built the excellence in semiconductor manufacturing [153].

"All I know is that I know nothing" (attributed to Socrates) is a required humble attitude in semiconductor manufacturing culture. Past successes do not correlate with future performance. There are diverse opinions of future paths from many vocal and proud past experts who had successes in semiconductor manufacturing. Current leaders in semiconductor manufacturing are very humble and quiet. There is no genius or lasting expert in semiconductor manufacturing.

The workforce is a critical part of the government's policy for semiconductor manufacturing. People do make a tremendous difference. A successful government industrial policy must include technology workforce policy, which will be discussed in Sec. 9.1, in semiconductor manufacturing. An industrial policy that selects winners in technology and company sectors for government support requires deep expertise and swift decisions in the rapidly changing semiconductor industry. However, a government may need to request public input, build consensus, and may not be able to make legally binding decisions quickly. The timely distribution of government funding is another challenge for the semiconductor industrial policy.

Continuity of R&D support and close partnership between industry and academia are two key elements of a successful technology workforce policy. The continuous R&D support is more effective than a single significant funding spike. When it comes to semiconductor policy, the slow and steady state tortoise may beat the hare. Taiwan is a model of such a steady policy. As shown in Fig. 6, Taiwan government has maintained a steady support over a 50-year span of research programs starting from the creation of the Industrial Technology Research Institute (ITRI) in 1973 and licensing of the 7.5 μm technology node (two and a half generations behind the state of the art at that time) from RCA in 1976 for the first 75 mm diameter wafer fab in Taiwan. In 1987, ITRI's workforce, equipment, and knowledge of the 150 mm diameter wafer fab were transferred to Taiwan Semiconductor Manufacturing Corporation (TSMC), a spin-off from ITRI, as TSMC's Fab 1, located inside and leased from ITRI.

Success in semiconductor manufacturing can be built with long term partnership among the IC industry, government (for industry and workforce policies), and academia (for workforce training). Such partnership needs several decades (not years) of continuing support to build the technology (to be discussed in Secs. 3–8), workforce (Sec. 9), and supply chain for semiconductor manufacturing.

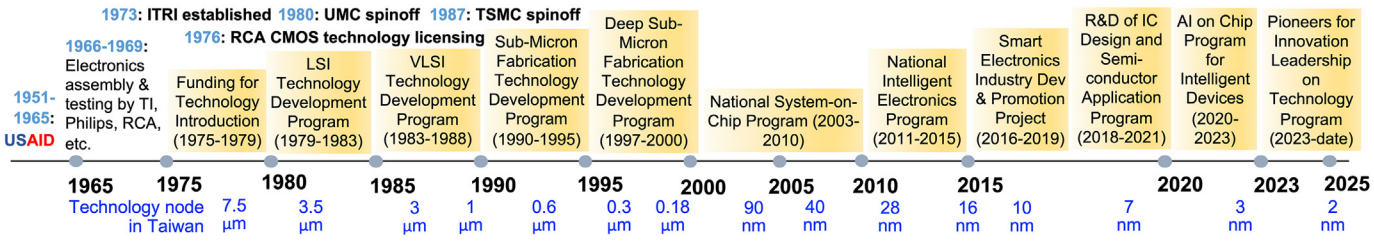

Fig. 6. Taiwan's 50 years of semiconductor manufacturing national R&D programs and advancements in the technology node of IC in mass production.

3. Semiconductor manufacturing processes

Silicon (Si), Silicon carbide (SiC), and Gallium nitride (GaN) are the three most common semiconductor substrate materials. Fig. 7 compares five key material properties of these materials. Si has good electrical and thermal properties as well as reasonable cost, making it the most versatile and popular semiconductor material. However, Si has a low breakdown field and energy gap under high voltage. On the contrary, SiC has a high breakdown field, energy gap, thermal conductivity, and melting point. SiC and GaN are ideal for high-power and high-frequency electrical switches for RF power amplifiers for 5G/6G communication and other applications. GaN is suitable for high-frequency applications but has poor thermal conductivity. Hence, semiconductor manufacturing processes for Si and SiC are elaborated next.

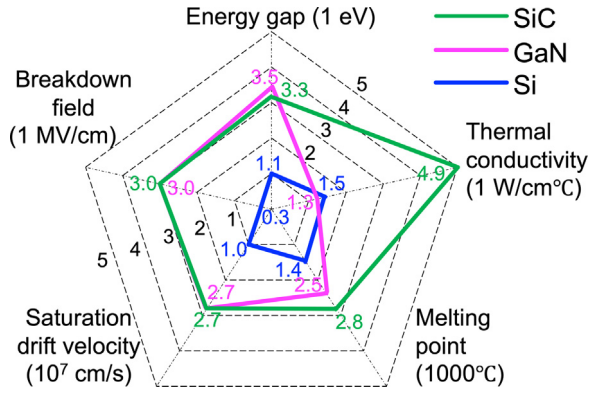

Fig. 7. Comparison of key properties of three semiconductor materials.

Fig. 8 summarizes the Si IC manufacturing and packaging processes. First is Si wafer manufacturing. Si is purified to 99.999999999% high purity—also known as nine nines, 9N, or 11N—and grown as high-quality single-crystal boules using the Czochralski (CZ) method. A series of processes, including outside diameter (OD) grinding, cropping, wire saw, face/notch grinding, and chemical mechanical planarization (CMP), are used to fabricate the 300 mm diameter Si wafer from the single-crystal Si boule to below 0.1 nm R_a to supply IC manufacturing fabs.

Lithography is the most critical IC manufacturing process due to the need for miniaturization in advanced technology nodes. Lithography R&D and production groups are the largest in fabs of the advanced technology node. Grinding, dicing, CMP, and etching are material removal processes. Physical vapor deposition (PVD), chemical vapor deposition (CVD), and ion implantation are material addition processes. Fabs require an automated material handling system (AMHS), cleaning, and inspection. In wafer fabs, an AMHS includes the OHT, wafer handling robots, and automatic guided vehicles (AGVs). Cleaning, which requires a substantial amount of water, is a necessary process and significantly impacts

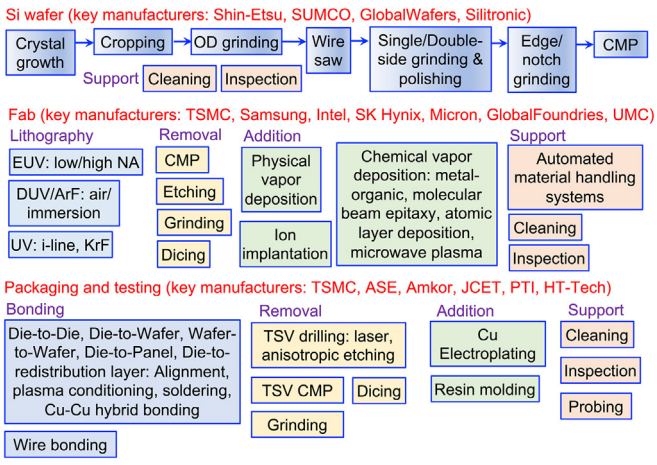

Fig. 8. Si-based semiconductor manufacturing processes.

yield. Inspection, to be discussed in Sec 6, utilizes advanced metrology for quality control. Each wafer takes many process steps and a long time to complete. For example, the 3 nm technology node logic IC often has over 100 layers (including 19–25 EUV layers [2]), requires over 1000 steps, and takes months to complete. At the end, yield is the outcome determining the IC fab effectiveness.

The front-end-of-line (FEOL) processes fabricate the individual features on a semiconductor wafer (via lithography, deposition, etching, etc.). The back-end-of-line (BEOL) processes connect or package FEOL-created components to manufacture IC chips.

Packaging is assembling an IC chip, also known as the die (a small functional block in the wafer), to the substrate (e.g., the printed circuit board (PCB)). Packaging has evolved from wire bonding and surface mount processes to solder ball joining and Cu-Cu hybrid bonding of IC chips to the substrate [121]. The needs of AI computing have driven new demands in advanced packaging, such as the panel-level packaging [1], high-density hybrid bonding, and silicon photonics, which will be discussed in Sec. 8.1. Removal processes in packaging include drilling an array of miniaturized high-density through-silicon via (TSV) holes. TSVs filled with Cu (by electroplating) will be polished by CMP. Molding to fill the thin gaps with resin is used to fix the chip. Cleaning, inspection, and probing for evaluation of electrical properties are three support processes in packaging.

The manufacturing and packaging processes for SiC power electronics are shown in Fig. 9. The single-crystal SiC boule is grown using the physical vapor transport method by sublimating the SiC powder and depositing the vaporized SiC onto a seed. After crystal growth and OD/surface grinding, slicing the SiC boule to wafer with minimal kerf loss and depth of subsurface damage is a critical process. SiC is hard and brittle compared to Si. The traditional loose diamond wire saw for wafer slicing [154] is slow. Hence laser slicing of a wafer from a SiC boule using many small and connected cracks generated by a pulsed laser focused underneath the ground flat surface has been invented [16]. The SiC wafer and boule surfaces after laser slicing are then flattened by grinding to prepare for the subsequent CMP and laser slicing. SiC transistors do not require advanced lithography for miniaturization like Si. Grinding and dicing are two processes that have long cycle times and are critical to the cost. Both will be discussed in Sec. 5.

There are differences and similarities in manufacturing of Si and SiC ICs. Single-crystal Si and SiC wafers grow very differently, Si based on CZ method and SiC based on physical vapor transport method. Transistor manufacturing is advancing regularly for Si technology node, but not for SiC. The wafer cost is critical for SiC. Both Si and SiC wafers require precision grinding and CMP processes. The type of grinding wheel, CMP slurry, and equipment are very different for Si and SiC wafers.

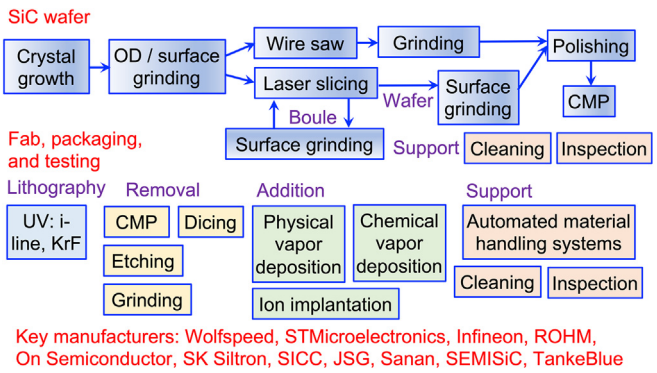

Fig. 9. SiC-based semiconductor manufacturing processes.

4. Equipment for semiconductor manufacturing

Equipment for semiconductor manufacturing (ESM) is a key enabler of high-volume production. Suppliers invest in substantial R&D to design and manufacture advanced ESM for precision, speed, and reliability [166]. This ESM supply chain is dominated by a small

number of companies [190] in three countries—the US, Japan, and the Netherlands [153].

Two key metrics of semiconductor manufacturing are quality (yield) and throughput, which have been improved due to advancements in ESM [125]. The following subsections introduce examples of recent ESM innovations that enable further increases in quality and throughput of IC production, specifically for lithography (motion and controls), deposition and etching (sensors and monitoring, handling and contamination), and packaging (hybrid bonding, grinding, and dicing).

4.1. Lithography ESM

Numerous lithographic techniques exist, each with their own advantages and disadvantages for patterning structures on a substrate. The patterning of photosensitive resist via light through a mask (photolithography (PL)) or scanning with electrons (e-beam lithography (EBL)) to define the geometrical configurations of transistors and interconnects is the most common lithography method [158]. To increase the resolution, the light wavelength for lithography has reduced from 365 nm (i-line UV) to 248 nm (KrF UV), 193 nm (ArF deep UV or DUV), 134 nm (ArF in water, also known as immersion DUV), and 13.5 nm (EUV). The numerical aperture (NA) is another key factor for miniaturization in lithography. For example, the Advanced Semiconductor Materials Lithography (ASML) has the 0.55 high NA EUV machine [3] with large optical mirrors and high cost (Table 2). The next generation 0.75–0.85 hyper NA EUV [95] needs to further enlarge the size of mirrors, which is challenging and expensive for mirror manufacturers (e.g., Zeiss). Two key equations for lithography are:

$$\mathrm{HP} = k_1\lambda/\mathrm{NA} \tag{1}$$

$$\mathrm{DOF} = k_2\lambda/\mathrm{NA}^2 \tag{2}$$

where HP is the half pitch, k_1 is the Raleigh parameter, λ is the wavelength of exposed light, DOF is the depth of focus, and k_2 is the non-paraxial DOF scaling coefficient.

The trend and cost of miniaturization using lithography are illustrated in Table 2. The HP for immersion DUV ($\lambda = 134$ nm) with $k_1 = 0.28$ is 40 nm. Using EUV ($\lambda = 13.5$ nm) with $k_1 = 0.40$ and a regular NA (= 0.33), the HP is reduced to 16 nm. The high NA (= 0.55) and hyper NA (= 0.8) EUV further reduce the HP to 10 nm and 7 nm, respectively.

Table 2
Half pitch and depth of focus of immersion DUV and EUV.

	λ (nm)	k_1	NA	k_2	HP (nm)	DOF (nm)	Equip. cost (US$M)
Immersion DUV	134	0.28	0.95	0.20	40	78	≅ 60
EUV	13.5	0.40	0.33	0.75	16	93	≅ 200
High NA EUV	13.5	0.40	0.55	0.75	10	33	≅ 400
Hyper NA EUV	13.5	0.40	0.80	0.75	7	16	–

The increase in NA reduces DOF [23] and affects yield. Low DOF makes wafer grinding and CMP critical to achieve a flat, smooth wafer surface during exposure for high-yield production. Optically, the low DOF is associated with yield and requires precision grinding and CMP to flatten the wafer. The DOF increases from 78 nm for immersion DUV to 96 nm for EUV with 0.33 NA, but then decreases to only 33 nm and 16 nm for 0.55 high NA and 0.80 hyper NA EUV, respectively. Low DOF will affect yield and cost.

Since EUV light does not propagate well through air, the EUV machine operates under a vacuum. The EUV mask reflects light, and the reflected image is then demagnified and focused onto the photoresist layer by a series of mirrors with multilayered coating [158]. High NA EUV can reduce multiple patterning (a technique to enhance the resolution) and the exposure dose while increasing the resolution. For example, the 0.55 NA EUV increases the geometrical density in single patterning by a factor of 2.8 compared to their 0.33 NA EUV [173].

For EUV, a pellicle membrane is required to protect the mask from contamination (such as residual tin droplets from the generation of EUV light) while allowing high EUV transmission. The composite pellicle, made of materials such as silicon nitride and a thin metal layer (e.g., rhenium), has achieved EUV transmission exceeding 90% at 400 W of power (Fig. 10) [173].

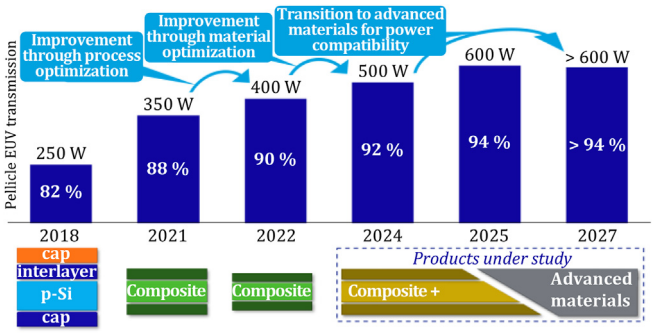

Fig. 10. ASML's roadmap for EUV pellicles [173].

One critical component of a high NA EUV machine is the reflective mask, the pattern that is created by EBL. Advancements include multi-beam mask writers (MBMW), such as NuFlare Technology's MBM-2000 MBMW, which achieves a throughput of 8.7 h in a writing area of 104 mm × 130 mm, a global positioning accuracy of 1.2 nm, a local positioning accuracy of 0.5 nm, and a local critical dimension (CD) uniformity of 0.61 nm for EUV masks [188]. As seen in Fig. 11, the electron beam emitted from the electron source is accelerated to 50 keV and split into about 250,000 square beamlets by a shaping aperture array. Each beamlet is then independently deflected in coordination with stage motion to form an image pattern in the resist on the mask [188]. Another innovation is the ZEISS e-beam based MeRiT® neXT repair tool that enables high-precision repair of ultra-small defects of EUV masks via low-energy electron beams due to the absence of physical sputtering and ion implantation [30].

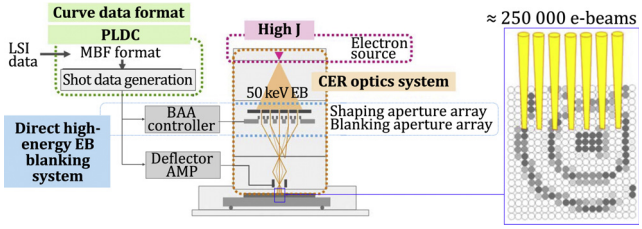

Fig. 11. Schematic of the multiple e-beam mask writing system [188].

EBL is a serial process with low throughput, which hinders its widespread use. It may be advantageous for applications with lower accuracy requirements, such as large multi-chip assemblies. Multicolumn electron beam lithography systems pattern substrates with resolutions less than 100 nm and high throughput due to up to 25 simultaneous electron beam columns [20].

Disadvantages arise from multiple constraints in the EUV ESM. For example, 10 nm HP patterning requires high-power EUV light sources to achieve high throughput, given the expected resist's sufficiently low photon-shot-noise-induced line edge roughness. Accordingly, an output of 400 W to 500 W has been achieved using a laser-produced plasma light source [98]. Polarized light is not required for the initial application of high NA EUV lithography; however, this may hinder the progress for technology nodes below 2 nm [98].

Nanoimprint lithography (NIL) is another advanced lithography process concept that can be seen as a potentially cost-effective alternative to EUV. In NIL, a substrate is coated with resist, and when the etched mask is stamped onto the resist-coated substrate, the resist fills the relief patterns via capillary action. After UV curing, the mask is removed by etching, leaving a patterned resist on the substrate [112]. The NIL (see Fig. 12) achieves a 14 nm linewidth resolution with an overall accuracy

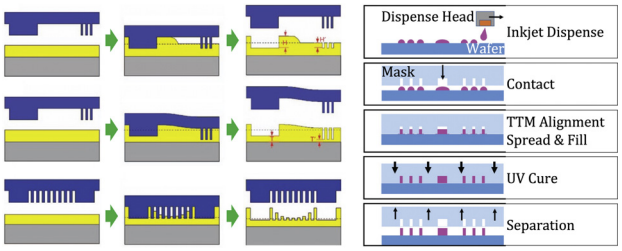

Fig. 12. (a) Defects for the case of spin-on imprinting [157] avoided in (b) Canon's nano-imprint lithography process with inkjet-printed pL-volume drops of low viscosity UV resist [25].

of less than 4 nm. It is expected to enable a minimum linewidth of 10 nm, corresponding to a 2 nm technology node [25]. NIL is a potentially cost-effective alternative to high NA EUV for replicating nanoscale features [98]. Another advantage of NIL is that it enables 2D or 3D circuit patterns to be formed in a single imprint with particle defects reduced to about one per every 10,000 wafers [25]. Defects in NIL can be reduced using the precision inkjet in the pL-level, as shown in Fig. 12.

4.2. Motion and controls

Precise, accurate, and efficient motion and controls are vital for quality and throughput in ESM, including lithography machines [55]. Meeting requirements for motion and power is becoming increasingly challenging, with a need for both higher speed (throughput) and greater accuracy. To increase exposure throughput, ASML targeted a 4 × increase in reticle/mask stage acceleration, up to 150 $m{\cdot}s^{-2}$ (15 g), and a 2 × increase in wafer stage acceleration, up to 69 $m{\cdot}s^{-2}$ (7 g), without causing detrimental vibration or heating. The EUV machine frames are electron-beam welded in vacuum to support the synchronized motion of wafers and reticles [98]. Mask scan speeds greater than 1 m/s are used for high NA EUV exposure tools while achieving sub-nm position control for sufficient overlay and imaging. In fact, synchronizing wafer and reticle motions to within a nanometer over a nanosecond, while achieving high wafer and reticle accelerations, enables a throughput of about 300 wafers per hour.

One key metric of lithography quality is the edge placement error (EPE), which represents a combination of overlay errors, scanner-to-scanner CD matching, global and local CD uniformity, and residual errors after optical proximity correction [173]. Minimizing the EPE to less than a nanometer, while aiming for increased wafer throughput, requires the use of controls to reduce each error component. Examples include ASML's DUV scanners with lens distortion manipulators, which provide additional freedom to correct overlay errors [173] and offer greater focus control, as the DOF at 0.55 NA is only about 33 nm (Table 2) [98]. To reduce positioning errors in mask writing, a thermal effect correction was developed and integrated into the MBMW platform to compensate for beam-induced heating of the mask substrate [151].

Increased quality also requires improved position measurements for control. High-resolution encoders and correction algorithms have enabled real-time accuracies of less than a nanometer in lithography stages [168] and in wafer positioning with an accuracy of 60 pm. For improved absolute 2D stage positioning, a commercial Fizeau phase-shifting interferometer (Zygo Verifire^{TM}) was used for self-calibration of a large-scale 2D variable-line-spacing grating for use as an absolute optical encoder, as shown in Fig. 13, e.g., in advanced multi-DOF motion control, which was able to mitigate the effects of tilt and noise errors [191].

As moving components become lighter for higher throughput, they are also more prone to non-rigid dynamics. Wafer-related motions in ESM are typically close to a fixed position, which could benefit from position-dependent and task-flexible control that is also automated and interpretable for industrial adoption [142]. Accordingly, Poot et al. [142] developed a motion control framework for

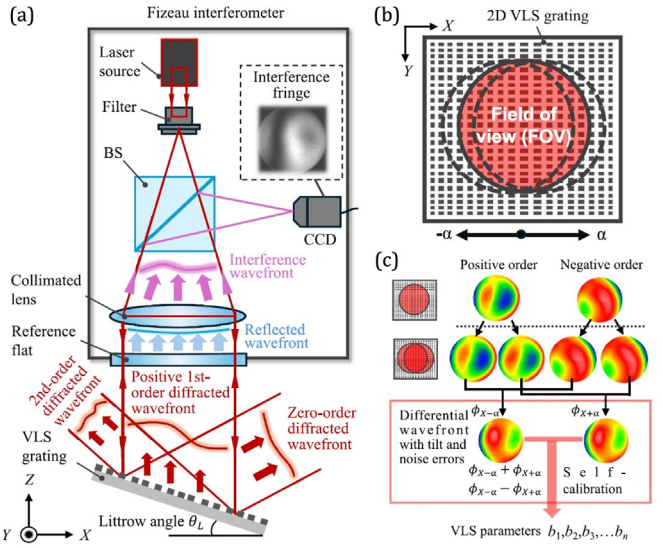

Fig. 13. (a) Measurement of the positive first-order diffracted wavefront from the target variable-line-spacing grating via a Fizeau phase-shifting interferometer, (b) measurement scheme of the conjugate differential self-calibration method in the X-direction, and (c) self-calibration of the variable-line-spacing parameters [191].

interpretable and task-flexible position-dependent feedforward control via Gaussian process regression, enabling accurate control over the ESM operating range.

Pushing the limits of precise, accurate, and efficient motion also benefits from increased temperature control. Temperature is a physical quantity that must be precisely controlled by chillers, which regulate the temperature of the heat transfer fluid. A step chiller for rapid wafer temperature control was developed to mix hot and cold fluids from two sources using feedforward control (in contrast to manual tuning for each temperature band, e.g., with a 5°C interval) to minimize temperature variations and ensure process consistency [193].

4.3. Sensors and monitoring

Semiconductor devices are created in strict environments, typically a vacuum, as an inherent process enabler to fabricate the individual features and to avoid device contamination by particles and gases. Hence, innovations for monitoring the dynamics of plasma for etching or deposition within a vacuum chamber have values in process development and diagnostics. Kim et al. [87] compared GHz-range broadband S-parameters, measured from the chamber viewport before the process runs, to those from the normal state to detect abnormal chamber conditions such as wafer shifts. A wafer-shaped sensor for in situ plasma monitoring has been created. The system thickness should be less than 1.6 mm for automated usage, enabling transfer via computerized robot arms. Park et al. [139] developed a 1.4 mm thick wafer-shaped sensor for 2D plasma monitoring with robot transfer capability, as shown in Fig. 14. The sensor contains 21 SiO_2-based plasma probes distributed on the wafer along with batteries, ICs, a coil for wireless charging, and a temperature sensor. Tests validated that the sensor's outputs for plasma density, ion flux, and electron temperature during

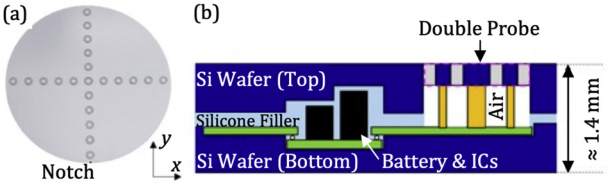

Fig. 14. (a) Wafer-shaped sensor for 2D plasma monitoring and (b) conceptual cross-sectional diagram of the sensor [139].

repetitive monitoring of argon plasma were similar to theoretical values. Wireless sensors for plasma monitoring for closed-chamber calibrations are similar in concept to those that have existed for monitoring tip/tilt, vibration, and humidity at the wafer and mask (reticle) levels [127].

However, if the vacuum does not reach the required level, such as 10 Pa for etching or 0.1 Pa during metal sputtering, the process control may fail due to insufficient etching or deposition rates. Thus, innovations for sensing the vacuum pressure at multiple locations inside the chamber, particularly near the wafer, may be advantageous, rather than at the vacuum pump. Xu et al. [186] developed a MEMS thermopile Pirani sensor that measures vacuum pressure through the pressure-dependent thermal conductivity of its small array of composite nano forests. The sensor has a high sensitivity of 50 mV/decade and a low detection limit of 0.025 Pa, making it capable of in-situ monitoring of ESM chamber pressure. In contrast, Qin et al. [143] created a MEMS diaphragm-based vacuum sensor with two weakly-coupled resonators that has a lower resolution of 0.1 Pa in the range of 0.3 Pa to 103 Pa, but also works in the range of 103 Pa to 105 Pa with a resolution of 2.0 Pa, thus enabling wide-range vacuum sensing.

Other equipment innovations for process sensing and monitoring are more indirect regarding the fabrication steps of etching and deposition but are still highly impactful. For example, cooling fluid is used for temperature control in ESM and the fluid leaks from small tubing cracks within the resist coater units can cause wafer process failures. Typically, tubing is manually inspected during regular maintenance, but perfect prevention is unattainable. Because small cracks can cause bubbles in the water flow, Aung et al. [10] developed fiber optic evanescent sensors to detect bubbles as an early detection of cracks in water tubing. The sensor system was implemented at GlobalFoundries for three years, resulting in 100% early detection of water jacket leaks and a subsequent reduction in yield excursions related to tubing issues to 0%. Manual water jacket inspections were eliminated.

4.4. Handling and contamination of ESM

ESM handles the liquid, gas, solid, and plasma states of matter. Many ESMs for processes such as dry etching use highly reactive fluoride plasma, which continuously consumes chamber components until they are eventually replaced. Hence, innovations in non-corrosive and durable chamber materials are required to reduce undesired particles during manufacturing processes.

Typically, stainless steel (SS) is used in ESM's gas delivery systems and chambers, and three types of anti-corrosion methods for SS exist: increased material corrosion resistance, surface smoothing, and protective films and coatings [116]. Other metallic contamination must be avoided, such as from metallic impurities in injected-molded perfluoroalkoxy alkane fittings used in wet etching and cleaning processes [104], or from ceramic parts, e.g., quartz windows, within etching chambers, using plasma-resistant glass [38]. Standard protective coatings on SS are produced by anodization or thermal spraying with alumina and yttria, which may degrade through cracking and porosity. Leppilahti [97] deposited atomic layer deposition (ALD) coatings of alumina and yttria to improve the uniformity, density, and purity of these protective layers. The high plasma resistance of the alumina and yttria coatings was demonstrated by sufficiently low etch rates, e.g., 3.9 nm/min of yttria in a trifluoromethane process.

New equipment was also developed to handle process materials and enable new processing techniques. Okabe and Somaya [132] developed a high-vacuum-compatible ionic liquid (IL) circulation system featuring a spiral-grooved rotational pump to address the need for wet processing under vacuum. The system was shown to circulate over 100 mL of IL under direct vacuum exposure, specifically at 10^{-5} Pa, and with 10^{-7} Pa partial pressures for the IL-derived products during circulation. The high-vacuum-compatible IL circulation system (see Fig. 15) may be used for wet processing under vacuum.

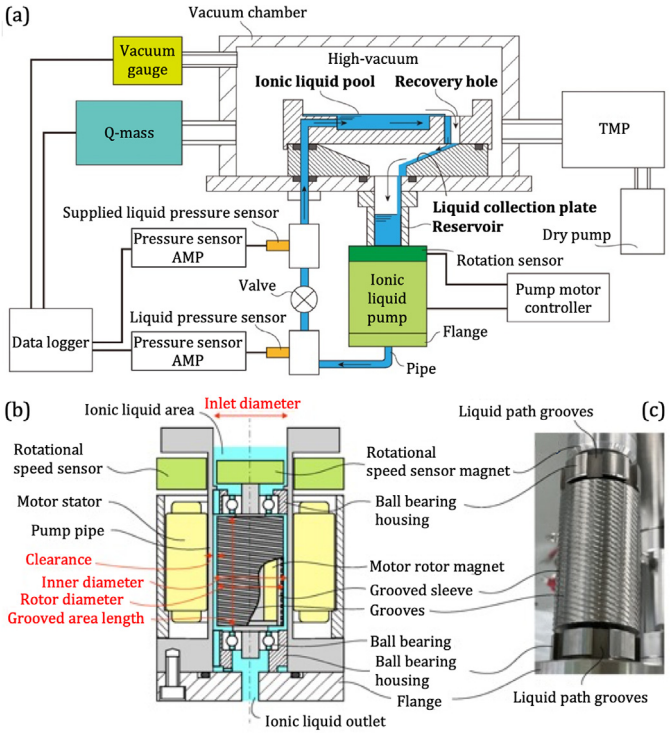

Fig. 15. Schematic of (a) high-vacuum-compatible IL circulation system with (b) spiral-grooved rotational pump, and (c) photograph [132].

4.5. Etching ESM

Improvements in plasma etching have focused on enhancing pattern fidelity, material selectivity, and minimizing plasma-induced damage. Atomic layer etching (ALE), which enables ultra-high materials etching selectivity and rate control, has made significant advances for high-volume manufacturing [130], while neutral beam etching (NBE) greatly reduces surface damage during etching by neutralizing plasma ions, which is utilized for GaN-based devices [96]. Recent ALE ESMs have smaller transition times between their ALE cycles (fast gas transition, fast plasma ignition, stabilization, etc.) and better ion energy control [130].

Further ESM advances include improvements in etching performance. For example, etching of deep aspect ratios up to 60:1 requires more uniform etch processes. Many plasma etchers use magnetic coupling to a plasma for etching, such as via the RF excitation (at 13.6 MHz) of copper coils in inductively coupled plasma (ICP) etching systems from Lam Research, which can create deep silicon trenches [120]. To improve the uniformity of the magnetic field for a more uniform plasma and etching process, coil design optimization reduced the standard deviation of the magnetic field magnitude above the wafer by 22%.

4.6. Hybrid bonding ESM

Recent innovations in IC packaging have focused on hybrid bonding. ESM for both wafer-to-wafer (W2W) and die-to-wafer (D2W) hybrid bonding must accurately align clean, flat surfaces to produce hybrid interconnections via CMP, cleaning, dicing, surface activation, alignment, pre-bonding, and low-temperature annealing [209]. The D2W/W2W ESM (by SUSS MicroTec) integrates all hybrid bonding processes in a single tool: W2W, collective D2W, and sequential D2W [160]. The high-precision flip-chip die bonder is fully automated, encompassing cleaning, activation, alignment, bonding, and measurement processes. Alignment accuracies achieved < 50 nm for W2W wafer bonding and < 100 nm for D2W/W2W die bonding. Another automated wafer bonding ESM has similar capabilities but with an additional hybrid bonding step for die partitioning [64].

Furthermore, CMP is a crucial process for obtaining planar surfaces, which are necessary for proper bonding of the silicon oxide

(SiOx) dielectric and Cu metal. The SiOx requires a surface roughness of less than 1 nm. A flat surface is needed to minimize Cu dishing and surrounding air gaps. Notable CMP machines include the Opta [6] and the F-REX [49] CMP platforms.

4.7. Grinding and dicing ESM

As part of the innovations for chip packaging, wafer thinning by grinding is necessary to achieve the specified dimensions in chip design to the low electrical resistance and capacitance after packaging. Wafer grinding is performed using a cup wheel with diamond segments (mostly with vitreous bond) on a spindle, as shown in Fig. 16. This is called back grinding because it is on the opposite side of the transistor layer of the wafer. The wafer is fixed on a rotating vacuum chuck made of porous ceramics. Air bearings are common for grinding wheels and chuck spindles, providing a damping effect to minimize surface damage on brittle Si and SiC wafers. The chuck typically tilts at a small angle α (see Fig. 16) to slightly bend and create a small elastic deformation in the wafer during grinding. An adjustable support with sub-μm accuracy is studied to deform the vacuum chuck and wafer during grinding, hereby adjusting the wafer's total thickness variation [101].

The wafer grinding machine in production has multiple spindles and chuck tables. For example, the DISCO DFG8830 wafer grinder [44], as shown in Fig. 17, has four spindles for a series of grinding wheels with different grit sizes, typically ending at ANSI mesh 30000 (or finer) for Si and SiC wafers. A wafer transfer robot uses an additional chuck table for loading and unloading. Wafers are stored in the wafer carrier case, also known as a cassette or front-opening unified pod (FOUP), which is transferred in and out of the grinding machine

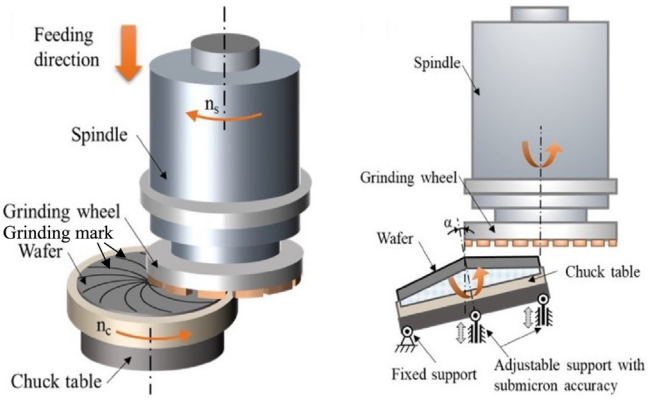

Fig. 16. Wafer grinding using a ring diamond wheel, the tilt angle of the vacuum chuck, α, and the bending of the vacuum chuck and wafer to adjust the wafer total thickness variation [101].

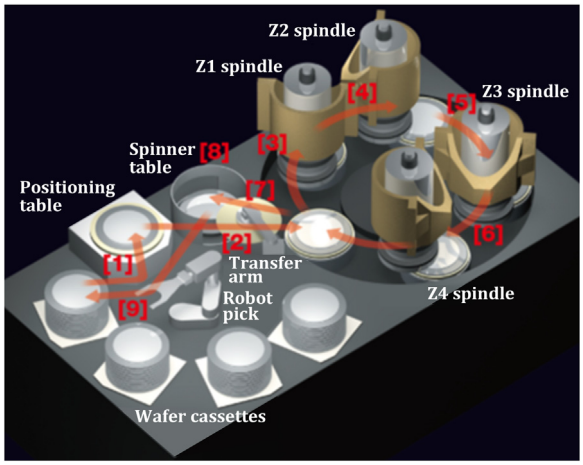

Fig. 17. The configuration of the 9-step wafer grinding machine with four grinding and five chuck spindles (DISCO DFG8830) [44].

from the OHT of AMHS in the fab. For example, the DISCO DFG8830 wafer grinder [44], as shown in Fig. 17, has four wheel spindles, five chuck spindles, and the capacity for four cassettes. An additional chuck is used for loading and unloading by a wafer transfer robot. Nine steps are outlined in Fig. 17 for this machine. The wafer is transferred to the positioning table of the machine using the robot (Step 1), loaded on to the chuck table by a transfer arm (Step 2), and ground (Steps 3–6) with four grinding operations of diamond grit sizes from large to small or the combination of three grinding and one polishing wheel (to reduce the time in subsequent CMP) at the same time. Finally, the wafer is moved to the spinner table (Step 7), cleaned and dried (Step 8), and then returned to the cassette (Step 9).

5. Material removal processes

Abrasive technology has advanced significantly since the 1990 review of abrasive machining of Si [169]. The wire saw slicing of Si wafers [154], the grinding wafers to the precise total thickness variation (TTV) and surface integrity, the dicing of wafer to die, and CMP all require precision abrasive removal processes. The very small DOF in EUV lithography (Table 2) needs very flat die surfaces to be achieved by grinding and CMP to have good yield. Advancements in abrasive grinding and dicing as well as laser slicing and dicing of SiC will be reviewed in this section.

5.1. Wafer grinding processes

Wafer back grinding, as shown in Fig. 16, is a process to thin the Si wafer. For example, the IC chip in a credit card is thinned by back grinding to between 0.05 and 0.1 mm in thickness on a 300 mm diameter Si wafer. The grinding wheel surface speed is commonly in the 40–80 m/s range. The Si wafer thickness needs to be further reduced to below 0.005 mm by back grinding to minimize the parasitic electrical resistance and capacitance of electrical circuits in packaged IC chips. Such wafer thinning may significantly affect the devices' power consumption and response time, as well as the battery life. Grinding a Si wafer down to 5 µm has been demonstrated (Fig. 18) [45].

A key application of back grinding is for backside power delivery (BPD), as illustrated in Fig. 19 [65]. The traditional frontside power delivery (left-hand side) has the power layer on the top of the interconnect layer, which is on the top of the transistor layer (about 1.3 μm in thickness). Buried power rails are then necessary in the transistor layer to deliver an electrical signal to all transistors. In contrast, the BPD design eliminates the buried power rails by providing the power directly to the transistor layer (right-hand side of Fig. 19). This design may improve the voltage drop by 30% and frequency response by 6% [65] due to lower electrical resistance and capacitance, respectively, by wafer thinning via back grinding.

Si wafer back grinding is critical for BPD. As shown in Fig. 20, the transistor and interconnect layers are first fabricated in Wafer 1. A hermetic layer is applied, and Wafer 2 is fusion-bonded to the top.

Fig. 18. A picture of 300 mm wafer with 5 µm thickness [45].

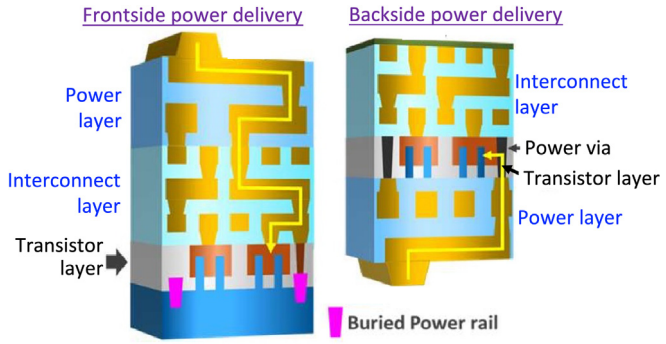

Fig. 19. Backside power delivery (modified from [65]).

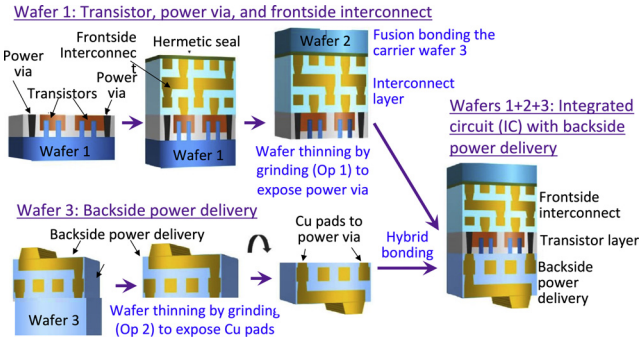

Fig. 20. Backside power delivery manufacturing steps and the need to grind a thin Si wafer (modified from [65]).

The wafer back grinding (marked as Op 1 in Fig. 20) is performed to remove almost all of Wafer 1 and expose the power via in the transistor layer for the subsequent wafer-to-wafer joining. The BPD circuit is fabricated on Wafer 3. That wafer is thinned by back grinding (marked as Op 2 in Fig. 20) to expose Cu pads for joining to power vias in Wafers 1 and 2. Diffusion bonding joins Wafers 1, 2, and 3 to form the IC with BPD, as shown in Fig. 20.

The grinding wheel feeds to the wafer to initiate the grinding. The edge of the ring grinding wheel is at the center of the wafer and creates grinding marks initiated from the center of the wafer (Fig. 21). The saw marks from diamond scribing on the Si wafer are determined by the grinding wheel and wafer rotating speeds. Diamond scribing in different directions of the anisotropic Si crystalline under the same depth of cut will result in varying levels of damage and phase transformation. Fig. 21a shows the top view of a cubic Si (Si-I) wafer with a notch representing the 90° [0 1̲ 0] direction. The 0° [0 1 1], 22.5°, and 45° [0 0 1] are also marked. Fig. 21c–d illustrate the change of Si crystal direction along saw marks on the wafer surface in three sets of grinding wheel and wafer rotational speeds. Using the color spectrum, the 45°, 22.5°, and 0°/90° directions are represented by colors red, green, and blue, respectively.

Since the single-crystal Si is anisotropic, the depth of subsurface damage varies in different crystalline directions under the same diamond cutting condition. This is best illustrated in Fig. 22, which shows single-point diamond cutting of Si with a gradually increasing depth of cut, illustrating the ductile and brittle (damaged) regions [187]. At a small depth of cut, material removal occurs in the ductile region without damage. Beyond a critical depth of cut, microfractures occur in the subsurface. The critical depth of cut varies depending on the cutting direction in the anisotropic Si. The cutting along 0° is the most prone to damage, and 22.5° has the largest critical depth of cut and is hence the least prone to damage. The phase transformation of Si due to the high stress and temperature during diamond scribing also depends on direction [178].

Due to this direction effect, the crystalline phase and damage region on the Si wafer after grinding are not uniform. In the grinding of very thin Si wafers for BPD (Op 1 and Op 2 in Fig. 20) or advanced packaging, a non-uniformly ground Si wafer surface with regions of

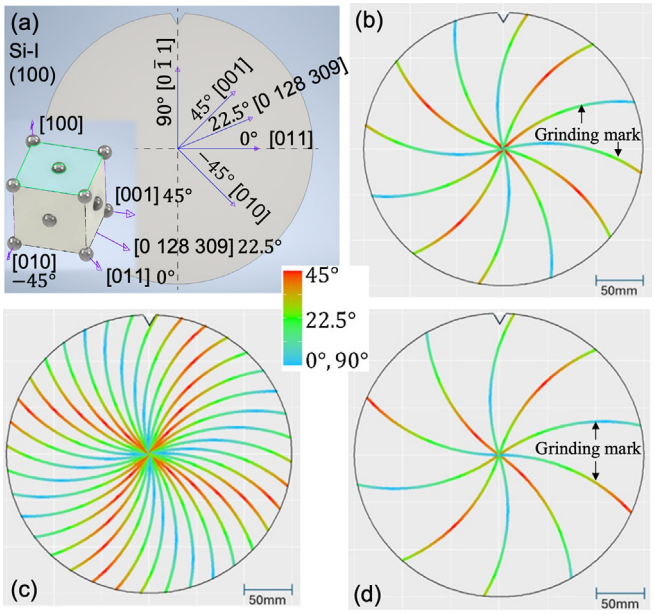

Fig. 21. Si wafer (a) crystalline directions and the saw marks and the Si crystalline directions for grinding wheel/wafer speed (rpm/rpm) of (b) 3600/300, (c) 3600/100, and (d) 1000/100.

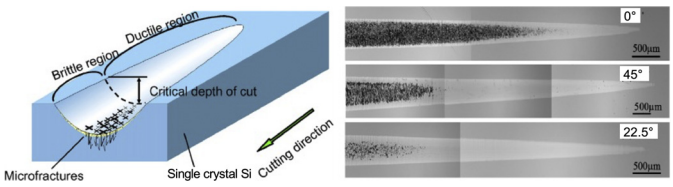

Fig. 22. Single-point diamond cutting of the single-crystal Si with a gradually increased depth of cut and the level of damage in 0° (largest), 22.5° (smallest), and 45° [187].

different phases and damages may occur. Such a crystalline direction effect on diamond scribing has been studied by molecular dynamics (MD) simulations for Si [165,207] and SiC [183] and by experimental investigations for Si [165,208]. Fig. 23 shows the MD modeling of the surface generation based on an 8 nm diamond engaging the Si workpiece 35 nm x 20 nm x 15 nm at 100 m/s in three Si crystal orientations using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) to find the atomic motion [207]. The Si crystal direction was shown to influence the surface shape, contact force, temperature, and stress.

Results of MD simulations using LAMMPS have identified the permanently deformed atoms (marked by light blue) in diamond scribing of the monocrystalline 6H-SiC in six crystal orientations, with two examples shown in Fig. 24 [183]. The crystal direction influences the deformed shape and subsurface damage of the SiC.

In finishing and fine grinding of Si or SiC wafers, the diamond grit size is small (ANSI mesh 30,000) to achieve ductile region grinding. The regional phase transformation on the Si surface after precision diamond grinding can be detected due to the volume change caused by phase transformation. This phenomenon has been studied in

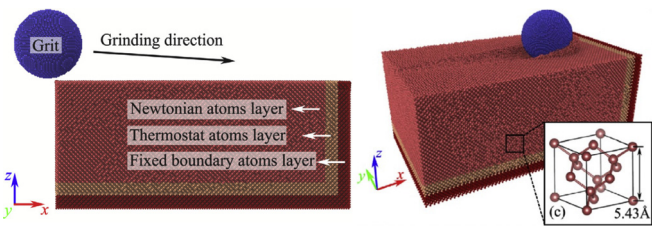

Fig. 23. MD model of the diamond scribed Si surface topology [207].

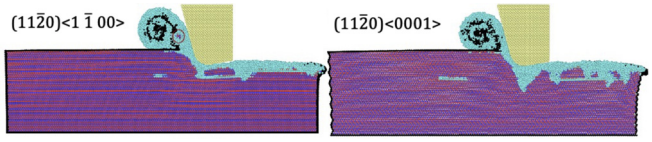

Fig. 24. Deformation of 6H–SiC in diamond scratching of two scratching planes and directions; magenta color: original atom, light blue color: permanently displaced atom, and black color: surface atom [183].

which an amorphous Si layer is formed at the topmost portion and followed by the pristine crystalline lattice of the Si-I phase beneath the bottom of the groove scratched by a diamond indenter [203]. The experimental study and modeling of the phase transformation from Si-I to metallic Si (β-Si) as well as the initiation and propagation of cracks underneath the diamond scribed single-crystal Si have also been conducted [178].

There are many other advancements in the modeling of Si wafer grinding. For example, an analytical model has been developed to predict subsurface damage [192]. Effects of grinding wheel speed and workpiece temperature in Si grinding were also investigated [204]. A high-shear and low-pressure material removal rate model was developed for Si wafer grinding [167].

The cost-effective grinding of SiC wafers to achieve good surface integrity and dimensional accuracy for reducing the time and cost required for subsequent CMP, is an ongoing competition among key SiC manufacturers. A review paper on single-crystal SiC machining [179] summarized recent advancements in grinding. The method to slice the SiC wafer, either by wire saw or laser [16] (see Fig. 25 by DISCO), greatly affects the grinding process. The laser creates a void near the focal point, and the large expanding region above the void, as shown in Fig. 26 [105]. For some SiC wafers, such as the N-type SiC, laser slicing creates 4°-angled cleave cracks and an uneven zig-zag surface. Rapidly scanning the surface to create a horizontal crack and minimize kerf loss is an active research topic. Plasma or laser can be applied to oxidize high spots on the surface of the SiC wafer after laser or wire saw slicing for more efficient grinding of SiC. In SiC wafer production, the CMP time is long and critical to the cost. Therefore, the slicing process needs to be tailored to minimize the damage layer thickness and reduce the CMP time.

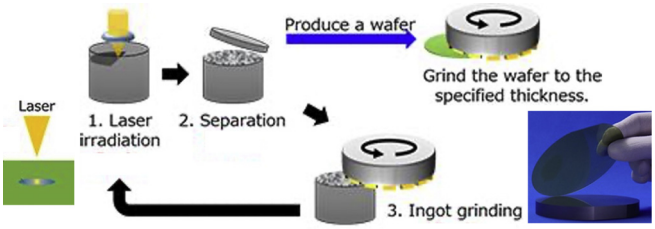

Fig. 25. Laser slicing and grinding of both the wafer and boule [16].

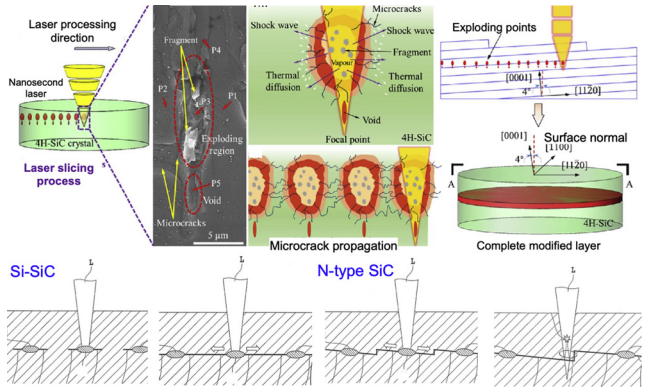

Fig. 26. Laser slicing of SiC wafer, the void and exploding region by laser in SiC [105], and the horizontal cracks in two types of SiC.

5.2. Dicing processes

Dicing is a critical process for SiC wafers. Three dicing methods, namely, blade dicing, laser (stealth) dicing, and laser ablation dicing, are all utilized in semiconductor manufacturing. Blade dicing is a mechanical removal process using a thin diamond disk. The dicing wheel width can be as thin as 0.01 mm for a nickel bond hubless wheel or 0.02 mm for the nickel or resin bond disk wheel in a hub. An air spindle is commonly utilized to minimize wafer damage. Ultrasonic-assisted blade dicing has also been studied for hard and brittle materials such as SiC [152]. In contrast to laser ablation dicing that uses laser energy to sublimate and evaporate wafer material, laser (stealth) dicing forms cracks inside a wafer using a transmissible laser beam, which applies thermal stress to the wafer. The wafer is broken up by tensile stress caused by rapid cooling [89]. This dicing method is free from fluids, debris, and contaminants [111].

Dicing before grinding (DBG) involves partially cutting the wafer followed by backside grinding to partition the wafers into dies. Stealth dicing before grinding (SDBG) replaces the dicing saw with a laser that causes internal cracks along the target dicing lines. SDBG (see Fig. 27) is a dry, non-contact process that reduces mechanical stresses and edge defects while increasing geometric accuracy and throughput [66].

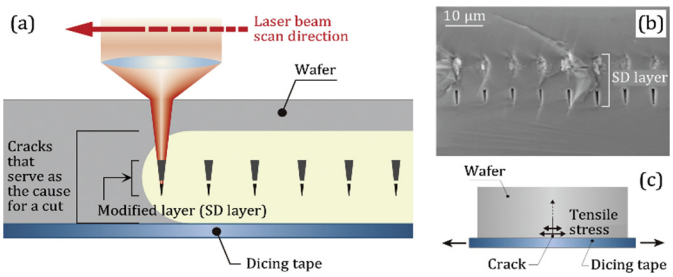

Fig. 27. (a) Schematic of the stealth dicing (SD) process, (b) photograph of the SD layer inside Si wafer, and (c) principle of crack propagation via extension of the dicing tape [66].

Blade and laser dicing both cause cracks or dislocations in SiC crystals, such as basal plane dislocations (BPDs), as shown in the X-ray topography (XRT) images in Fig. 28a, because of the mechanical and/or thermal stress applied on a wafer. These damages may result in the formation of stacking faults inside the dies (devices). In SiC power devices, these faults should be avoided to ensure device performance and reliability.

A damage-free dicing method using a waterjet-guided laser (WGL) [79], which applies a pulsed laser and a high-pressure water jet, as shown in Fig. 29, has been developed for SiC wafers. A water jet with a diameter of less than 0.1 mm is formed with a pressure greater than 10 MPa. The laser is coupled with the water jet through an objective lens and an optical window. The laser light propagates through the water jet as a waveguide with total internal reflection. The light energy reaches the end of the water jet, keeping the same diameter. WGL dicing differs

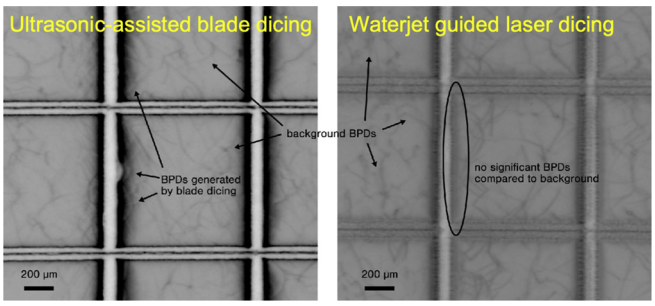

Fig. 28. XRT images of SiC wafer diced by an ultrasonic-assisted blade (with basal plane dislocation) and waterjet-guided laser dicing [79].

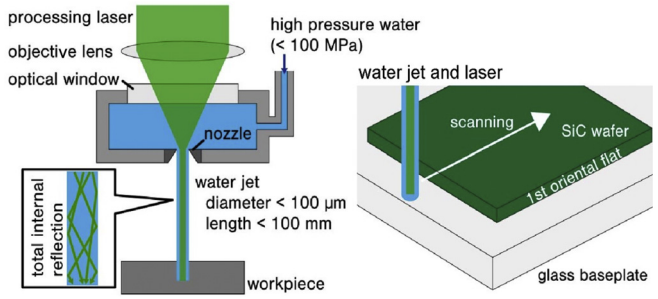

Fig. 29. The concept of waterjet guided laser dicing of SiC wafer [79].

from conventional laser dicing, which utilizes a focused beam. Instead of thermally-induced cracking, the workpiece is ablated by the thermal energy delivered by the laser light and water jet, and simultaneously cooled by water. Debris is effectively ejected by the water jet. There could be almost no generation of cracks or dislocations in crystals such as BPDs, since no mechanical stress is applied to the wafer. Additionally, the processed part is not tapered, but instead, is perpendicular to the wafer surface, as the water jet can behave as a straight waveguide.

XRT of WGL-dicing and blade-dicing processed wafers are shown in Fig. 28b. Some background BPDs were seen throughout the specimens as gray line contrasts. There were many arc-shaped BPD contrasts around the blade dicing grooves. In contrast, although diffracted contrasts appeared around the holes depending on the processing conditions, no contrasts due to BPDs were observed for the WGL-dicing processed wafer.

6. In-line metrology for process control

Measurement is essential in IC manufacturing, which is practiced across the entire value chain composed of wafer manufacturers, IC fab, and packaging. Fabrication, packaging, and measurement are the three major portions of IC manufacturing, where measurement is as important as the other two. Each of the three portions contributes about one-third of the total cost [113]. Measurement is not only necessary for verifying the quality and the functionality of the circuits after the FEOL and BEOL through electrical testing, but also important for control and management of IC manufacturing processes through integrated metrology [7]. Compared with electrical testing, integrated metrology involves many diverse measurands and measurement technologies, which are more challenging. It takes months to go through over 1000 steps in the fab.

In-fab (in situ) measurements can be categorized as in-line, on-machine, and in-process measurements, as shown in Fig. 30 [56]. In-line measurement, the focus of this section, is effective for making sure the requirements for IC products are satisfied in each step. It can also identify product defects and failures in a timely manner so that wasteful subsequent steps can be avoided and costs can be reduced. Some defects can even be repaired based on in-line measurement, such as scanning electron microscopy (SEM) and the aerial image based at-wavelength, or actinic, metrology system for EUV mask repair with gas-assisted e-beam lithography methods [70]. Results of in-line measurements can also be used for diagnosing the process and equipment [56], allowing for the optimal status of the process and equipment to be maintained.

Meanwhile, the recent trends of scaling down, stacking up, and 3D heterogeneous integration in IC devices have brought significant challenges to semiconductor metrology [76,136]. Among the quantities measured in IC manufacturing, a few key dimensional quantities

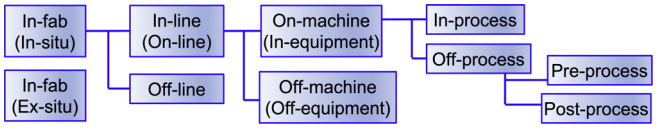

Fig. 30. Measurement conditions in IC fab, modified from [56].

can be focused on as measurands from the perspective of process control and management. Thickness is the most fundamental and common measurand [57]. For example, the thickness of Si wafers is measured during the grinding and polishing processes. In IC fab, thickness is measured against various layers, such as Si substrates, insulator layers, and metal layers, during back-grinding, deposition, and etching processes. Thickness measurements of solder layers and mold resins are performed in the packaging fab during bonding and molding processes. Other measurands include CD and overlay in the lithography process, as well as the overlay and internal defects in bonding and molding processes.

For in-line measurement of such a measurand, it is necessary to select a proper measurement method based on the material, structure, required accuracy, throughput, and compatibility with the manufacturing process [58,59]. Optical methods can make accurate measurements of thickness and CD at a fast speed. CD can also be measured by mechanical methods such as atomic force microscope (AFM), electrical methods of SEM/transmission electron microscopes (TEM), and X-ray based systems such as small-angle x-ray scattering (SAXS). X-ray microscope (XRM) and scanning acoustic microscope (SAM) are effective for tomographic measurement of internal defects. In addition, X-ray reflectometry (XRR) is a method for thickness measurement. In this section, in-line measurements of the measurands shown in Table 3 are presented.

Table 3
In-line measurement methods for process control.

Process	Measurand		
	Thickness	Critical dimension	Internal defect
Grinding	Reflectometry	N/A	N/A
CMP	Reflectometry Eddy current	N/A	N/A
Deposition	Ellipsometry XRR	N/A	N/A
Patterning	N/A	Optical CD (OCD), CD-SEM/TEM, AFM, SAXS	N/A
Packaging	N/A		XRM, SAM

6.1. Wafer thickness measurement for grinding processes

A broadband near-IR light source, such as a super luminescent diode (SLD), is typically employed in a spectral reflectometer, also called a spectral interferometer, for Si substrate thickness measurement. In Fig. 31, a light beam from the light source is projected onto the Si substrate. The beam is reflected at the surface and from the bottom surface of the Si substrate. The two reflected beams S_1 and S_2 are then interfered with each other to produce an interferogram. The wavelength spectrum $I(\lambda)$ of the interferogram is obtained by a spectrometer, from which an output $i(z)$ is evaluated from the inverse Fourier transform of $I(\lambda)$ [60]. The $i(z)$ has two pulses corresponding

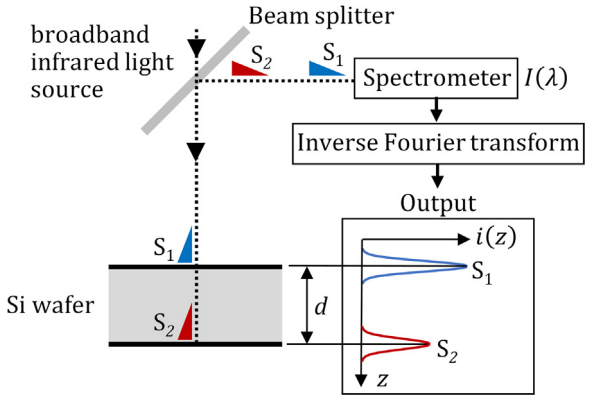

Fig. 31. Thickness measurement using spectral reflectometry for a Si wafer.

to the interference of signals S_1 and S_2. The thickness d of the Si wafer can be obtained from the peak distance between the two pulses. The minimum detectable thickness is determined by the width of the pulses $i(z)$, which is inversely proportional to the spectral bandwidth of the light source [102,103].

As an early work on on-machine thickness measurement of Si wafers, Onuki et al. [135] built a spectral interferometer by combining a commercial white-light source and a commercial spectrometer (Fig. 32). It was verified that the spectrometer was a key element for determining the thickness measurement range and resolution. By selecting a spectrometer with a sensitive range of 900 nm to 2600 nm and a spectral resolution of 8 cm^{-1} (2 nm at a wavelength of 1700 nm), the thickness measurement range and resolution were confirmed to be 88 μm and 0.2 μm, respectively.

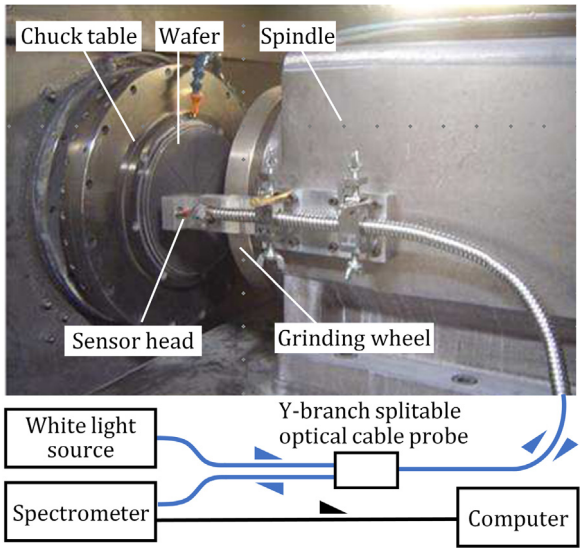

Fig. 32. Setup of a prototype spectral reflectometer on a grinding machine for thickness measurement of a Si wafer, modified from [135].

Since then, significant progress has been achieved in spectral reflectometry for measuring Si wafer thickness. Spectral reflectometers with excellent performance are commercially available for different applications. Table 4 shows two sensors for thick and thin wafers, both designed to have a long working distance suitable for use on grinding, lapping, and polishing machines for Si wafer thickness measurement. The wafer thickness variations can be measured by scanning the sensor over the wafers. The spectral reflectometer in Fig. 31 can be expanded from point to line measurement by using a line-shaped light beam and a line spectrometer. Since the thickness data along a line across the wafer can be simultaneously obtained by the line spectral reflectometer, the thickness of the entire wafer can be obtained by scanning the interferometer along a single axis, as shown in Fig. 33a, to improve the measurement speed. Fig. 33b shows the Si wafer measurement result using a commercial line spectral reflectometer system

Table 4
Specifications of commercial spectral reflectometry for thickness measurement of Si wafer and coatings.

	Long range type *	Short range type **
Thickness range	50 μm–1050 μm	4.3 nm–43 μm
Thickness resolution	1 nm	1 nm
Measuring rate	100 Hz–6 kHz	200 Hz
Accuracy	±100 nm	±0.4%
Working distance	24 mm ± 3 mm	10 mm
Multi-layers	Up to 5 layers	Up to 10 layers
Light source and wavelength	SLD λ_c 1.1 μm	Halogen lamp 0.4 μm–1.1 μm
Light spot diameter	20 μm	10 mm

* Micro-Epsilon interferoMETER IMS5420.
** Hamamatsu Photonics Optical NanoGauge C13027-12.

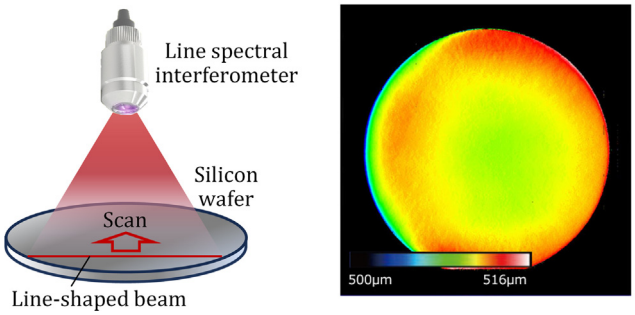

Fig. 33. A line spectral reflectometer for in-line inspection of Si wafer thickness (a) schematic of the line scan mode and (b) a measurement result of thickness variation of a Si wafer, modified from [77].

[77]. The system can make thickness measurements with a resolution of 1 nm, and a spatial resolution of 100 μm for a 100 mm diameter wafer and 300 μm for a 300 mm diameter wafer. The measurement time is less than 2 minutes. The thickness measurement range is from 20 μm to 800 μm. The measurement error of spectrometer and the uncertainty in the refractive index of Si are the major sources of measurement uncertainty for spectral interferometry.

6.2. Endpoint detection for CMP processes

Endpoint detection is a common practice in CMP for in-process measurement of the thickness of metal or dielectric layers [118]. Fig. 34 shows a setup of endpoint detection with a thickness sensor embedded in the turntable of the CMP machine where the polishing pad is mounted [180]. This sensor is rotated by the turntable to scan across the center of the wafer mounted on the carrier ring over an arc path on the wafer surface. For a metal (e.g., Cu or W) layer on a Si substrate, an eddy current thickness sensor is used as shown in Fig. 35 [180]. The eddy current sensor is based on the principle of electromagnetic induction. The magnetic field generated by the eddy current flowing in the metal layer is proportional to the metal layer thickness and can be measured by a detection coil. In Fig. 35, the top-ring is composed of four donut-shaped airbags aligned along the radial directions for controlling the polishing pressures in four areas on the wafer surface. The pressure in each area is controlled by the corresponding airbag based on the eddy current thickness sensor measurement.

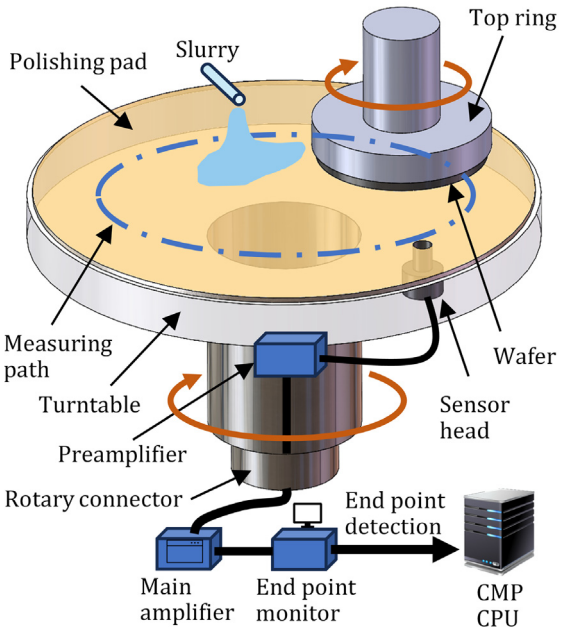

Fig. 34. Endpoint detection in CMP based on in-process thickness measurement, modified from [180].

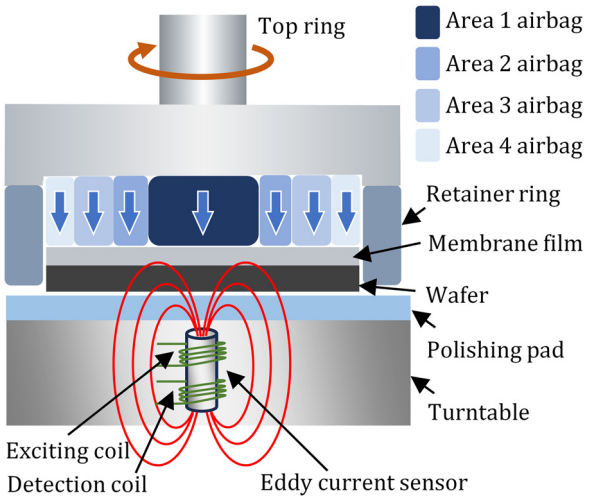

Fig. 35. Eddy current sensor for CMP endpoint detection [180].

The variations in the electromagnetic properties and temperatures of the substrate and the layer materials as well as the slurry would be the major sources of measurement uncertainty for the eddy current measurement.

A spectral reflectometer can be used for endpoint detection in CMP of dielectric layers, such as SiO_2. Fig. 36 shows an example of the light from a halogen lamp (400 nm to 800 nm) being guided by an optical fiber (input fiber) to the center of the wafer through a small aperture opened on the polishing pad to measure the dielectric layer deposited on the wafer. The aperture had an elliptical shape with a longer and shorter diameter of 6 mm and 3 mm, respectively. The size of the aperture was small enough to avoid any influence on the CMP process. The light was reflected at the dielectric layer when the aperture moved across the center of the wafer. The reflected light was collected and guided by another optical fiber (output fiber) to a spectrometer for data processing. Two optical fibers were in a water supply pipe where pure water flowed onto the polishing pad through the aperture. A film of pure water was formed between the ends of two fibers and the wafer dielectric layer when the aperture was covered by the wafer. The water functions as a stable transparent medium to guide the reflected beam from the wafer to the output fiber. The water flow rate was controlled at a very low level when the aperture was not covered by the wafer, so that the influence of the

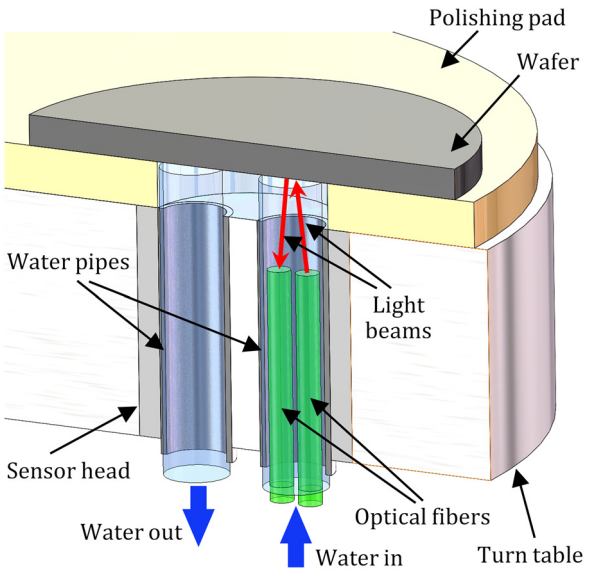

Fig. 36. Two pure water pipes for guiding the light between a spectral interferometer and the water surface in the CMP machine [180].

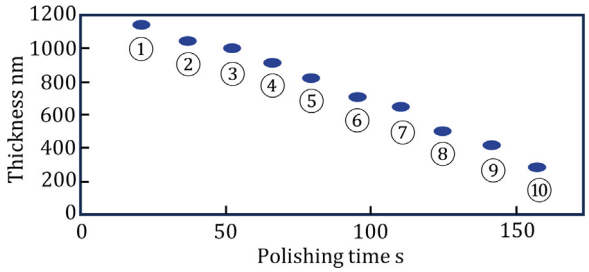

Fig. 37. The result of CMP in-process thickness measurement of dielectric layer by spectral reflectometry [131].

water on the slurry could be minimized. Fig. 37 shows the thickness measurement results of a borophosphosilicate glass layer in an interlayer dielectric during a CMP process, where the output of the reflectometer at a specific light wavelength changed periodically with the change of the layer thickness. The circled number in the figure shows the sampling number of the polishing time at every half period of the reflectometer output.

The data-driven techniques for CMP endpoint detection have also been developed. For example, a sequential Bayesian approach has been used to correlate wireless vibration signal features with material removal and thereby determine CMP endpoints [145].

6.3. Thin film measurement for deposition process

Ellipsometry is another popular method for dielectric and metal thin film thickness measurement. Fig. 38 shows a schematic of spectroscopic ellipsometry [161]. A linearly polarized light is projected onto a single-layer or a multilayer thin film coated on a substrate. The reflected beam typically has an elliptical polarization comprising a *p*-polarization component in the plane of incidence and an *s*-polarization component vertical to the plane of incidence. The ratio ρ of r_p and r_r, which are the complex amplitude reflection coefficients in two directions, respectively, is referred to as the ellipsometric function. The amplitude Ψ and phase Δ of ρ are determined by the thickness of the thin film and the optical constants of the thin film and the substrate. In a spectroscopic ellipsometer, a wideband UV and visible light source is employed. The spectrum of the reflected beam after passing through a rotating analyzer is measured. A model-based analysis algorithm is then applied to the spectral data to obtain the thickness of the thin film.

Spectral reflectometry and spectroscopic ellipsometry for thin film measurement are fundamentally the same method, being based on multi-beam interference among the reflected beams at the surfaces of the thin film and substrate. Reflectometry only utilizes the intensity of the interference signal, i.e., the reflectance of the

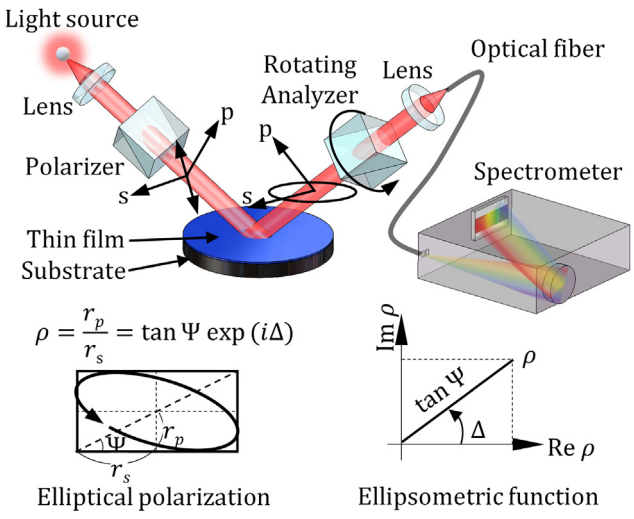

Fig. 38. A schematic of spectroscopic ellipsometry, modified from [161].

multiple-beam interference, in either the *p*- or *s*-polarization direction, which is basically a simple cosine function of the thin film thickness. Ellipsometry uses not only the reflectance but also the phase difference between the electrical fields of the *p*- and *s*-polarization components, which greatly improves the measurement resolution of thickness. Figs. 39a and b show the outputs of spectral reflectometry and spectroscopic ellipsometry for the thickness measurement of a SiO_2 layer on a Si substrate, respectively. In Fig. 39a, the nanometric thickness of the SiO_2 layer can be identified from the phase output of spectral ellipsometry, which shows the phase difference between the *p*- and *s*-polarization components of the reflected beam. It is difficult for spectral reflectometry to evaluate the thickness of the SiO_2 layer when the thickness is below 10 nm, as shown in Fig. 39b. In spectroscopic ellipsometry, it is necessary to set the incident angle to be close to the principal angle of the substrate, which Δ is 90°, to get a high resolution in thin film thickness measurement. In Fig. 39, the incident angle was set to be 75°, which is close to the principal angle of 75.6° for the Si substrate at a wavelength of 632.8 nm. For this reason, it is necessary to adopt an oblique incident design, which results in a bulky ellipsometer. This is a major shortcoming of ellipsometry compared with spectral reflectometry, especially for on-machine and in-process measurement. Table 5 shows the comparison between a commercial ellipsometer (Model FS-8 of Film Sense) and a commercial reflectometer (Model C15151-01 of Hamamatsu).

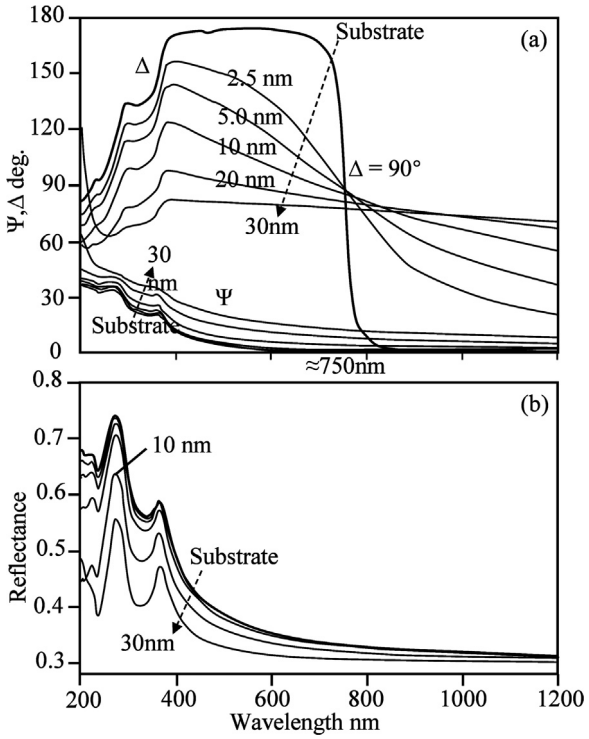

Fig. 39. Thickness sensitivity in measurement of a SiO_2 layer on a Si substrate (a) spectral reflectometry and (b) spectroscopic ellipsometry [161].

Table 5
Ellipsometry and reflectometry for thin film thickness measurement.

	Ellipsometer*	Reflectometer**
Thickness range	0.1 nm–5 μm	4.3 nm–43 μm
Thickness resolution	0.01 nm	1 nm
Measuring rate	590 Hz	200 Hz
Accuracy	0.18 nm for 90 nm SiO_2 on Si	±0.4%
Incident angle	60°–75°	0°
Working distance/ Beam length	Beam length 150 mm to 1 m	Working distance 10 mm
Multi-layers	Up to 10 layers	Up to 10 layers
Light source	8 LEDs in 0.37 μm–0.95 μm range	Halogen lamp 0.4 μm–1.1 μm
Light spot diameter	6 mm–20 mm	1 mm

* Film Sense, Multi-Wavelength Ellipsometer FS-8.
** Hamamatsu Photonics Optical NanoGauge C15151-01.

The penetration capability of X-rays is being utilized to measure not only dielectric thin films but also metal thin films. The thickness measurement of ellipsometry relies on accurate values of optical constants, which can vary in ultrathin films, resulting in inaccurate thickness measurements [146]. In contrast, the optical constants of an ultrathin film for X-rays are almost consistent with bulk materials. This makes X-ray reflectometry (XRR) a suited tool for thickness measurement of films thinner than 10 nm. XRR is also an absolute measurement method that does not need calibration. This is an advantage for in-line metrology. On the other hand, both are model-based methods where modelling errors would be the major source of thickness measurement uncertainty.

6.4. CD measurement for patterning processes

Lithography-based patterning is carried out by repeating several processes, including photoresist coating, exposure, development, and etching. In situ metrology of the CD of a patterned feature in patterning processes can be made by a variety of measuring methods based on optical (scatterometry),·electrical (SEM/TEM), and mechanical (AFM) methods. Fig. 40 shows a simple comparison of these methods in terms of throughput and measurable parameters [4]. CD-SEM is a top-view 2D imaging technology, which has long been a standard tool for measuring linewidth, hole diameter, and overlay between layers [119]. It offers intuitive images of individual structures with a high throughput. CD-SEM is a probe-scan method where a focused electron spot scans across the sample surface. The sub-nanometric diameter of a focused electron spot makes the lateral resolution of a CD-SEM as high as 1 nm. A model-based CD-SEM also provides 3D shape information of a patterned feature. From the process control point of view, the automatic operation of CD-SEM is essential for in-line measurement and inspection following a predetermined measurement sequence. It is necessary to automatically move the field of view of the CD-SEM to the evaluation point based on pattern recognition for image matching of a measured SEM image with a designed pattern. The decrease in feature size and the increase in pattern complexity have brought challenges for conventional 2D pattern recognition. Fig. 41 shows a new in-line CD with two detectors for 3D pattern recognition [189]. This technique can automatically measure all CDs, while the success rate of the conventional method varied from 65% to 100%.

TEM is currently used for off-line inspection rather than as a routine metrology tool for in-line CD measurement due to its destructive nature and time-consuming sample preparation. However, the sub-angstrom spatial resolution and 3D measurement capability of TEM make it a tool for CD measurement of future 3D IC devices with physical feature sizes in the single nm range. Automation of TEM measurement is essential for this purpose [159,163]. Fig. 42 shows a

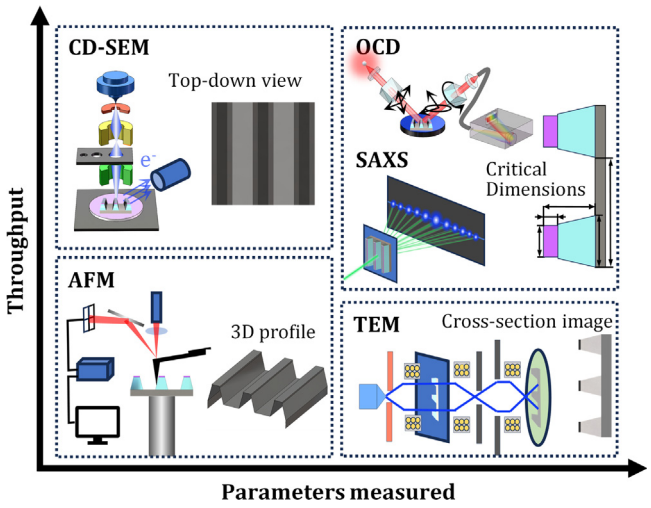

Fig. 40. A comparison of the CD measurement technologies.

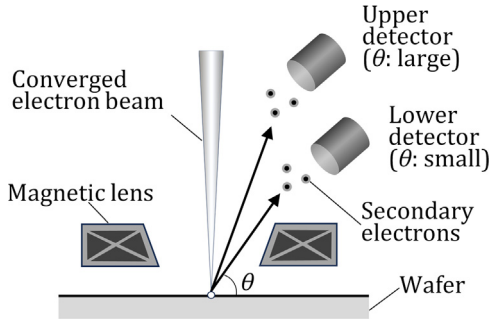

Fig. 41. An in-line CD-SEM with two detectors [189].

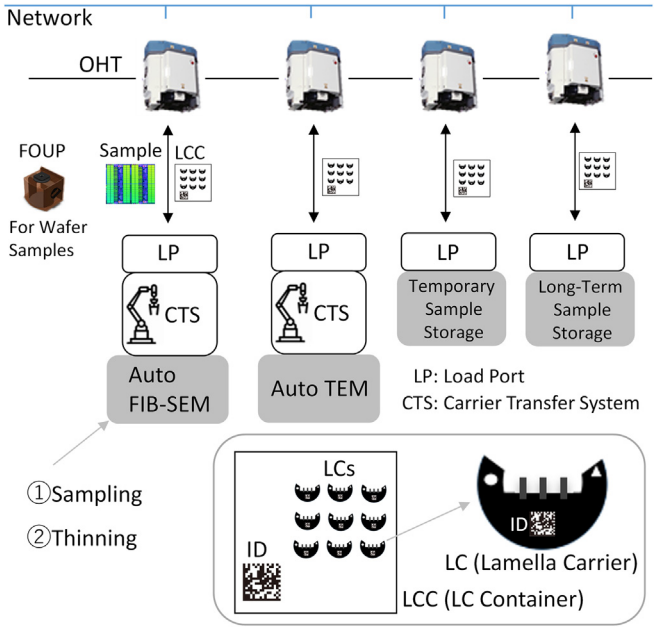

Fig. 42. Automatic in-line TEM [Courtesy of Tsuyoshi Onishi of Hitachi High-Tech Corp].

proposed automatic in-line TEM [134]. Automation of the sample preparation using FIB is one of the key tasks for in-line TEM.

CD-AFM is also a probe-scan imaging system where a Si probe tip with a tip radius of several nm, which determines the lateral resolution of imaging, is employed (Fig. 43). An XY scanning stage and a Z-servo stage are employed in an AFM for the scanning, which determines the measurement speed of imaging as well as the 3D imaging accuracy. Automatic CD-AFM has been developed for in-line CD measurement [75]. An on-axis optical image recognition system is embedded in the CD-AFM for rapid and accurate global alignment. A thermo-actuated probe cantilever with a drive frequency of 600 kHz is employed to achieve a data collection rate of up to 260 kHz, resulting in a throughput of up to 170 wafers per hour. The AFM probe is oscillated with a sawtooth profile to increase the response of the probe so that the interaction force between the probe and sidewalls can be reduced. As a result, the probe life is good for 10,000 images, which is 10 times higher than conventional AFMs. The edge/top/bottom roughness can also be detected. Similar system can be found in [11]. For AFM measurements, the edge radius of the probe tip is one of the significant uncertainty factors, which can be self-calibrated by using an edge-reversal method [28,54].

Optical scatterometry utilizes the light scattered or diffracted from a periodic grating structure, which follows the grating equation [29]. A model-based analysis of the scattered light pattern provides averaged measurements of the grating geometries, including

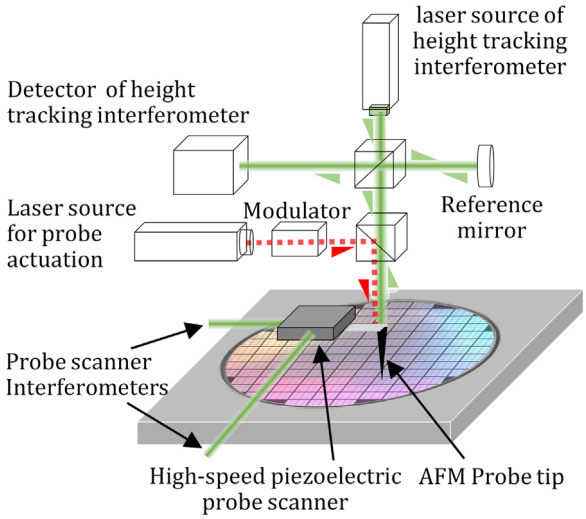

Fig. 43. A high-speed automatic AFM for in-line measurement [75].

line and space, trench depth, and side angle, over the grating area illuminated by the incident light, typically on the order of 10 μm to 50 μm. Optical scatterometry for CD measurement is often referred to as OCD. It has the advantage of high throughput and has good compatibility for automatic in-line metrology [129]. Fig. 44 shows the results of CDs of the top layer of tall stacks measured by vertical travelling scatterometry (VTS). The irrelevant spectral information from buried layers was filtered out for the data processing of scatterometry. Four wafers exposed with different doses were employed for the measurement. Three were exposed with fixed dose values (Doses 1–3) and one with variable dose across the wafer (Dose stripe). In addition to optical scatterometry, SAXS is also available for in-line CD measurement. Fig. 45 shows the results of a 3D profile of a high-aspect-ratio memory hole structure near the edge of the wafer measured by SAXS.

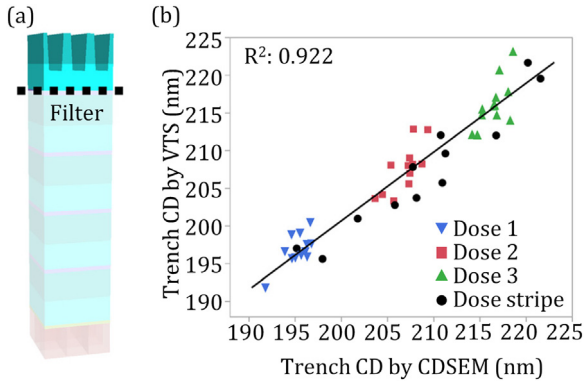

Fig. 44. Measurement results of trench CD by a VTS in comparison with CD-SEM [148].

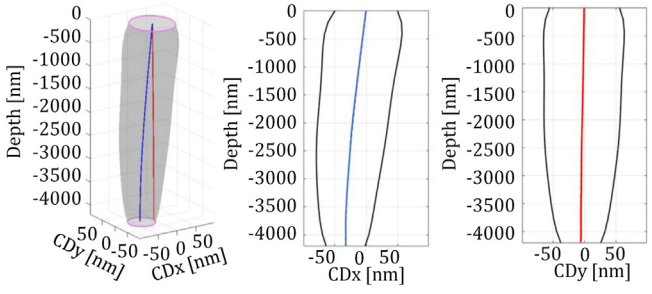

Fig. 45. Measurement results of small angle X-ray scatterometry [61].

The calibration errors in CD-SEM/TEM, the probe tracking errors and probe wear in AFM, the modeling errors in OCD and SAXS are the major uncertainty sources in each of the methods.

6.5. Interconnect inspection for packaging processes

X-ray computed tomography (XCT) is a powerful imaging tool for high-resolution and non-destructive inspection of interconnects such as micro-bumps and TSVs in 3D packaging [114]. In dimensional CT metrology, the transmitted X-rays from a sample at different angular positions are projected onto an X-ray detector. The output of each pixel on the detector is a result of attenuation of the input X-ray due to the absorption or scattering of materials that the X-ray passes through. A mathematical reconstruction of the projected images then provides a 3D image of the internal structure of the sample because the amount of attenuation is a function of the length and the properties of the material along the X-ray path. Spatial resolution is a key issue when XCT is applied for interconnect inspection [33]. An XCT with a micrometric resolution is often called X-ray microscope (XRM). A recently released in-line XRM for quality control of the solder mounting process of C4 Bump and μBump in 3D packaging has a high spatial resolution of 0.2 μm and a fast inspection time of 30 s [133].

The geometric magnification of a CT system has a trade-off relationship with the working distance. To ensure a sufficiently large working distance, which is crucial for in-line inspection, the resolution of the CT system will be reduced. To improve the resolution while maintaining the working distance, an optical magnifying system is incorporated within the detector assembly to further magnify the X-ray image [69]. It can reach spatial resolutions of 0.70 μm and 0.75 μm at working distances of 50 mm and 100 mm, respectively, with a minimum voxel size of 40 nm for Zeiss VersaXRM 730. Fig. 46 shows an XRM image of a 2.5D interposer package, which is the intermediate layer used to connect multiple chiplets or dies side-by-side [194].

Scanning acoustic microscopy (SAM) is another regular tool used for in-line inspection of interconnects in bonded wafers, chip-on-wafer, stacked wafers, and over-molded wafers. Differing from XCT which operates in transmission mode, SAM works in reflection mode, as shown in Fig. 47 [78]. Acoustic echo components are reflected to the detector from the material or structural interfaces inside the package. The defects in the interconnects would generate irregular changes in the echo components, from which the defects can be detected non-destructively. The working frequency is a key parameter in SAM. It determines the lateral resolution of imaging and the penetration depth of the acoustic signal in the package. The frequency bandwidth of an in-line SAM is typically from 100 MHz to 500 MHz for a high penetration depth in packaging inspection, where a lateral resolution of 5 μm is achievable for a frequency of 500 MHz [126].

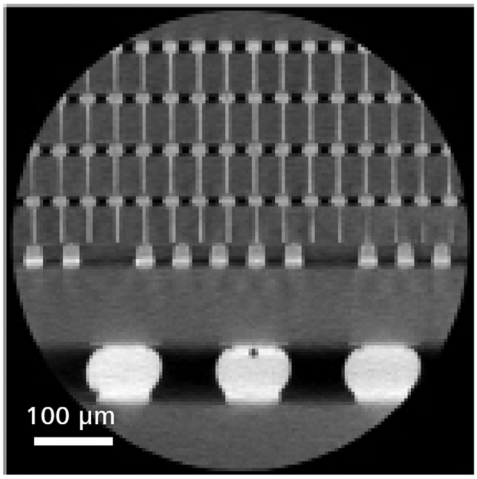

Fig. 46. An XRM image of a 2.5D interposer package. [194].

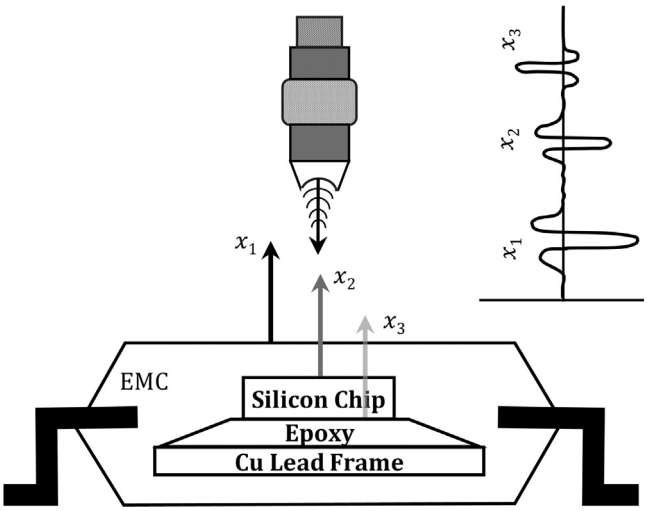

Fig. 47. Operation of scanning acoustic microscopy, modified from [78].

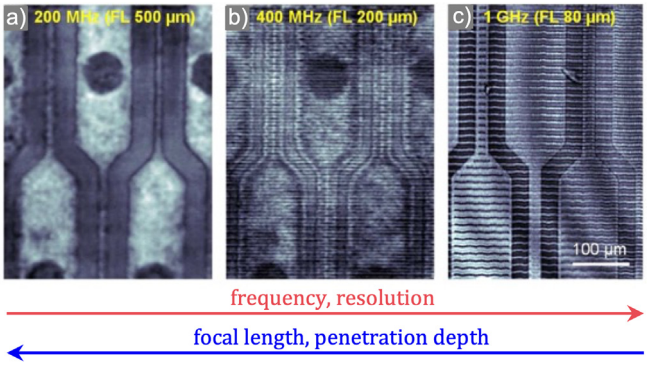

Fig. 48. Back-side SAM images of a bottle-shaped power metal finger structure with ball bonds formed from 30 μm bond wires with frequencies of 200 MHz, 400 MHz, and 1 GHz and corresponding focal lengths of 500 μm, 200 μm, and 80 μm, respectively [174].

In Fig. 48, the lateral resolution of SAM can be improved by increasing the signal frequency to GHz [174]. A lateral resolution of 1 μm could be achieved with a frequency of 2 GHz [114]. The lateral resolution of SAM has a trade-off relationship with the penetration depth. An increase in signal frequency will improve the lateral resolution but decrease the penetration depth due to acoustic attenuation rising exponentially with the frequency. An acoustic interferometry approach has been conducted using a frequency of 100 MHz that can realize not only sufficient penetration depth but also high resolution for fast in-line inspection of TSVs [140] (Fig. 49).

Calibration errors and uncertainties of material properties are the major uncertainty sources for both XRM and SAM.

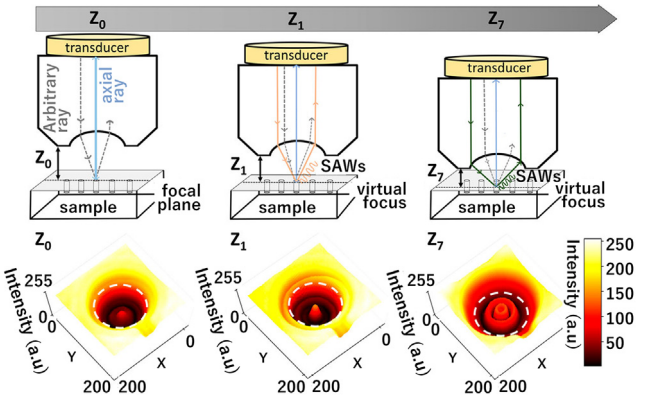

Fig. 49. Scanning acoustic microscope interferometry for TSV inspection [140].

7. Data analytics

The research in data analytics for semiconductor manufacturing aims to advance the transformation of data from fabs into useful knowledge on processes and equipment, and timely decisions in fab operations and control. Modern fabs generate a significant amount of data in production. Sources of those data include sensor readings, the status of each equipment in production, in-process metrology, positions and moves of the work in progress (WIP), production and maintenance schedules, process parameters, etc. Fig. 50 illustrates per-minute data rates from a 300 mm wafer fab in 2019. For example, current etch equipment generates data from several hundred sensors, typically with sample rates from 5 to 10 Hz, whereas 1 Hz sampling rates were standard just a few years ago. In 2022, this led to 2–3 TB of data per day just from etch operations in a mid-size fab. The density and wealth of this data will grow with higher sampling rates and more sensors, all providing better monitoring and control of processes to improve yield. Modern fab operations generate several PB of data each day [117]. These data rates are already staggering and doubled every 2–3 years. Traditional operation research (OR), estimation and control theory, and first principles-based physics and analytical traceability methods do not work well with such a high daily data rate and need advanced data analytics to optimize semiconductor fab operations.

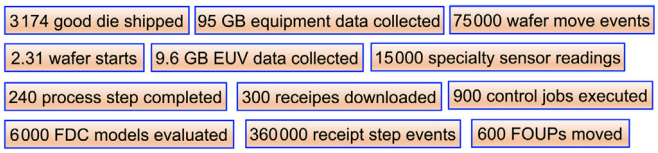

Fig. 50. Data sources and per-minute data rates observed in a 300 mm fab in 2019 [117], FDC: fault detection and classification, FOUP: front-opening unified pod.

Fig. 51 summarizes the functions of data analytics in fabs. Equipment and process-related data are utilized for the detection of anomalous behaviors (fault detection), identification of the root causes of such behaviors (fault diagnosis), and prediction of the remaining life of equipment and processes (fault prognostics). Virtual metrology (VM) employs equipment and process data to estimate the outgoing quality characteristics in the relevant process. VM reduces the need for time-consuming, expensive, and potentially destructive physical metrology and inspection. Advanced process control (APC), also known as the automatic process control, fuses physical metrology and VM results with equipment- and process-related data and models to close the process control loop via adaptations of process parameters. APC aims at maintaining the desired quality characteristics with minimal variations. Finally, equipment reliability information provided either via equipment usage history or the fault-detection and diagnosis (FDD) related condition monitoring data, is fused with fab-state related data, such as WIP-levels, spare parts and maintenance staffing availability, queueing times and lengths, for system-level optimization of fab operations, which include scheduling and prioritization of maintenance, WIP routing, job assignments, and other production operations.

This section is organized according to the functional aspects of data analytics, focusing on FDD, VM, APC, and fab operations in semiconductor manufacturing. Due to space constraints, this section will address the aforementioned functionalities only in the etch process.

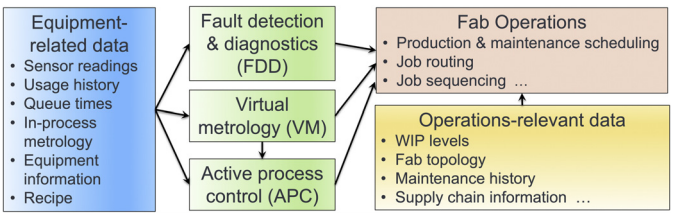

Fig. 51. Functionalities of data analytics in semiconductor manufacturing.

The key challenges observed in data analytics in etching are representative of what is seen of FDD, VM, APC, and fab operation in other processes as well.

7.1. Data analytics for FDD in etch processes

FDD research in semiconductor manufacturing is reviewed in [9]. This section covers the advancements afterward. A distinguishing trait of FDD is its process-centric (vs. equipment-centric) focus because IC manufacturing equipment is highly flexible, routinely performing different processes, involving different materials and chemistries, and having highly variable process parameters. Challenges associated with an equipment are therefore primarily driven by its processes. Furthermore, designs and models of equipment are highly confidential. This drives the FDD research towards being process-centric, which is why this subsection is process-centric, with a focus on etch processes.

FDD in etch processes predominantly relies on readings from classical sensors, such as pressure, flow, and radio frequency (RF), in the equipment. Data gathered from these sensors are also known as the status variable identification data (SVID). The use of raw SVID data for FDD is now more prevalent than the previously dominant reliance on fault detection and classification (FDC) data, which are elaborately curated from SVID in the form of informative data statistics extracted from windowed segments of signals. Such curation requires involvement of experts familiar with the etch process [81], which motivated the transition towards directly working with the raw SVID.

The increased numbers and sampling rates associated with SVID and the emergence of the use of complete optical emission spectroscopy (OES) sensor readings for etch process FDD [34,35,137] created a need for data analytics using deep learning (DL) methods. DL is a subset of machine learning (ML) methods, which use artificial neural networks (ANNs) with multiple hidden layers, enabling them to automatically learn the sensor reading features from the raw data [90]. Only 2 out of 40 papers on etch FDD employed DL concepts in [9]. About half of the recent papers on etch FDD employed DL.

One of the main challenges in FDD using sensor data is aligning informative signatures, because sensor data may vary in length or contain different numbers of process features. Among recent papers, the data alignment challenge is often tackled using DL models, which are agnostic to input signal lengths. The DL methods included attention-based [81] and convolutional neural network (CNN) based models [195,196], as well as a long short-term memory (LSTM) recurrent neural network [71]. Another methodological paradigm to feature extraction and alignment in the etch FDD is the use of dynamic time warping (DTW), which generates a warping path, along with a non-metric distance between two time series [124]. A central time series is used, with all other SVID being aligned to it via DTW to keep a consistent alignment of the extracted SVID features [53,91]. A more elaborate DTW-based approach, where a time series alignment kernel form of a DTW distance is introduced and employed to advance etch FDD via transfer learning across different etch equipment [210,211].

The use of entire SVID for etch FDD also exacerbated challenges associated with the dimensionality reduction. Principal component analysis (PCA) remains popular as a tractable and statistically interpretable method [47,197]. Nonlinear versions of PCA, such as kernel-PCA [149], have been employed for dimensionality reduction in research on etch FDD. Kernel-PCA approaches use a nonlinear mapping into a higher dimensional space, often with Gaussian or sigmoid kernels, and were followed in studies [211,197,175]. A non-linear PCA-based dimensionality reduction approach, referred to as principal polynomial analysis, was performed individually on each substep of the etch process [198]. Another nonlinear statistics-based approach to dimensionality reduction in etch FDD [199] utilized the classic independent component analysis [92] to cope with the non-Gaussian nature of the data.

Outside of various versions of PCA, dimensionality reduction in the etch FDD has been pursued using statistics-based embeddings, such as neighborhood preserving Gaussian mixture model-based embeddings [200], as well as via DL-based embeddings, which were

implemented using convolutional autoencoders [195,196] or stacked LSTM blocks inside a larger autoencoder network [71].

When it comes to recent research on etch FDD based on OES sensor [34,35,137], one can note that it typically relies on engineering knowledge about the underlying process and equipment physics, thus focusing on informative OES wavelengths and time-windows. The SVID curation in recent research seems to be like that of earlier research [9]. This will change and shift towards ML and DL based approaches, especially if one considers the more elaborate OES curation methods recently employed in etch VM (discussed in Sec. 6.2).

Besides the dimensionality reduction problems discussed above, significant label imbalances are a major challenge in etch FDD, because typical fab operations view fault occurrences as extremely rare events. Research focuses on anomaly or novelty detection methods that use only normal operational data to train the relevant models and detect faults as departures from those nominal patterns. Recent works in etch process anomaly (fault) detection accomplish this through a simple k-nearest neighbors (kNN) based approach, relying on detection when the distance to k nearest neighbors in a golden dataset exceeds a threshold [196,47,197,199]. Additional measures for novelty detection have included Mahalanobis distance [71], time series alignment kernel [210,211], and support vector data description [47].

An emerging paradigm on etch FDD as a response to class imbalance and data scarcity is data augmentation. The concept of generative adversarial networks was used to improve class imbalance by generating additional training samples, leading to performance improvements in OES-based etch FDD [35]. Another example is data augmentation. The problem of missing data labels was addressed via a DTW-based assignment of soft labels to unlabeled time series traces from an etch process, leading to improved anomaly detection performance [91].

Classification studies continue to be prominent in the recent research on etch FDD, with most research in anomaly detection based on a normal vs. faulty process data classification or accomplishing full FDD via multi-class classification differentiating between normal and multiple classes of faulty process data. The standard linear models, such as linear discriminant analysis [48], partial least squares (PLS) [39], and least absolute shrinkage and selection operator (LASSO) [91], have been used, although the problems may encourage the use of nonlinear classification methods. A significant growth has occurred in the use of DL-based methods, e.g., classification using multi-layer perceptron [34], CNNs [12,31], and self-attentive CNNs [81,155]. Ensemble tree approaches, such as explainable boosting machines [34], XGBoost [91], random forest (RF) [35,91,52,82], and AdaBoost [51], are popular methods which perform well with relatively small training sets. Discriminant analysis methods have also been explored, including multi-domain discriminant analysis [210] and kernel fisher discriminant analysis [47]. Finally, the classical support vector-based classification continues to be utilized as the primary classification approach [196,211,21,83], as one of multiple classification approaches to demonstrate novelties in data curation techniques [35,52], or as a comparison with the feature sub-space neighbor vote [47].

The etch FDD is studied with the emerging concept of artificial immune systems [137] by fusing signatures from sensor SVID and OES data, and employing mechanisms of negative, positive, and clonal selections to detect and characterize faults in etch processes. Another research in etch FDD tackles the problem of quick domain adaptation [12]. In this study, FDD was enabled for a new, previously unseen etch process recipe by training a CNN that projects new recipe data into the space of existing recipes with well-trained FDD models.

7.2. Virtual metrology in etch processes

Physical metrology of nm-scale quality characteristics in advanced technology node semiconductor manufacturing is both cost- and time-prohibitive. This leads to physical measurements being available in only a few locations on a very small portion of wafers. As a remedy to this situation, the concept of VM emerges as a solution. VM pursues a mapping that connects signatures extracted from the equipment sensor data gathered during processing of a wafer to estimate the corresponding outgoing geometric or electrical quality characteristics of features produced on that wafer. Fig. 52 illustrates the increase in the density of available in-process metrology enabled by VM. This section focuses on advancements in VM in etch processes.

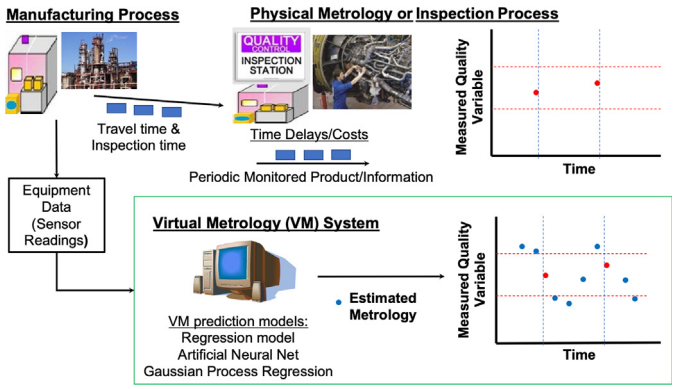

Fig. 52. The concept of VM increasing the density of available metrology. The missing quality information between two physical metrology measurements (red dots) vs. VM estimates of quality (blue dots).

A recent review of VM in IC manufacturing is in [109]. Due to the inherent nonlinearities of the governing physical phenomena, the most recent research in VM in etch processes has focused on the use of nonlinear regression model forms, like CNNs, transformers, and regression trees [109]. Another significant trend is the incorporation into the relevant VM models of complete OES sensor data emitted by plasma during etch processes. An example of OES data from 3 stages of a 9-stage industrial etch process is shown in Fig. 53. The current OES temporal sampling rate is 10 Hz and a wavelength resolution is finer than 0.5 nm. Both are expected to increase. The multi-step etch processes in modern semiconductor manufacturing may emit 10 or more OES readings per single etch process, extracting useful information from such a dense and large dataset is a challenge. This was addressed in the past by expertly selected OES wavelengths depicting relevant plasma chemistries. Recent research has clearly shown that more VM-informative signatures can be extracted by analyzing all OES wavelengths across all OES time samples [32,42,84,172]. OES readings are treated as images and OES-based VM is approached using a purely DL-based method [172]. A more analytically tractable approach to OES-based VM in reactive ion etching processes has key OES wavelength bands identified via a PCA-based procedure [32] with VM model inputs being steady-state and transient features extracted from the temporal evolution of those wavelength bands using a process-agnostic dynamics-inspired feature extraction method [68]. Tractability was pursued in [84], where OES data were decomposed into temporal and frequency contributions using functional singular value decomposition (SVD) [73], with those contributions serving as inputs for functional linear regression [22] based VM models of post-etch CDs. An SVD-based procedure was also used in [42], where VM-related information was extracted from OES sensor readings via numerical implementation of a classical SVD applied to flattened OES data. The ridge regression-based VM model employing SVD-enabled OES projections yielded the highest accuracy, well surpassing a coefficient of determination (R^2) of 60% in some portions of the data. Furthermore, the tractability of the OES processing and VM-model enabled the identification of time-wavelength patterns in OES data, which contributed to the VM model and illustrated the importance of concurrent analysis of temporal evolutions of all OES wavelengths [42]. This analysis of OES contributions to VM models showed that curation of OES signals using classical image filtering and processing techniques, which focus on intensity, is not adequate for OES-based VM.

Driven by both the increased use of OES data, as well as the increased number and density of SVID-type sensors, the latest research in etch VM led to methodological improvements that enable VM

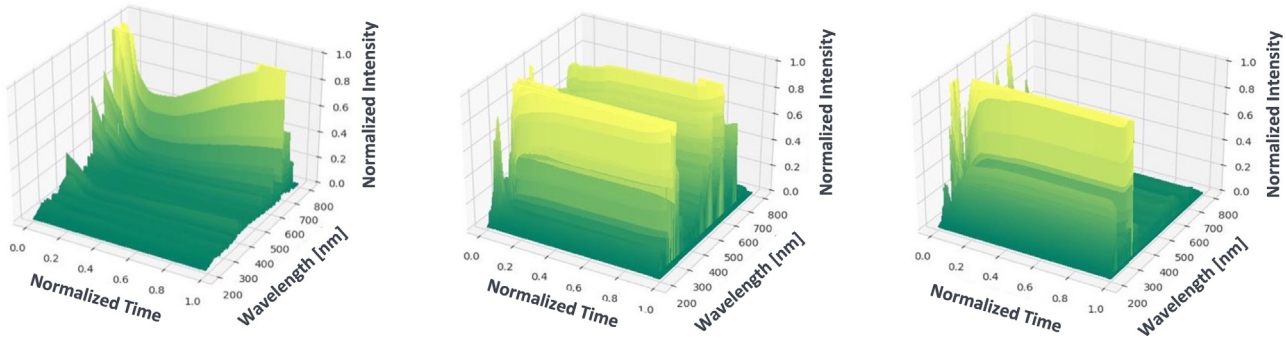

Fig. 53. Complete OES readings from 3 stages of a 9-stage etch process performed on a single wafer with normalized time-axis and spectral intensities.

models with large-dimensional feature vectors extracted from equipment data. A PCA-modification referred to as true sparse PCA (TS-PCA) is proposed to address the challenge of large-dimensional VM-model inputs [184]. The number of sensors employed to build a VM model is reduced by using a dynamic algorithm that maximizes the explained variance within the dataset, while constraining the number of sensors utilized in all principal components identified via PCA [176]. The LASSO regularization framework was employed to accomplish input dimensionality reduction for a deep neural network-based VM model, with the LASSO-based regularization being superior when compared to other sparse regularization methods [36].

Recent literature indicates that performance of etch VM can be improved by increasing the numbers of sensor readings and features extracted from them [37,185,201]. There are benefits of jointly modeling multiple post-etch CDs in the VM framework via a single CNN-based model referred to as a multiple-input-multiple-output (MIMO) CNN [37]. Robustness of an etch process VM model is maximized by randomly sampling different subsets of sensor data and thus maximizing the variance in a multi-sensor VM-model input [201]. A VM model is built for each individual sensory input, with those models being fused into a final regression tree based VM model [185].

Improvements in etch VM were also achieved via adaptations of the training process [72], where a cosine-distance based clustering method was employed to identify informative data samples for VM model training and upkeeping. When applied to etch and diffusion processes, this led to increased VM model robustness to inherently present temporal drifts and shifts in those processes.

A unique VM problem was considered in [93], where focus was on estimating the variance of electrical testing metrology rather than the metrology values themselves. A modified bagging trees-based regression model was employed to estimate the variance of post-etch electrical tests using 250 sensor readings gathered during the etch process as inputs [14]. Such a focus on the variance naturally lends this VM-based approach to be also used for process monitoring.

Deeply related to VM-based process monitoring or any kind of decision-making based on VM is the notion of explainability of VM models. Explainability in etch VM is pursued with attention networks used to curate equipment sensory features informing a VM model in an atomic layer etching process, and process engineers confirming the importance of those features [67]. Using a physics-based approach, the LASSO regression was coupled with domain expert knowledge by allowing the domain expert to penalize the coefficients of features from irrelevant sensors to ensure that the model agrees with their physical intuition [80]. This work was extended to enable online implementation, leading to long-term efficacy of the etch VM model over the course of a year [162].

Another methodological direction that can be highly beneficial for one's ability to close the process loop based on VM results is the use of first-principles models of the etching processes to realize VM [85]. VM was pursued using physics-based features and models explaining the complex plasma interactions in a $C_4F_8/Ar/O_2$ plasma etch process [85]. The same group employs a similar methodological paradigm to realize VM models of etching trenches in $Ge_2Sb_2Te_5$ in Ar/SF_6 plasma [156]. Such an approach to VM obviously requires deep knowledge of the underlying etch process to curate adequate sensor features and models for every new recipe, which limits the generalizability of this paradigm. However, remarkable VM model accuracy and long-term robustness in industrial implementation reported in [85,156] was driven by the physics-based knowledge embedded in those approaches. Such high accuracy and robustness of the resulting VM models enabled the VM-based APC of those processes in [138]. The concept of APC in etch processes will be discussed in detail next.

7.3. Data analytics for APC in etch processes

Due to the lack of adequate sensing needed to characterize the complex process physics coupled with the chemically and physically prohibitive environments inside the equipment, a majority of APC has been in the form of ex situ process adaptations from one wafer to another or one lot to another, rather than in situ adaptations, as the wafer is being processed. Furthermore, due to the lack of sufficiently fast metrology, these ex-situ APC adaptations are done sparsely in a run-to-run (RtR) fashion, i.e., lot-to-lot or batch-to-batch, rather than wafer-to-wafer [123].

The early RtR research focused on developing methods for building and maintaining process-level models that relate controllable process parameters with the outgoing product quality, as well as on methods to predict the ever-present random differences between desired and realized control commands, which are often referred to as process biases. Through these capabilities, the RtR paradigm enables compensation of process biases, thus reducing variability in the outgoing product quality and increasing yield. Those early developments focused on methodological advancements that relied on tractable forms of process and prediction models, such as exponentially weighted moving average and Kalman filter-based predictive models, which enabled theoretical findings regarding the stability and performance of the relevant RtR solutions [123].

Review of the RtR research [164,100] and the analysis of the research published after these survey papers indicate that methodological RtR developments continue to be vibrant. Recent progress has been mainly driven by the increased availability of data and advancements in AI/ML techniques. Such a methodological shift led to a greater emphasis on practical implementations in processes and a lesser emphasis on theoretical results on controller stability and performance.

In this context, active closed-loop control of etch processes plays a significant role, with the main challenge being the sparsity of metrology data. Physical measurements of post-etch CDs are usually available only once per lot of 25 wafers or even less frequently. Research in etch VM has been active due to the complexity of the underlying physics. VM has been used for closed-loop process control, with VM-based CD estimates informing the process control in a wafer-to-wafer manner, rather than as RtR or batch-to-batch [138,94].

Specifically, a PLS-based VM model was used to estimate center-to-edge CD variability in plasma etching utilizing expert-curated features extracted from OES sensors, voltage-current probe readings,

and SVID sensor traces into estimates of center-to-edge CD variability [94]. A high R^2 of 0.89 of the resulting VM model was exploited to facilitate wafer-to-wafer APC by actuating center-to-edge gas flow rates in the etch chamber, with APC experiments involving more than 60 wafers showing a significant reduction in CD ranges.

A uniquely detailed model of the underlying physics and chemistry was used to facilitate VM-based APC in a plasma-etch process employed for the fabrication of optical light-emitting diode (OLED) displays [138]. The domain knowledge and physics principles are combined to fuse models of electron energy distribution functions of plasma inside the etch chamber, with hundreds of condition characterization variables extracted from the densely sampled SVID traces and key OES wavelengths. This connection between big data gathered from the equipment and physics-based models enabled predictions of etch rates with R^2 well over 85%, as well as highly accurate FDD capabilities. The industrial implementation of FDD and APC solutions led to 25% savings in reduced yield losses over a 5-year period. Such uniquely successful results are an illustration of how the fusion of physics and AI/ML-based methods can be used to transition lab-scale research into industrially relevant solutions for challenging problems of data analytics in OLED manufacturing.

7.4. Data analytics for optimization of operations in etch processes

Converting raw process and fab data into operational decisions in vacuum processes, such as etching, is a uniquely challenging problem. Etch tools are usually realized in the form of so-called cluster tools, consisting of multiple process chambers, each of which executes processes that may be significantly different in terms of their chemistries and underlying parameters. These chambers are connected via a robotic manipulator capable of rapidly delivering wafers from the load-lock of the tool, where wafers which need to be processed on the tool are stored, to any of the processing chambers. Consequently, each etch tool can be seen as a job-shop style "mini" manufacturing system, as illustrated in Fig. 54.

With this unique level of flexibility and integration come unique challenges in terms of operational decision-making. Traditionally, decisions in terms of production scheduling and sequencing, and maintenance scheduling are done separately. However, on etching, the fact that each chamber can perform different operations, each of which degrades that chamber and deteriorates outgoing product quality with different dynamics, means that production scheduling and sequencing have a direct and strong impact on degradation in each of those chambers. In such an environment, one can respond to chamber degradation not just in a traditional way, by scheduling preventive maintenance, but also via reconfiguration, by sending wafers for processing in a different chamber and, if possible, sequencing products to come out of the load-lock in a way that the degraded chamber ends up executing less-demanding operations, or being avoided completely.

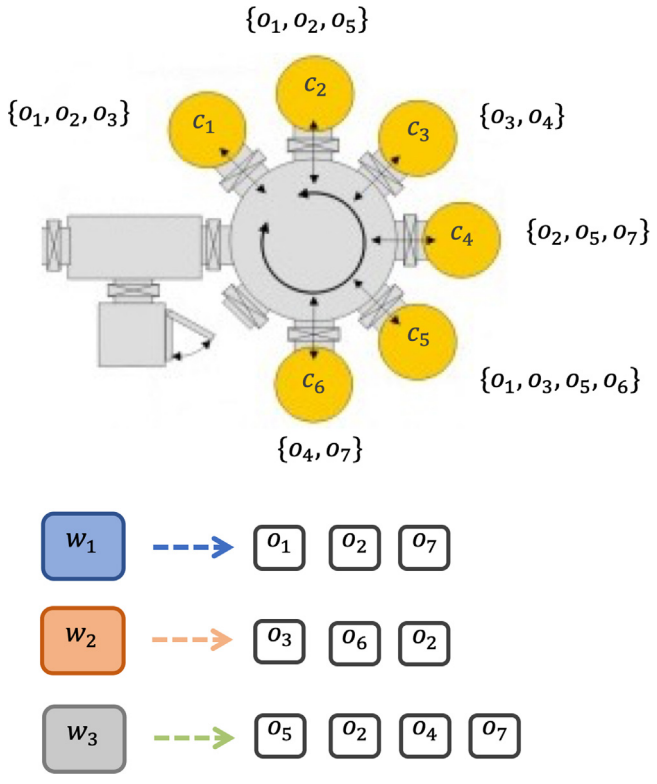

Fig. 54. Illustration of a cluster tool consisting of 6 chambers (labeled with $c_1 - c_6$), with each chamber being able to execute operations (labeled with $o_1 - o_7$). Tool needs to process three types of wafers (labeled with $w_1 - w_3$), with each of those wafer types requiring a certain sequence of operations.

Consequently, integrated decision-making across production and maintenance scheduling domains is of paramount importance when it comes to optimization and control of etch process operations. Recent research has shown that when it comes to flexible job-shop style manufacturing systems, such as etch systems illustrated in Fig. 54, there exist significant benefits of integrated decision-making in the domains of production sequencing and maintenance scheduling relative to the traditional, fragmented approach to optimization of those operations [26]. It was shown that benefits of integrated decision-making grow with higher production benefits and with higher penalties for unscheduled machine downtimes. These findings emphasize the importance of this paradigm in manufacturing, especially when it comes to production of high-end products, like GPUs and AI chips.

It should be noted that the existing research relied on discrete-event simulation (DES) based models of system operations, which are then used to inform meta-heuristic optimization of operational decisions. Building such models and tuning them to actual parameters of the manufacturing system is a time-consuming process. In addition, DES-based optimization is a computationally demanding paradigm, which further hampers real-time adaptations and decision-making necessary for full realization of digital twin functionalities [15]. Recent research which pursues operational decision making using surrogate modeling and reinforcement learning paradigms holds significant promise in overcoming this weakness [15,88,202].

8. Technical trends in future semiconductor manufacturing

Advanced packaging, yield optimization, sustainability, and in-space are timely emerging in semiconductor manufacturing. We discuss their technical trends in the following subsections.

8.1. Advanced packaging

Packaging is the assembly or bonding of the IC die with the Si wafer, glass panel, or PCB as the substrate. Bonding can be conducted in die-to-die, die-to-wafer, wafer-to-wafer, and die-to-panel size levels. Three key advanced packaging topics are the panel-level packaging (PLP), high-density hybrid bonding, and Si photonics, which are all leaving the labs into production and will be covered in this section.

In conventional packaging, the system on a chip (SoC) is bonded to organic substrates—face-up for wire bonding and face-down for flip-chip bonding—to connect processors and dynamic random-access memory (DRAM). As shown in Fig. 55, modern GPU and CPU modules for AI and high-performance computing use Si interposers with chip-to-wafer wafer-level packaging (WLP). These interposers connect processors and 3D DRAM (also known as high bandwidth memory (HBM)) via short redistribution layers (RDLs) and TSVs.

This approach aligns with the chiplet design paradigm, where functional blocks are partitioned and interconnected on Si interposers, improving yield by enabling small-die fabrication across both advanced (e.g., 3 nm for xPU (CPU, GPU, or NPU)) and legacy (e.g., 28 nm for 3D DRAM) technology nodes. A key trend is the increasing size of Si interposers. The standard reticle size is 26 mm × 33 mm (1 ×) in lithography. Growing AI workloads require more layers of HBM and a larger size compute die, pushing the interposer size up to 12 × and yielding only four dies per 300 mm wafer.

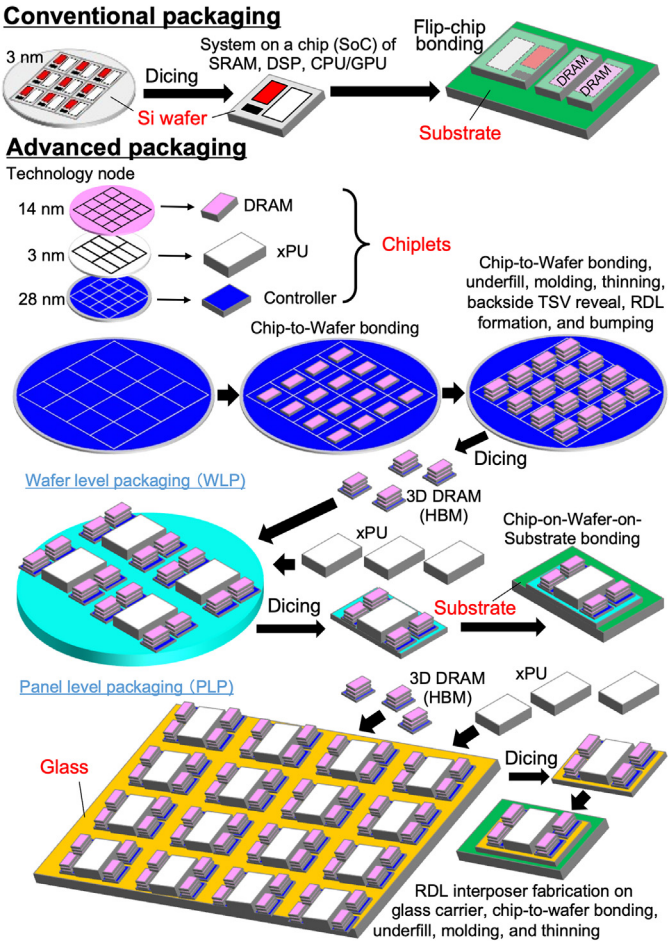

Fig. 55. Chip-to-wafer integration from WLP to PLP.

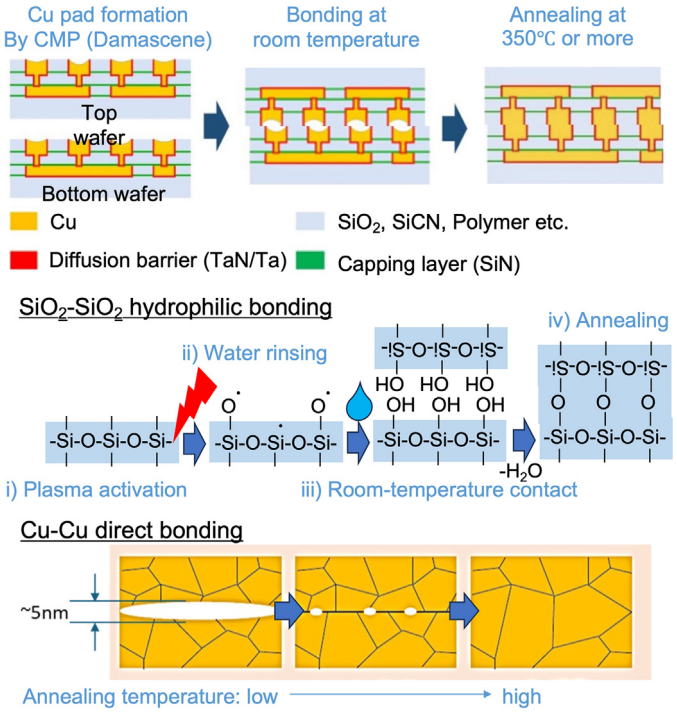

Fig. 56. Hybrid bonding process with dishing after CMP.

To overcome this, PLP is being explored to produce more RDL organic interposers [19]. While organic substrates struggle with 2 µm line width and spacing (2 µm/2 µm L/S) scaling, Si interposers support 0.5 µm/0.5 µm line width and spacing. Despite limitations in high-speed signaling for PLD due to resolution, pitch and line space dimensions, and overlay accuracy, PLP offers area advantages, and glass carriers are being adopted for production. To address signal integrity, Si bridges [108] or glass core substrates [53] are being innovated for fine-pitch interconnects.

The chiplet architecture enhances the performance in memory-processor systems. Recently, both memory and xPU have been stacked in 3D configurations, such as static random-access memory (SRAM) on AI accelerator or AI accelerator on SRAM dies [182]. However, since DRAM must operate below 95°C [86] and low thermal conductivity in underfill layers (≅ 15 µm) and microbump structures with Sn-based solder and Cu pillars, HBM is not directly stacked above or below processors due to thermal constraints.

To mitigate thermal issues, Cu-Cu and SiO_2-SiO_2 hybrid bonding techniques are employed, eliminating solder and enabling direct bonding via CMP Cu pads and SiO_2 surfaces (Fig. 56). Hybrid bonding is a technique for permanent wafer-to-wafer or die-to-wafer bonding through the simultaneous bonding of dielectric (e.g., SiO_2 hydrophilic bonding) and metal (e.g., Cu) materials. These advantages allow for significant device size and weight reductions, which make hybrid bonding a key enabler for advanced packaging of future IC devices.

Despite CMP-induced Cu pad concave dishing due to the lower hardness of Cu relative to SiO_2, controlled barrier layers (e.g., Ta) may improve bonding. However, gap-filling remains difficult when dishing exceeds 5 nm [147]. Hybrid bonding proceeds with room-temperature alignment of SiO_2 layers, followed by atomic-level bonding and annealing to enable Cu diffusion. Cu oxidation (CuO, Cu_2O, $Cu(OH)_2$) affects diffusion length. While the 350°C annealing temperature is typical, annealing temperatures below 250°C are preferred for memory chips (e.g., HBM) due to concerns about material degradation and reliability.

Hybrid bonding is still evolving, with the pitch and Cu pad size reducing steadily (with R&D in 150 nm level by Tokyo Electron), and its performance approaching that of monolithic ICs. Key challenges for hybrid bonding are: 1) low-temperature bondability and connectivity, 2) high bonding energy, 3) Cu contamination, 4) CMP (dishing, erosion, corrosion), 5) pretreatment (activation) and wait time, 6) trade-off of throughput vs. accuracy, 7) warpage and residual stress, 8) miniaturization, 9) electroplating the nano-crystalline Cu in small feature size and with high thermal conductivity, 10) Cu passivation against CuOx (Cu diffusion length), 11) Cu pad density and fine-pitch capability, 12) dielectric property and low power loss (SiO_2, SiCN, polymers, etc.), 13) reliability in electromigration tests, 14) water absorption, 15) heat dissipation, 16) tolerability of containment (e.g., particle), and 17) cost.

Co-packaged optics (CPO) is an emerging packaging trend under active R&D for AI computing data transmission, aiming to overcome the data bandwidth, latency (delay), and power consumption/heat barriers of traditional electrical interconnects. As shown in Fig. 57, an optical fiber connects two photonic ICs (PICs), which switch optical and electrical signals for fast and energy-efficient data transmission between two electric ICs (EICs).

CPO utilizes two structural designs, as shown in Fig. 57. Design 1 is the embedded PICs/EICs. The PIC is embedded in the substrate and interconnected with the EIC via a 2.3D RDL organic interposer [128], which is a high-density organic packaging solution designed to bridge the gap between lower-cost organic substrates and high-performance 2.5D silicon interposers. Optical coupling between the fiber unit array (FUA) and PIC can be achieved via edge, grating, or mirror coupling. In Fig. 57, edge coupling connects fibers to polymer waveguides that link to SiN waveguides in the PIC via mirror coupling. SiN waveguides interface with Ge photodiodes for optical-to-electrical conversion. To shorten the distance between the EIC and PIC, fan-out WLP is studied based on the embedding of either the PIC or EIC in a 2.3D RDL interposer [18].

Design 2 is stacked PIC/EIC via hybrid bonding. The stacked approach, exemplified by the compact universal photonic engine (COUPE) of TSMC, stacks EIC on PIC via hybrid bonding, mounted on a substrate using solder bumps (Fig. 57) [107]. Grating coupling connects FUA to SiN waveguides. Hybrid bonding enables fine-pitch scale joining, but it presents challenges such as

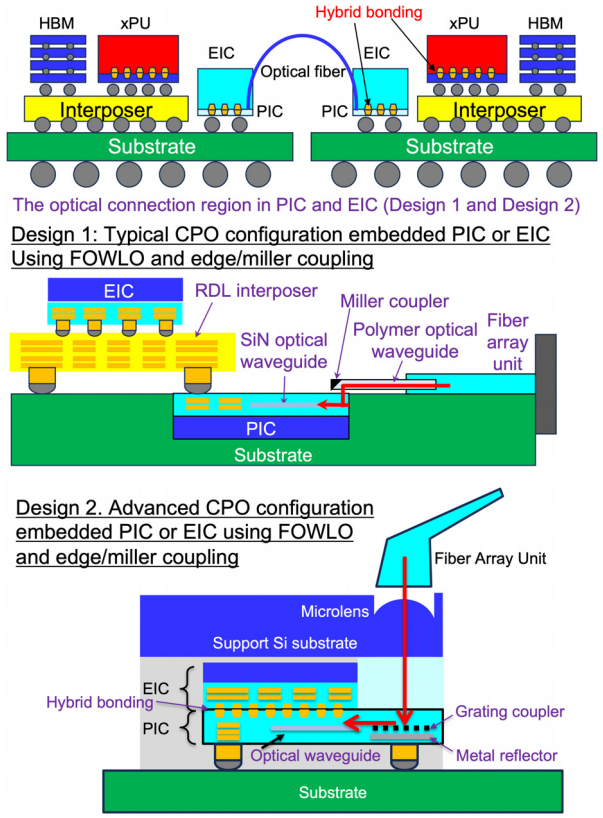

Fig. 57. xPU-to-xPU link with a CPO with embedded PIC and EIC using the fan-out wafer-level packaging and PIC/EIC stack using hybrid bonding.

dimensional mismatch between the FUA and waveguides. This method directly connects the fiber and PIC, resulting in low transmission loss, but it has limited design flexibility and requires technical improvements in terms of both energy efficiency (in pJ/bit) and data transfer density (in Tbps/mm).

8.2. Yield and productivity optimization

A major direction one can expect in future research and practice in all aspects of data analytics in etch processes is the enablement and proliferation of the use of densely sampled temporal evolution of OES. OES is a reliable sensing technique, which enables the determination of the elemental composition of plasma with low detection limits, high precision, and high accuracy [43]. Currently, its predominant use is based on analyzing the temporal evolution of a very limited set of wavelengths identified in such a way that they are representative of key elements or compounds relevant to the underlying process. Thus, the etch process can be optimized, monitored, and controlled via the concept of end-point detection. Namely, ideal etch results can be obtained by switching off the plasma and thus stopping the etch process when the concentration of the compound of interest reaches its local optimum, as determined by analyzing the temporal evolution of OES wavelengths that characterize the concentrations of those compounds in the plasma. Such confinement to analyzing just several wavelengths hampers utilization of full information content in OES readings, since useful information can be potentially found in all wavelengths [32,172,68]. Each chemical element emits energy at multiple wavelengths, with spectral peaks from different elements potentially overlapping and hampering analysis based solely on a narrow wavelength band. Theoretical wavelengths at which elements emit energy when excited into a plasma state are broadened due to the high speeds at which particles move in plasma (Doppler broadening), particle collisions (pressure broadening), as well as the influence of strong electrical (Stark broadening), and magnetic fields (Zeeman effect) [50]. The reason why earlier research and current industrial practice are limited to analyzing just a few OES wavelengths is the fact that analyzing the entire OES reading inevitably leads to data overload. For example, with the standard temporal resolution of 10 Hz with which each wavelength is sampled, as well as the wavelength resolution, which can easily reach 0.5 nm, a complete OES reading from a single stage of an etch process can be larger than 50 MB. Consequently, in multistage etch processes where each etch stage produces an OES reading, their analysis would involve seeking useful information in 100 s of megabytes of data per wafer. This quickly becomes unmanageable in a modern fab where hundreds of thousands of wafers pass through an etch tool in just a few months, leading to solutions that focus on just a few wavelengths.

The recent rapid advance of computing technologies, as well as the growth of capabilities of the big data mining algorithms based on AI/ML, are now turning the attention of the researchers and practitioners towards further enabling of mining of complete OES readings in FDD [81,34,35,137] and etch VM [42,84,73,22]. This trend is expected to grow, enabling further improvements in FDD and VM, as well as the pursuit of APC in etch processes through mining of complete OES signals. Such advancements are certainly going to positively impact data analytics and process control in plasma deposition processes as well, since modeling, optimization, and control of plasma are crucial problems in this domain. Finally, besides algorithmic advancements aimed at better usage of OES readings in etch processes, one can expect the incorporation of a larger number of OES sensors for each etch process stage, which would enable better representation of the distributed nature of plasma in etch chambers. The challenge of data overload, which previously prevented the exploration of such concepts, is now less of a limiting factor, which should lead to exciting research and discoveries in the future.

8.3. Sustainable semiconductor manufacturing

IC manufacturing is highly energy and water intensive [177], yet significant reductions are achievable through efficiency and reuse measures without degrading product quality and throughput [206]. For example, a 0.33 NA EUV machine requires 1.2 MW, and a 0.55 high NA EUV machine needs 1.4 MW of electricity to operate. TSMC consumed about 9% of the electricity in Taiwan in 2025 and is expected to rise to 24% in 2030 [170]. Generating and maintaining plasma in a chamber for in-reactive ion etching processes requires well over 60 kW, with 1000 or more chambers in a fab and this power being expected to reach 100 kW with next-generation processes. The potential to reduce power requirements through operational optimization can exceed 25%.

Various fluorocarbons are used in fab processes, such as dry etching and chamber cleaning, for which no proven substitutes exist [63]. Volatile organic compounds and air contaminants are also emitted from IC manufacturing processes. Multiple tools exist to aid ESM design and its processes toward greater efficiency. The SEMI S23 [150] standard, published by Semiconductor Equipment and Materials International (SEMI) in the US, provides a method for analyzing the use of energy, utilities, and materials by ESM, along with energy equivalent values to assess energy consumption. Similarly, the imec.netzero “virtual fab webapp” platform [74] analyzes quantitatively the environmental impact of high-volume semiconductor manufacturing. Applied Materials’ EcoTwin™ digital twin platform [5] is an onboard software model of the tool that monitors real-time energy and chemical usage and automatically reports estimated environmental impacts using measured data, domain knowledge, and models, enabling users to choose optimal recipes [63].

These tools have been utilized in industry. Ma et al. [115] utilized the SEMI S23 standard [150] and field measurements to demonstrate a 49% reduction in energy consumption from their first- to third-generation EUV vacuum system by minimizing the number of vacuum pumps. An EUV hydrogen recycling system (HRS) has also been developed and tested to replace gas-fired abatement, demonstrating more than 80% recovery of H_2, with purified H_2 containing less than 1 ppm of N_2 and 1 ppb of H_2O. With large-capacity vacuum pumps and the HRS, each EUV system can reduce equivalent CO_2 outputs by 536 tons/-year [13]. In another example, Applied Materials utilized EcoTwin™ and imec.netzero to estimate the environmental impacts of full wafer

production and concluded that a new pattern-shaping system uses 10% less equivalent energy at the system level [63]. Finally, Baek et al. [13] outlined how Samsung applied efficient chemistry and engineering in CVD. The hibernation operation of systems may save up to 15% in electrical and chemical consumption costs.

8.4. In-space semiconductor manufacturing

The future of semiconductor manufacturing is poised to extend beyond Earth's surface, leveraging unique environments, such as microgravity, ultra-high vacuum, and reduced contamination, to create unprecedented opportunities for defect-free crystal growth, advanced thin-film deposition, and lithography. These advantages are particularly promising to overcome limitations in terrestrial fabrication for ultra-fast processors and radiation-hardened electronics, which are essential for deep space missions [40].

In-space semiconductor manufacturing will necessitate reimagined process systems that are autonomous, miniaturized, and adaptive to the extreme variability of orbital and planetary environments. Additive microfabrication, directed self-assembly, and photonic curing techniques are emerging as enablers to produce high-performance components in orbit. Integration with the in-situ resource utilization, such as refining Si or Ga from lunar regolith, could reduce earth-launch dependencies.

The US National Aeronautics and Space Administration [40], the European Space Agency, and the commercial sector are already piloting early-stage electronics fabrication on the International Space Station. The next phase involves scaling to fully integrated, robotically managed fabs within orbital platforms, such as Starlab, Axiom Station, or future Moon bases. Pioneering research, e.g., [17,144], has been conducted. This evolution promises resilient, distributed electronics supply chains supporting space infrastructure, while also offering insights for ultra-clean, potentially zero-emission terrestrial fabs.

The convergence of space-based manufacturing and semiconductor innovation will catalyze a paradigm shift, transforming how we design, fabricate, and deploy critical electronics.

9. Societal trends in future semiconductor manufacturing

The technology workforce policy and training ethical manufacturing leaders are important topics that will determine the future competitiveness and will be elaborated further.

9.1. Workforce policy—current and future

A trained workforce (about 2000–4000 per fab) is required to run the continuous twenty-four hours a day, seven days a week (24/7) fab and some R&D operations. For example, the immersion DUV and EUV operations in a plant of the fab require a team of over 60 and 100, respectively, of well-trained engineers and technicians. With reportedly over 100 EUV machines at TSMC and more than 40 at Samsung in 2023 [99], an extensive trained team and associated workforce training program are required for EUV and other operations. This workforce needs to be well educated in basic knowledge and communication skills before the professional training in various fab operations. A good example is TSMC's Newcomer Training Center where all new employees need to be trained for four two-week modules in 1) equipment, 2) processes, 3) operations, and 4) data/AI before entering the fab.

National policies have been developed to educate and train such large workforces. China and Taiwan have established Semiconductor Colleges. For example, in Taiwan, the College of Semiconductor Research at National Tsing Hua University is focused on post-graduate education with four fields of study in semiconductor 1) design, 2) device, 3) process, and 4) material. Imec in Belgium is renowned in semiconductor research and training, particularly in EUV. Such focused education and training programs in semiconductor manufacturing can be adapted and modified to other regions where challenges and opportunities may differ. Training homegrown leaders is also critical. For example, over two-thirds of top executives at TSMC were educated solely in Taiwan, without any overseas education.

The technology workforce policy focuses on the timely and precise supply of educated talent with fundamental knowledge to the semiconductor manufacturing industry. A strong semiconductor industry typically has a strong technology workforce policy behind its success. A technology workforce policy fosters an industry-academia-government partnership for education programs in areas of employment needs. Unlike an industry policy, which inevitably needs to select winners of technology and companies, the technology workforce policy focuses on education and training. Because it takes 3–4 years to build and start a new fab, workforce needs in specific technical areas (e.g., lithography, metrology, and others) can be forecasted. Students with a manufacturing background will be recruited to study in these technical areas at universities and community colleges, with mentorship and internship opportunities provided by industry.

The partnership between industry and academia provides R&D topics and funding to educate talented individuals. An effective technology workforce policy has been shown to 1) recruit top students with the culture and career aspirations in semiconductor manufacturing, 2) identify an industry-relevant topic, and 3) match students with mentors to learn, grow, and contribute to the semiconductor manufacturing industry. In Taiwan, the 50-year government programs (Fig. 6) were both industrial and technology workforce policies. Higher education institutions benefited to educate and supply talents to the semiconductor industry [27].

The technology workforce policy could impact the semiconductor industry within 1–2 years. In comparison, the applied R&D projects may take 2–3 years to develop a technology. The basic research project may take 3–4 years for discovery. Both basic and applied research are still important for innovating knowledge and technology, and they can coexist well with technology workforce policy.

9.2. Training ethical semiconductor manufacturing leaders

Key manufacturing decisions in the semiconductor industry have significant financial consequences. These decisions are challenging and will have a significant impact on the cost of a fab over the next 3–4 years, as well as the fab yield and cost of IC production over the next 15–20 years. The right directions in manufacturing may be missed when business leaders make ill-advised decisions based on wrong advice. For example, Intel's decisions to use multipatterning immersion DUV, instead of EUV, in lithography for the 7 nm technology node has perhaps cost their manufacturing leadership.

Ethical decisions are essential for customer thrust, the core in Fig. 1, which leverages extensive experience and training in manufacturing and a humble and continuous learning attitude. TSMC's core values begin with "Integrity is our most basic and most important core value. We tell the truth. We believe the record of our accomplishments is the best proof of our merit. Hence, we do not brag. We do not make commitments lightly. Once we make a commitment, we devote ourselves completely to meeting that commitment," is an ethical benchmark in semiconductor manufacturing excellence [171]. Hiring and culture building are essential to build and train an ethical manufacturing workforce, especially on leaders making key manufacturing decisions.

10. Concluding remarks

This study outlined advancements in equipment, removal processes, in-line measurements, and data analytics for etching in semiconductor manufacturing as well as the technological, cultural, policy, workforce, and financial challenges to achieve the yield and long-term profitable operation in semiconductor manufacturing. The steady, decade-long government, industry, and workforce policies are essential to reach the goal of semiconductor manufacturing excellence. Future technical and societal trends were discussed. Late Professor S.M. Wu's three characteristics 1) knowledgeable leadership, 2) superior technology, and 3) empowered workforce of globally competitive manufacturing align well with our conclusions.

For CIRP Scientific and Technical Committees (STCs), there are many potential fields of study and research directions in semiconductor manufacturing, e.g., advanced packaging for AI computing (STC-A), digital twins for design in advanced packaging for AI (STC-Dn), additive manufacturing of electronics (STC-E), forming and joining in packaging (STC-F), abrasive processes for power electronics (STC-G), automated material handling systems in semiconductor manufacturing (STC-M), optical inspections for semiconductor manufacturing (STC-P), operations and supply chain modeling and optimization in semiconductor manufacturing (STC-O), and surface metrology in semiconductor manufacturing (STC-S). Adopting such a research topic will expand the CIRP community into emerging areas of semiconductor manufacturing and broaden its impact. A potential CWG is on *Manufacturing for AI Infrastrure,* which includes the immersion-cooling, Si photonics, optical communication, energy efficiency, and operational optimization. Technologies, such as AI, are continuing to evolve. The manufacturing has been driven to fulfill the technical needs.

NIST disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIST. Certain commercial equipment, instruments, or materials are identified to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Albert J. Shih: Writing – review & editing, Writing – original draft, Visualization, Project administration, Investigation, Conceptualization. **Wei Gao:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **Gregory W. Vogl:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **Dragan Djurdjanovic:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **Takafumi Fukushima:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization. **Ajay Malshe:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Conceptualization.

Acknowledgments

Authors thank Harold Bosse, Chunlei Karl Song, Roberto Dailey, Samuel Bertelson, Zhiyang Zhang, Tao Liu and Xin Xiong for their assistance and industrial partners, including Imec, Intel, ITRI, Samsung, and TSMC. This work is partially supported by the National Science Foundation in the US and the Yushan Scholar Program in Taiwan. We acknowledge CIRP Council and members for their support of the formation and attending this CWG.

References

[1] ACM Research (2024) AI Spurring Growth of Panel-Level Packaging. https://www.acmr.com/ai-spurring-growth-of-panel-level-packaging/; 2024.
[2] Ahmad M (2023) TSMC Upends 3-nm Roadmap With Three New Nodes, EDN. https://www.edn.com/tsmc-upends-3-nm-roadmap-with-three-new-nodes/#google_vignette; 2023 Accessed May 30, 2026.
[3] Aleksandar K (2024) ASML High-NA EUV Twinscan EXE Machines Cost $380 Million, 10-20 Units Already Booked TechPowerUp https://www.techpowerup.com/319071/asml-high-na-euv-twinscan-exe-machines-cost-usd-380-million-10-20-units-already-booked; 2024.
[4] Antonelli GA, Keller N (2016) Challenges in Optical Metrology of 3D Memory. *American Vacuum Society 63rd International Symposium & Exhibition*, Nashville, Tennessee, USA, .
[5] Applied Materials (2026) Ecotwin™ Eco-Efficiency Software. https://www.appliedmaterials.com/il/en/semiconductor/solutions-and-software/ai-x/ecot-win.html; 2026 Accessed May 30, 2026.
[6] Applied Materials (2026) Opta™ CMP. https://www.appliedmaterials.com/us/en/semiconductor/products/semiconductor-products/opta-cmp.html; 2026 Accessed May 30, 2026.
[7] Archenti A, Gao W, Donmez A, Savio E, Irino N (2024) Integrated Metrology for Advanced Manufacturing. *CIRP Annals* 73(2):639–665.
[8] Arden MW (2002) The International Technology Roadmap for Semiconductors—Perspectives and Challenges for the Next 15 Years. *Current Opinion in Solid State & Materials Science* 6(5):371–377.
[9] Arpitha V, Pani A (2022) Machine Learning Approaches for Fault Detection in Semiconductor Manufacturing Process: A Critical Review of Recent Applications and Future Perspectives. *Chemical and Biochemical Engineering Quarterly* 36 (1):1–16.
[10] Aung S, Houdong L, Sani F (2024) Real-time water leak detection and preventing yield excursions in lithography. *2024 35th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–4.
[11] Automated High-Throughput 3D Scanning Probe Metrology—QURDRA (2026) Nearfield Instruments. https://www.nearfieldinstruments.com/quadra/; 2026 Accessed May 30, 2026.
[12] Azamfar M, Li X, Lee J (2020) Deep Learning-Based Domain Adaptation Method for Fault Diagnosis in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 33(3):445–453.
[13] Baek S, Park J, Koh T, Kim D, Woo J, Jung J, Park S, Lee C, Choi C (2024) Achievement of Green and Sustainable CVD Through Process, Equipment and Systematic Optimization in Semiconductor Fabrication. *International Journal of Precision Engineering and Manufacturing-Green Technology* 11(4):1295–1316.
[14] Baskin II, Marcou G, Horvath D, Varnek A (2017) Bagging and Boosting of Regression Models. in Varnek A, (Ed.) *Tutorials in Chemoinformatics*, John Wiley & Sons, Ltd., New Jersy, 249–255.
[15] Behrendt S, Altenmüller T, May MC, Kuhnle A, Lanza G (2025) Real-to-Sim: Automatic Simulation Model Generation for a Digital Twin in Semiconductor Manufacturing. *Journal of Intelligent Manufacturing* 37(2):829–848.
[16] Belforte D (2016) Disco Develops High-Speed Laser Wafer Slicing Technology. *Laser Focus World* https://www.laserfocusworld.com/industrial-laser-solutions/article/14217072/disco-develops-high-speed-laser-wafer-slicing-technology.
[17] Bhundiya H, Marshall M, Cordero Z (2024) Fabrication Time Diagrams for in-Space Manufacturing of Large Reticulated Structures. *ASME Journal of Manufacturing Science and Engineering* 146(12):121004.
[18] Bhuvanendran N, Gourikutty S, Wu J, Lim T, Sandra S, Jong M, Yee C, Long L, Ho D, Choong C, Liang D, Tu X, Wang W, Tan C, See A, Wang H, Coccioli R, Tan R, Nagarajan R, Bhattacharya S (2024) A compact wafer-level heterogeneously integrated scalable optical transceiver for data centers. *IEEE 2024 74th Electronic Components and Technology Conference (ECTC)*, 1398–1403.
[19] Braun T, Becker K, Raatz S, Bader V, Bauer J, Aschenbrenner R, Voges S, Thomas T, Kahle R, Lang K (2015) From fan-out wafer to fan-out panel level packaging. *IEEE 2025 European Conference on Circuit Theory and Design (ECCTD)*, 1–4.
[20] Byambadorj T, Ceballos AC, MacWilliams KP, Prescop TA, Lam DK, Michalka TL, Bishop C, Sandstrom C, Olson T (2024) Transcending the reticle limit in on-wafer die integration and advanced packaging: full-wafer patterning with high-productivity electron beam lithography. *2024 74th IEEE Electronics Components and Technology Conference (ECTC)*, 11–15.
[21] Cai H, Feng J, Moyne J, Iskandar J, Armacost M, Li F, Lee J (2020) A framework for semi-automated fault detection configuration with automated feature extraction and limits setting. *IEEE 2020 34th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–6.
[22] Cai T, Hall P (2006) Prediction in Functional Linear Regression. *Annals of Statistics* 34(5):2159–2179.
[23] Calado V, Mathijssen S, Van Setten E, Finders J, Wittebrood F, Bouman W, Bhattacharyya K, Op'T Root W, McNamara E, McLaren M (2024) Focus Sensing Using Placement and CD Variation for High NA EUV Lithography. *Proceedings of SPIE* 12953:62–69.
[24] Camps JV, Saz-Carranza A (2023) *The European Chips Act: Europe's Quest for Semiconductor Autonomy*, Esade Geo-Center for Global Economy. Barcelona Policy Note.
[25] Canon U.S.A. (2026) FPA-1200NZ2C Nanoimprint Lithography Systems for Fine Patterning Applications. https://s7d1.scene7.com/is/content/canon/FPA-1200NZ2Cpdf; 2026.
[26] Celen M, Djurdjanovic D (2020) Integrated Maintenance and Operations Decision Making With Imperfect Degradation State Observations. *Journal of Manufacturing Systems* 55:302–316.
[27] Chang MF, Lin C, Shen C, Wang S, Chang K, Chang R, Yeh W (2021) The Role of Government Policy in the Building of a Global Semiconductor Industry. *Nature Electronics* 4(4):230–233.
[28] Chen YL, Cai YD, Xu M, Shimizu Y, Ito S, Gao W (2017) An Edge Reversal Method for Precision Measurement of Cutting Edge Radius of Single Point Diamond Tools. *Precision Engineering* 50:380–387.
[29] Chen X, Wang C, Yang T, Liu J, Luo C, Liu S (2022) Inline optical measurement and inspection for IC manufacturing: state-of-the-art, Challenges, and Perspectives. *Laser & Optoelectronics Progress* 59(9):0922025.
[30] Chen CK, Weng PE, Chiang CW, Chen PS, Wen VCW, Gerson M, Hermanns CF, Thorsten H, Klaus E, Li YF (2024) Repair Processes on Next Generation EUV Materials. :132160Q.
[31] Chien C, Hung W, Liao E (2022) Redefining Monitoring Rules for Intelligent Fault Detection and Classification Via CNN Transfer Learning for Smart Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 35(2):158–165.

[32] Chien KC, Graff N, Chang C, Djurdjanovic D (2023) In-Situ Monitoring of Sapphire Nanostructure Etching Using Optical Emissions Spectroscopy. *Journal of Vacuum Science & Technology B* 41(6):062807.

[33] Chiffre LD, Carmignato S, Kruth J, Schmitt R, Weckenmann A (2014) Industrial Applications of Computed Tomography. *CIRP Annals* 63(2):655–677.

[34] Choi JE, An S, Lee Y, Lee Y, Kim D, Hong S (2024) Explainable Artificial Intelligence-Based Evidential Inferencing on Process Faults in Plasma Etching. *Journal of Physics D: Applied Physics* 57(18):185201.

[35] Choi JE, Seol D, Kim C, Hong S (2023) Generative Adversarial Network-Based Fault Detection in Semiconductor Equipment With Class-Imbalanced Data. *Sensors* 23(4):1889.

[36] Choi J, Son Y, Kang J (2024) Group-Exclusive Feature Group Lasso and Applications to Automatic Sensor Selection for Virtual Metrology in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 37(4):505–517.

[37] Choi J, Zhu M, Kang J, Jeong M (2024) Convolutional Neural Network Based Multi-Input Multi-Output Model for Multi-Sensor Multivariate Virtual Metrology in Semiconductor Manufacturing. *Annals of Operations Research* 339(1-2):185–201.

[38] Choi JH, Bin Im W, Kim HJ (2023) Plasma Resistant Glass (PRG) for Reducing Particulate Contamination During Plasma Etching in Semiconductor Manufacturing: A Review. *Materials Today Communications* 34:105267.

[39] Chouichi A, Blue J, Yugma C, Pasqualini F (2020) Chamber-to-Chamber Discrepancy Detection in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 33(1):86–95.

[40] Courtright Z (2023) On-Demand Manufacturing of Electronics. *NASA*. Accessed May 30, 2026.

[41] Cunningham JA (2012) Management: Using the Learning Curve as a Management Tool: The Learning Curve Can Help in Preparing Cost Reduction Programs, Pricing Forecasts, and Product Development Goals. *IEEE Spectrum* 17(6):45–48.

[42] Dailey R, Bertelson S, Kim J, Djurdjanovic D (2024) Virtual Metrology of Critical Dimensions in Plasma Etch Processes Using Entire Optical Emission Spectrum. *IEEE Transactions on Semiconductor Manufacturing* 37(3):363–372.

[43] Devia D, Rodriguez-Restrepo L, Restrepo-Parra E (2015) Methods Employed in Optical Emission Spectroscopy Analysis: A Review. *Ingeniería y Ciencia* 11 (21):239–267.

[44] DISCO (2026) DFG8830 Fully Automatic Grinder for Thinning Difficult-to-Process Materials. https://disco.co.jp/eg/products/grinder/dfg8830.html; 2026 Accessed May 30, 2026.

[45] DISCO (2026) Ultra-Thin Grinding. https://www.disco.co.jp/eg/solution/library/grinder/thin.html; 2026 Accessed May 30, 2026.

[46] Dréan G (2019) The Chips Industry: Moore and Rock's Laws. in Chamoux J, (Ed.) *The Digital Era 2: Political Economy Revisited*, Wiley, New Jersy, 125–135.

[47] Du X (2019) Fault detection Using Bispectral Features and One-Class Classifiers. *Journal of Process Control* 83:1–10.

[48] Du X, Yan J, Ma R (2020) Fault Classification of Nonlinear Small Sample Data Through Feature Sub-Space Neighbor Vote. *Electronics* 9(11):1–21.

[49] EBARA (2026) Model F-Rex CMP. https://www.ebara.com/global-en/products/FREX/; 2026 Accessed May 30, 2026.

[50] Fantz U (2006) Basics of Plasma Spectroscopy. *Plasma Sources Science and Technology* 15(4):137–147.

[51] Fan S, Hsu C, Tsai D, Chou M, Jen C, Tsou J (2022) Key Feature Identification for Monitoring Wafer-to-Wafer Variation in Semiconductor Manufacturing. *IEEE Transactions on Automation Science and Engineering: A Publication of the IEEE Robotics and Automation Society* 19(3):1530–1541.

[52] Feng J, Zhu F, Liu Z, Zhang J, Hua L, Wang S, Xie M (2021) Trace abstraction: a novel method to enhance fault detection in semiconductor manufacturing processes with an optimization approach. *IEEE 2021 Global Reliability and Prognostics & Health Management (PHM)* : 1–6.

[53] Fujimoto K, Okawa Y, Tai T, Kuramochi S (2024) Development of glass core substrate with the stress analysis, transmission characteristics and reliability. *2024 IEEE 74th Electronic Components and Technology Conference (ECTC)*, 546–550.

[54] Gao W (2021) *Surface Metrology for Micro- and Nanofabrication*, Springer, Singapore.

[55] Gao W, Kim SW, Bosse H, Haitjema H, Chen YL, Lu XD, Knapp W, Weckenmann A, Estler WT, Kunzmann H (2015) Measurement Technologies for Precision Positioning. *CIRP Annals* 64(2):773–796.

[56] Gao W, Haitjema H, Fang F, Leach R, Cheung C, Savio E, Linares J (2019) On-Machine and In-Process Surface Metrology for Precision Manufacturing. *CIRP Annals* 68(2):843–866.

[57] Gao W, Huang P, Yamada T, Kiyono S (2002) A compact and Sensitive Two-Dimensional Angle Probe for Flatness Measurement of Large Silicon Wafers. *Precision Engineering* 26(4):396–404.

[58] Gao W (2010) *Precision Nanometrology Sensors and Measuring Systems for Nanomanufacturing*, Springer, London.

[59] Gao W, Shimizu Y (2021) *Optical Metrology for Precision Engineering*, Walter De Gruyter, Berlin.

[60] Gao W, Kim SW, Bosse H, Minoshima K (2025) Dimensional Metrology Based on Ultrashort Pulse Laser and Optical Frequency Comb. *CIRP Annals* 74(2):993–1018.

[61] Gin P, Wormington M, Amasay Y, Sorkhabi S (2023) Inline Metrology of High Aspect Ratio Hole Tilt and Center Line Shift Using Small-Angle X-Ray Scattering. *Journal of Micro/Nanopatterning, Materials, and Metrology* 22(3):031205.

[62] Gooding M. xAI Targets One Million GPUs for Colossus supercomputer in Memphis: Massive Expansion of Data Center Planned, Data Center Dynamics, Accessed May 30, 2026. https://www.datacenterdynamics.com/en/news/xai-elon-musk-memphis-colossus-gpu/, 2024.

[63] Gross BJ, Reyes E, Kapadia S (2023) Sustainability-aware technology development at applied materials. *IEEE 2023 International Electron Devices Meeting (IEDM)* : 1–4.

[64] Group EV (2026) Gemini®: Automated Production Wafer Bonding System. https://www.evgroup.com/products/bonding/permanent-bonding-systems/gemini; 2026 Accessed May 30, 2026.

[65] Hafez W, Agnihotri P, Asoro M, Aykol M, Bains B, Bambery R, Bapna M, Barik A, Chatterjee A, Chiu P, Chu T (2023) Intel PowerVia technology: backside power delivery for high density and high-performance computing. *2023 IEEE Symposium on VLSI Technology and Circuits* : 1–2.

[66] Hamamatsu Photonics (2026) Stealth Dicing™ Technology. https://www.hamamatsu.com/jp/en/product/semiconductor-manufacturing-support-systems/stealth-dicing-technology.html; 2026 Accessed May 30, 2026.

[67] Han S, Min J, Ma J, Hwang G, Heo T, Kim Y, Kang S, Kim H, Park S, Sung K (2023) Deep learning-based virtual metrology in multivariate time series. *2023 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 30–37.

[68] Haq AA, Djurdjanovic D (2019) Dynamics-Inspired Feature Extraction in Semiconductor Manufacturing Processes. *Journal of Industrial Information Integration* 13:22–31.

[69] Hartfield C, Harris W, Gu A, Terada M, Viswanathan V, Jiao L, Rodgers T (2022) Emerging Technologies for Advanced 3D Package Characterization to Enable the More-than-Moore Era. *ECS Transactions* 109(2):15–29.

[70] Heil T, Waldow M, Capelli R, Schneider H, Ahmels L, Tu F, Schöneberg J, Marbach H (2021) Pushing the Limits of EUV Mask Repair: Addressing Sub-10 nm Defects With the Next Generation E-Beam-Based Mask Repair Tool. *Journal of Micro/Nanopatterning, Materials, and Metrology* 20(3):031013.

[71] Hosseinpour F, Ahmed I, Baraldi P, Zio E, Behzad M, Lewitschnig H (2025) A Novel Methodology Based on Long Short-Term Memory Stacked Autoencoders for Unsupervised Detection of Abnormal Working Conditions in Semiconductor Manufacturing Systems. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 239(5):1115–1133.

[72] Hsieh Y, Liu C, Huang S, Li C, Wilch J, Vogel-Heuser B, Cheng F, Chen C (2024) Developing the Keep-Important-Samples Scheme for Training The Advanced CNN-Based Automatic Virtual Metrology Models. *IEEE Robotics and Automation Letters* 9:7931–7938.

[73] Huang J, Shen H, Buja A (2009) The Analysis of Two-Way Functional DATA Using Two-way Regularized Singular Value Decompositions. *Journal of the American Statistical Association* 104(488):1609–1620.

[74] IMEC (2026) imec.netzero. https://netzero.imec-int.com/introduction; 2026 Accessed May 30, 2026.

[75] Infinitesima Ltd (2026) Metron3D Atomic Force Microscope. https://www.infinitesima.com/index.php/products/metron3d; 2026 Accessed May 30, 2026.

[76] Institute of Electrical and Electronics Engineers (2024) IEEE International Roadmap for Devices and Systems: 2024 IRDS Metrology. https://irds.ieee.org/images/files/pdf/2024/2024IRDS_MET.pdf; 2024 Accessed May 30, 2026.

[77] JFE Techno-Research Corporation (2026) Thickness Distribution Measuring Device FiDiCa. https://www.jfe-tec.co.jp/product/hikari/FiDiCa.html; 2026 Accessed May 30, 2026.

[78] Jhang K, Jang H, Park B, Ha J, Park I, Kim K (2002) Wavelet Analysis Based Deconvolution to Improve the Resolution of Scanning Acoustic Microscope Images for the Inspection of Thin Die Layer in Semiconductor. *NDT & E International: Independent Nondestructive Testing and Evaluation* 35:549–557.

[79] Kadoya S, Takahashi S, Michihata M, Ohtani N, Abe K, Arimura S (2024) Proposal of Damage-Free SiC Wafer Dicing Using Water Jet Guided Laser. *Materials Science Forum* 1124:85–89.

[80] Kikuchi K, Takada M, Xueting W, Ito T, Oomuro Y, Li G (2024) Quick update of virtual metrology using weighted transfer lasso. *IEEE 2024 International Symposium on Semiconductor Manufacturing (ISSM)*, 1–4.

[81] Kim E, Cho S, Lee B, Cho M (2019) Fault Detection and Diagnosis Using Self-Attentive Convolutional Neural Networks for Variable-Length Sensor Data in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 32(3):302–309.

[82] Kim E, An J, Cho H, Cho S, Lee B (2023) A Sensor Data Mining Process for Identifying Root Causes Associated With Low Yield in Semiconductor Manufacturing. *Data Technologies and Applications* 57(3):397–417.

[83] Kim D, Kang S, Cho S (2020) Expected Margin–Based Pattern Selection for Support Vector Machines. *Expert Systems with Applications* 139:112865.

[84] Kim S, Kwon Y, Kim J, Bae K, Oh H (2024) A Model Averaging Prediction of Two-Way Functional Data in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 37(1):76–86.

[85] Kim G, Kwon J, Lee I, Seo H, Park J, Shin J, Kim G (2024) Application of Plasma Information-Based Virtual Metrology (PI-VM) for Etching in $C_4F_8/Ar/O_2$ Plasma. *IEEE Transactions on Semiconductor Manufacturing* 37(4):602–614.

[86] Kim T, Lee J, Kim Y, Park H, Hwang H, Kim J, Jung H, Kim D (2023) Thermal improvement of HBM with joint thermal resistance reduction for scaling 12 stacks and beyond. *IEEE 2023 73th Electrnoic Components and Tecnology Conference (ECTC)*, 767–771.

[87] Kim S, Jung H, Lee H, Lim S, Lee J, Yu J, Kim J, Lee S (2023) A method for diagnosing the process chamber prior to plasma processes. *IEEE 2023 Asia-Pacific Microwave Conference (APMC)*, 366–368.

[88] Kuhnle A, May MC, Schäfer L, Lanza G (2022) Explainable Reinforcement Learning in Production Control of Job Shop Manufacturing System. *International Journal of Production Research* 60(19):5812–5834.

[89] Kumagai M, Uchiyama N, Ohmura E, Sugiura R, Atsumi K, Fukumitsu K (2007) Advanced Dicing Technology for Semiconductor Wafer—Stealth Dicing. *IEEE Transactions on Semiconductor Manufacturing* 20(3):259–265.

[90] LeCun Y, Bengio Y, Hinton G (2015) Deep Learning. *Nature* 521:436–444.

[91] Lee G, Kim K (2023) Abnormal chamber Detection in the Etching Process Using Time-Series Data Augmentation and Soft Labeling. *IEEE Sensors Journal* 23 (5):5084–5093.

[92] Lee TW (1998) Independent component analysis. *Independent Component Analysis Theory and Applications*, Springer, Dordrecht, 27–66.

[93] Lee G, Lim H, Wang T, Zhang G, Jeong M (2024) Double bagging Trees With Weighted Sampling for Predictive Maintenance and Management of Etching Equipment. *Journal of Process Control* 135:103175.

[94] Lee H, Baek K, Shin K (2017) Resolving Critical Dimension Drift Over Time in Plasma Etching Through Virtual Metrology-Based Wafer-to-Wafer Control. *Japanese Journal of Applied Physics* 56(6):066502-1-6.

[95] Lee I, Franke JH, Philipsen V, Ronse K, De Gendt S, Hendrickx E (2024) Best Focus Alignment Through Pitch Strategies for Hyper-NA EUV Lithography. *Proceedings of SPIE* 12953:226–244.

[96] Lee TY, Chen PT, Huang CC, Chen HC, Chen LY, Lee PT, Chen FC, Horng RH, Kuo HC (2025) Advances in core Technologies for Semiconductor Manufacturing: Applications and Challenges of Atomic Layer Etching, Neutral Beam Etching and Atomic Layer Deposition. *Nanoscale Advances* 7(10):2796–2817.

[97] Leppilahti L (2023) Plasma corrosion resistant ALD coatings for semiconductor manufacturing process equipment. *2023 34th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–5.

[98] Levinson HJ (2022) High-NA EUV Lithography: Current Status and Outlook for the Future. *Japanese Journal of Applied Physics* 61(SD):SD0803.

[99] Liang A, Ke W (2024) TSMC Now Reportedly Operates Over Half of Global EUVs, Weighs High-NA Adoption. *DIGITIMES Asia*. Accessed May 30, 2026 https://www.digitimes.com/news/a20241112PD204/euv-tsmc-adoption-2023-technology.html.

[100] Liu K, Chen Y, Zhang T, Tian S, Zhang X (2018) A Survey of Run-to-Run Control for Batch Processes. *ISA Transactions* 83:107–125.

[101] Liu Y, Tao H, Zhao D, Lu X (2022) An Investigation on the Total Thickness Variation Control and Optimization in the Wafer Backside Grinding Process. *Materials* 15(12):4230.

[102] Liu T, Matsukuma H, Suzuki A, Sato R, Gao W (2024) An Improved Data Processing Algorithm for Spectrally Resolved Interferometry Using a Femtosecond Laser. *Sensors* 24(9):2869.

[103] Liu T, Wu J, Suzuki A, Sato R, Matsukuma H, Gao W (2023) Improved Algorithms of Data Processing for Dispersive Interferometry Using a Femtosecond Laser. *Sensors* 23(10):4953.

[104] Li X, Liu F, Xie J, Zhang C, Gao R, Lyu C (2023) Research on Metallic Contamination Control of PFA Injection Molded Fittings. *Journal of Applied Polymer Science* 140(21):e53940.

[105] Li Y, Zhang Z, Song Q, Shi H, Hou Y, Yue S, Wang R, Cai S, Zhang Z (2024) Surface Micromorphology and Nanostructures Evolution in Hybrid Laser Processes of Slicing and Polishing Single Crystal 4H-SiC. *Journal of Materials Science & Technology* 184:235–244.

[106] Luo Y, Van Assche A (2023) The Rise of Techno-Geopolitical Uncertainty: Implications of the United States CHIPS and Science Act. *Journal of International Business Studies* 54(8):1423–1440.

[107] Lu C, Lin C, Tang T, Lin C, Chang J, Tsai C, Hsia H, Twu J, Liu C, Wu G, Yee K, Yu D (2024) High bandwidth and energy efficient electrical-optical system integration using coupe technology. *IEEE 2024 74th Electrnoic Components and Technology Conference (ECTC)*, 893–897.

[108] Mahajan R, Sankman R, Patel N, Kim D, Aygun K, Qian Z, Mekonnen Y, Salama I, Sharan S, Iyengar D, Mallik D (2016) Embedded multi-die interconnect bridge (Emib)—a high density, high bandwidth packaging interconnect. *IEEE 2016 66th Electronic Components and Technology Conference (ECTC)*, 557–565.

[109] Maitra V, Su Y, Shi J (2024) Virtual Metrology in Semiconductor Manufacturing: Current Status and Future Prospects. *Expert Systems with Applications* 249:123559.

[110] Markov IL (2014) Limits on Fundamental Limits to Computation. *Nature* 512 (7513):147–154.

[111] Marks M, Cheong K, Hassan Z (2022) A Review of Laser Ablation and Dicing of Si Wafers. *Precision Engineering* 73:377–408.

[112] Maruyama N, Sato K, Suzaki Y, Jimbo S, Yamashita I, Yamamoto K, Yamamoto K, Hiura M, Takabayashi Y (2023) Advances and Applications in Nanoimprint Lithography. Novel Patterning Technologies 2023. *Proceedings of SPIE* 12497:86–97.

[113] May GS, Spanos CJ (2006) *Fundamentals of Semiconductor Manufacturing and Process Control*, Wiley, New Jersy.

[114] Ma ZY, Seiller DG (2016) *Metrology and Diagnostic Techniques for Nanoelectronics*, Jenny Stanford, Singapore.

[115] Ma Z, Salloum R, Keen A, Feng J, Ji H (2022) EUV lithography vacuum system energy and footprint reduction. *IEEE 2022 International Workshop on Advanced Patterning Solutions (IWAPS)* : 1–4.

[116] Ma C, Du Z, Wang X, Zhou P, Zhao Y, Hua Y (2024) Corrosion of Stainless Steels and Corrosion Protection Strategies in the Semiconductor Manufacturing Industry: A Review. *Corrosion Reviews* 42(2):127–161.

[117] Meixner A (2021) Too Much Fab and Test Data, Low Utilization, Semiconductor Engineering. https://semiengineering.com/too-much-fab-and-test-data-low-utilization/; 2021 Accessed May 30, 2026.

[118] Mitani R (2012) Eddy Current Sensor for Endpoint Monitoring of a Chemical-Mechanical Polishing System. *Ebara Report* 234:3–6 https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=201202275322564275.

[119] Miyamoto A, Kawahara T (2019) Automatic Generation of CD-SEM Imaging Sequence Including Optimization of Addressing Region of Interest. 927–935.

[120] Molles J, Popović Z, Shukla P, Peng S (2024) Electromagnetic design of excitation coils for an inductively-coupled plasma etching system. *2024 IEEE Wireless and Microwave Technology Conference (WAMICON)*, 1–4.

[121] Moreau S, Jourdon J, Lhostis S, Bouchu D, Ayoub B, Arnaud L, Fremont H (2022) Hybrid Bonding-Based Interconnects: A Status on the Last Robustness and Reliability Achievements. *ECS Journal of Solid State Science and Technology* 11 (2):024001.

[122] Moss S (2025) OpenAI Announces 'The Stargate Project:' $500bn Over Four Years on AI Infrastructure: A New Company Building Huge Data Centers. *Data Center Dynamics* https://www.datacenterdynamics.com/en/news/openai-announces-the-stargate-project-500bn-over-four-years-on-ai-infrastructure/.

[123] Moyne J (2021) Run-to-Run Control in Semiconductor Manufacturing. in Baillieul J, Samad T, (Eds.) *Encyclopedia of Systems and Control*, Springer, Cham; 2021. https://doi.org/10.1007/978-3-030-44184-5_255.

[124] Müller M (2007) Dynamic Time Warping. *Information Retrieval for Music and Motion*, Springer, 69–84 https://www.doi.org/10.1007/978-3-540-74048-3_4.

[125] Neisser M, Orji NG, Levinson HJ, Celano U, Moyne J, Mashiro S, Wilcox D, Libman S (2024) How Lithography and Metrology are Enabling Yield in the Next Generation of Semiconductor Patterning. *Computer* 57(1):51–58.

[126] Nordson Test & Inspection (2026) Nordson Immersion Optical Inspection AMI AW Series. https://www.nordson.com/en/products/test-and-inspection-products/ami-aw-series; 2026 Accessed May 30, 2026.

[127] Nordson Test & Inspection (2026) WaferSense® Wireless Measurement Devices. https://www.nordson.com/en/divisions/test-and-inspection/our-technologies—wafer-sense; 2026 Accessed May 30, 2026.

[128] Noriki A, Uemura H, Kuwatsuka H, Matsui N, Motoii R, Maeda D, Sugita T, Nakamura F, Suda S, Kurosu T, Amano T (2024) Development of all-photonics-function embedded package substrate using 2.3D rdl interposer for co-packaged optics. *IEEE 2024 74th Electronic Components and Technology Conference (ECTC)*, 96–100.

[129] Nova (2026) Prime Optical Critical Dimension Measurement. https://www.novami.com/nova-product/prism/; 2026 Accessed May 30, 2026.

[130] Oehrlein GS, Brandstadter SM, Bruce RL, Chang JP, DeMott JC, Donnelly VM (2024) Future of Plasma Etching for microelectronics: Challenges and Opportunities. *Journal of Vacuum Science & Technology B* 42(4):041501.

[131] Ohta S, Nakai S, Tsuji H (2007) Optical Endpoint Monitoring System for Oxide-CMP. In: *Proceedings of the General Meeting of the Kanto Branch of the Japan Society of Mechanical Engineers*, 243–247. https://doi.org/10.1299/jsmekanto.2007.13.243.

[132] Okabe T, Somaya K (2023) Development of Ionic Liquid Circulation System in High-Vacuum Chamber for Semiconductor Device Fabrication. *Vacuum* 207:111562.

[133] Omron Automatic (2026) X-ray Inspection AVI VT-X950. https://inspection.omron.eu/en/products/vt-x950; 2026 Accessed May 30, 2026.

[134] Onishi T (2023) Automation of Electron Microscope Sample Preparation. *Microscopy: The Journal of the Quekett Microscopical Club* 58(3):106–110.

[135] Onuki T, Ono R, Suzuki A, Ojima H, Shimizu J, Zhou L (2013) A Thin Silicon Wafer Thickness Measurement System by Optical Reflectometry Scheme Using Fourier Transform Near-Infrared Spectrometer. *Advanced Materials Research* 797:549–554.

[136] Orji N, Badaroglu M, Barnes B, Beitia C, Bunday B, Celano U, Kline R, Neisser M, Obeng Y, Vladar A (2018) Metrology for the Next Generation of Semiconductor Devices. *Nature Electronics* 1(10):532–547.

[137] Park H, Choi J, Kim D, Hong S (2021) Artificial Immune System for Fault Detection and Classification of Semiconductor Equipment. *Electronics* 10(8):944.

[138] Park S, Seong J, Jang Y, Roh H, Kwon J, Lee J, Ryu S, Song J, Roh K, Noh Y, Park Y, Jang Y, Cho T, Yang J, Kim G (2022) Plasma Information-Based Virtual Metrology (PI-VM) and Mass Production Process Control. *The Journal of the Korean Physical Society* 80:647–669.

[139] Park H, Kim J, Cho S, Kim K, Jang S, Choi Y, Lee H (2024) Development of Wafer-Type Plasma Monitoring Sensor With Automated Robot Arm Transfer Capability for Two-Dimensional in Situ Processing Plasma Diagnosis. *Sensors* 24(6):1786.

[140] Paulachan P, Hammer R, Siegert J, Wiesler I, Brunner R (2024) Fast in-Line Failure Analysis of Sub-Micron-Sized Cracks in 3D Interconnect Technologies Utilizing Acoustic Interferometry. *Communication Engineering* 3(1):100.

[141] Perry TS (2018) Move Over, Moore's Law. Make Way for Huang's Law. *IEEE Spectrum* 55(5):7.

[142] Poot M, Van Haren M, Kostić D, Portegies J, Oomen T (2024) Position-Dependent Motion Feedforward Via Gaussian Processes: Applied to Snap and Force Ripple in Semiconductor Equipment. *IEEE Transactions on Control Systems Technology: A Publication of the IEEE Control Systems Society* 32(6):1968–1982.

[143] Qin J, Xia W, Wang J, Chen D, Lu Y, Huo X, Xie B, Chen J (2025) A novel High-Performance Wide-Range Vacuum Sensor Based on a Weak-Coupling Resonator. *Microsystems & Nanoengineering* 11(1):98.

[144] Raffel J, Böhm T, Düsing J, Röhl M, Schilde C, Malshe A, Overmeyer L, Lotz C (2024) Ultrasonic Levitation as a Handling Tool for in-Space Manufacturing Processes. *ASME Journal of Manufacturing Science and Engineering* 146(12):121001.

[145] Rao PK, Bhushan MB, Bukkapatnam ST, Kong Z, Byalal S, Beyca OF, Fields A, Komanduri R (2013) Process-Machine Interaction (PMI) Modeling and Monitoring of Chemical Mechanical Planarization (CMP) Process Using Wireless Vibration Sensors. *IEEE Transactions on Semiconductor Manufacturing* 27(1):1–5.

[146] Rigaku Corporation (2026) XTRAIA MF-3000 Catalogue. https://rigaku.com/products/semiconductor-metrology/xrr-edxrf-and-optical-tools/xtraia-mf-3000; 2026 Accessed May 30, 2026.

[147] Sart C, Estevez R, Fiori V, Lhostis S, Deloffre E, Parry G, Gonella R (2016) Cu/Sio_2 hybrid bonding: finite element modeling and experimental characterization. *IEEE 2015 6th Electronic Systems-Integration Technology Conference (ESTC)*, 1–5.

[148] Schmidt D, Medikonda M, Rizzolo M, Silvestre C, Frougier J, Greene A, Breton M, Cepler A, Ofek J, Kaplan I, Koret R, Turovets I (2023) Spectral Interferometry for Fully Integrated Device Metrology. *Journal of Micro/Nanopatterning, Materials, and Metrology* 22(3):031203.

[149] Schölkopf B, Smola A, Müller K (1997) Kernel Principal Component Analysis. *International Conference on Artificial Neural Networks*, Springer, Berlin, 583–588.

[150] SEMI (2023) SEMI S23—Guide for Conservation of Energy, Utilities and Materials Used by Semiconductor Manufacturing Equipment. *SEMI Standards* .

[151] Shamoun B, Chandramouli M, Liu B, Juday RK, Bucay I, Sowers AT, Abboud FE (2021) Multi-Beam Mask Writer in EUV Era: Challenges and Opportunities. *Novel Patterning Technologies 2021* 11610:64–82.

[152] Shen J, Zhu X, Chen J, Tao P, Wu X (2019) Investigation on the Edge Chipping in Ultrasonic Assisted Sawing of Monocrystalline Silicon. *Micromachines* 10(9):616.

[153] Shih AJ (2023) Multicultural Diversity Workforce and Global Technology Collaboration Empowered Semiconductor Manufacturing Excellence in Taiwan: A Manufacturing Engineer's Perspective. *Journal of Manufacturing Science and Engineering* 145(9):090801.

[154] Shih AJ, Denkena B, Grove T, Curry D, Hocheng H, Tsai HY, Ohmori H, Katahira K, Pei ZJ (2018) Fixed Abrasive Machining of Non-Metallic Materials. *CIRP Annals* 67(2):767–790.

[155] Shim J, Cho S, Kum E, Jeong S (2021) Adaptive Fault Detection Framework for Recipe Transition in Semiconductor Manufacturing. *Computers & Industrial Engineering* 161:107632.

[156] Song J, Lee M, Ryu S, Jang Y, Park S, Kim G (2023) Investigation of Ion-Induced Etch Damages on Trench Surface of $Ge_2Sb_2Te_5$ in High Density Ar/SF_6 Plasma. *Current Applied Physics: The Official Journal of the Korean Physical Society* 45: 105–113.

[157] Sreenivasan SV (2017) Nanoimprint Lithography Steppers for Volume Fabrication of Leading-Edge Semiconductor Integrated Circuits. *Microsystems & Nanoengineering* 3(1):17075.

[158] Stokes K, Clark K, Odetade D, Hardy M, Oppenheimer PG (2024) Advances in Lithographic Techniques for Precision Nanostructure Fabrication in Biomedical Applications. *Discover Nano* 57(1):153.

[159] Strauss M, Arjavac J, Horspool D, Nakahara K, Deeb C, Hobbs C (2012) Automated S/TEM metrology on advanced semiconductor gate structures. *Metrology, Inspection, and Process Control for Microlithography XXVI*, SPIE, 346–357.

[160] SUSS MicroTec (2026) XBC300 Gen2 D2W/W2W Hybrid Bonding Platform. https://www.suss.com/en/products-solutions/wafer-bonder/xbc300-gen2-d2w-w2w; 2026 Accessed May 30, 2026.

[161] Tadokoro T (2011) Spectroscopic Ellipsometry: Fundamentals and Applications. *Journal of the Imaging Society of Japan* 50(5):439–447.

[162] Takada M, Kikuchi K, Wang X, Ito T, Li G, Akihiro I, Oomuro Y (2024) Virtual metrology using transfer learning with domain knowledge. *IEEE 2024 International Symposium on Semiconductor Manufacturing (ISSM)*, 1–4.

[163] Tan HY, Weng W, Rai R, Kang C, Dumas L, Brooks I, Katnani A, Zhong Z, Hakala C, Lu Y, Fretwell J (2018) Advanced industrial S/TEM automation and metrology: boundary of precision. *2018 IEEE 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 131–135.

[164] Tan F, Pan T, Li Z, Chen S (2015) Survey on Run-to-Run Control Algorithms in High-Mix Semiconductor Manufacturing Processes. *IEEE Transactions on Industrial Informatics* 11(6):1435–1444.

[165] Tao H, Zeng Q, Liu Y, Zhao D, Lu X (2023) Influence of Anisotropy on Material Removal and Deformation Mechanism Based on Nanoscratch Tests of Monocrystal Silicon. *Tribology International* 187:105527.

[166] Thadani A, Allen GC (2023) Mapping the Semiconductor Supply Chain. Center for Strategic and International Studies, JSTOR, .

[167] Tian Y, Zhao S, Zhang G, Wang P, Liu S (2025) A Material Removal Rate Prediction Model for High-Shear and Low-Pressure Grinding of Single Crystal Silicon. *Journal of Manufacturing Science and Engineering* 147(6):061011.

[168] Tonoy A, Ahmed M, Khan M (2025) Precision Mechanical Systems in Semiconductor Lithography Equipment Design and Development. *American Journal of Advanced Technology and Engineering Solutions* 1(1):71–97.

[169] Tönshoff HK, Schmieden WV, Inasaki I, König W, Spur G (1990) Abrasive Machining of Silicon. *CIRP Annals* 39(2):621–635.

[170] Trueman C (2024) TSMC Could Account for 24% of Taiwan's Electricity Consumption by 2030, Data Center Dynamics. https://www.datacenterdynamics.com/en/news/tsmc-could-account-for-24-of-taiwans-electricity-consumption-by-2030/ . Accessed May 30, 2026.

[171] TSMC Core Values and Business Philosophy (2026) https://www.tsmc.com/english/abouTSMC/values; 2026 Accessed May 30, 2026.

[172] Tsutsui T, Matsuzawa T (2019) Virtual Metrology model Robustness Against Chamber Condition Variation Using Deep Learning. *IEEE Transactions on Semiconductor Manufacturing* 32(4):428–433.

[173] Van Den Brink M, Yen A, Van Wijnen P, Lercel M, Sluijk B (2022) Holistic Patterning to Advance Semiconductor Manufacturing in the 2020s and Beyond. *IEEE Symposium on VLSI Technology and Circuits* 2022:3–7.

[174] Vogg G, Heidmann T, Brandb S (2015) Scanning Acoustic GHz-Microscopy Versus Conventional SAM for Advanced Assessment of Ball Bond and Metal Interfaces in Microelectronic Devices. *Microelectronics and Reliability* 55:1554–1558.

[175] Wang T, Qiao M, Zhang M, Yang Y, Snoussi H (2020) Data-Driven Prognostic Method Based on Self-Supervised Learning Approaches for Fault Detection. *Journal of Intelligent Manufacturing* 31(7):1611–1619.

[176] Wang T, Xie Y, Jeong Y, Jeong M (2024) Dynamic Sparse PCA: A Dimensional Reduction Method for Sensor Data in Virtual Metrology. *Expert Systems with Applications* 251:123995.

[177] Wang Q, Huang N, Chen Z, Chen X, Cai H, Wu Y (2023) Environmental Data and Facts in the Semiconductor Manufacturing Industry: An Unexpected High Water and Energy Consumption Situation. *Water Cycle* 4:47–54.

[178] Wang P, Wang B, Melkote SN (2020) Modeling and Simulation of phase Transformation and Crack Formation During Scribing of Mono-Crystalline Silicon. *International Journal of Mechanical Sciences* 175:105527.

[179] Wang H, Kang R, Dong Z, Gao S (2025) Ultraprecision Machining for Single-Crystal Silicon Carbide Wafers: State-of-the-Art and Prospectives. *Journal of Advanced Manufacturing Science and Technology* 5:2025010.

[180] Watanabe K (2018) Thickness Control Technology for CMP Process. *Journal of the Japan Society for Precision Engineering* 84(3):239–242.

[181] Wodecki B (2025) Nscale to Invest £2bn in UK Data Centres for AI Computing Expansion. *Capacity* https://capacityglobal.com/news/article-nscale-to-invest-2bn-in-uk-data-centres/.

[182] Wuu J, Mantor M, Loh G, Smith A, Johnson D, Fisher D, Johnson B, Henrion C, Schreiber R, Lucas J, Dussinger S, Tomlinson A, Walker W, Moyer P, Kulkarni D, Ng D, Jung W, Swaminathan R, Naffziger S (2023) Coevolution of chiplet technology and cache architecture for AI and compute. *IEEE 2024 70th International Electro Devices Meeting (IEDM)*, 1–4.

[183] Wu Z, Zhang L, Liu W (2021) Structural Anisotropy Effect on the Nanoscratching of Monocrystalline 6H-Silicon Carbide. *Wear: An International Journal on the Science and Technology of Friction Lubrication and Wear* 34:105125.

[184] Xie Y, Wang T, Jeong Y, Tosyali A, Jeong M (2024) True Sparse PCA for Reducing the Number of Essential Sensors in Virtual Metrology. *International Journal of Production Research* 62(6):2142–2157.

[185] Xie Y, Wang T, Jeong M, Lee G (2024) Multi-Source Adaptive Thresholding Adaboost With Application to Virtual Metrology. *International Journal of Production Research* 62(17):6344–6359.

[186] Xu S, Shi M, Huang R, Zhang C, Li S, Huang C, Zhou N, Mao H (2023) A MEMS Thermopile Pirani Sensor Integrated With Composite Nanoforests for Vacuum Monitoring in Semiconductor Equipment. *IEEE Transactions on Electron Devices* 70(10):5300–5305.

[187] Yan J, Asami T, Harada H, Kuriyagawa T (2009) Fundamental Investigation of Subsurface Damage in Single Crystalline Silicon Caused by Diamond Machining. *Precision Engineering* 33(4):378–386.

[188] Yasuda J, Nomura H, Matsumoto H, Nakayamada N, Yamashita H (2023) Recent Progress and Future of Electron Multi-Beam Mask Writer. *Japanese Journal of Applied Physics* 62(sG):SG0803.

[189] Yasui K, Osaki M, Miyamoto A, Namai H (2020) Three-Dimensional Structure Recognition of Circuit Patterns on semiconductor Devices Using Multiple SEM Images Detected in Different Electron Scattering Angles. *Microelectronics and Reliability* 108:113628.

[190] Yeung HWC, Huang S, Ying X (2023) From fabless to fabs everywhere? semiconductor global value chains in transition. *Global Value Chain Development Report 2023: Resilient and Sustainable GVCS in Turbulent Times*, World Trade Organization, Geneva, Switzerland132–187.

[191] Yin CG, Xiong X, R S, Matsukuma H, Tamiya H, Gao W (2025) Conjugate Differential Self-Calibration of a large-Scale Two-Dimensional Variable-Line-Spacing Grating for an Absolute Optical Encoder Using a Fizeau Phase-Shifting Interferometer. *Measurement Science & Technology* 36(7):075023.

[192] Yin J, Bai Q, Goel S, Zhou P, Zhang B (2021) An Analytical Model to Predict the Depth of Sub-Surface Damage for Grinding of Brittle Materials. *CIRP Journal of Manufacturing Science and Technology* 33:454–464.

[193] Yun H, Kwon H, Yeh S, Lee M, Kim J, Kim Y, Cha S, Kang M, Nam J (2023) High-Speed Wafer Temperature Control Approach of Step Chiller for Semiconductor Manufacturing Equipment. *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, 1–6.

[194] Zeiss (2026) Basics of ZEISS 3D X-ray Microscopy for Semiconductor Package Analysis. www.zeiss.com/semiconductor-microscopy; 2026 Accessed May 30, 2026.

[195] Zhang H, Wang P, Gao X, Gao H, Qi Y (2020) Automated fault detection using convolutional auto encoder and k nearest neighbor rule for semiconductor manufacturing processes. *IEEE 2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)* : 83–87.

[196] Zhang H, Wang P, Gao X, Qi Y, Gao H (2022) Data Visualization and Fault Detection Using Bi-Kernel T-Distributed Stochastic Neighbor Embedding in Semiconductor Etching Processes. *IEEE Transactions on Semiconductor Manufacturing* 35(3):522–531.

[197] Zhang C, Gao X, Li Y, Feng L (2019) Fault detection Strategy Based on Weighted Distance of k Nearest Neighbors for Semiconductor Manufacturing Processes. *IEEE Transactions on Semiconductor Manufacturing* 32(1):75–81.

[198] Zhang C, Peng K, Dong J, Zhang K (2020) Nonlinear Fault Detection Based on Fault-Related Multiphase Principle Polynomial Analysis for Al Stack Etch Process. IFAC-Pap. *Online* 53(2):11860–11865.

[199] Zhang C, Xu T, Li Y (2020) A Novel Fault Detection Scheme Based on Difference in Independent Component for Reliable Process Monitoring: Application on the Semiconductor Manufacturing Processes. *Journal of Chemical Engineering of Japan* 53(7):313–320.

[200] Zhang C, Dai X, Zheng X, Li Y (2020) A Novel monitoring Strategy Combining the Advantages of NPE and GMM. *IEEE Access: Practical Innovations, Open Solutions* 8:82989–82997.

[201] Zhang G, Wang T, Baek J, Jeong M, Seo S, Choi J (2024) Multi-Source Ensemble Method With Random Source Selection for Virtual Metrology. *Annals of Operations Research* 17:1–24.

[202] Zhang X, Zhu G-Y (2025) A Literature Review of Reinforcement Learning Methods Applied to job-Shop Scheduling Problems. *Computers & Operations Research* 175:106929.

[203] Zhang Z, Wang B, Kang R, Zhang B, Guo D (2015) Changes in Surface Layer of Silicon Wafers From Diamond Scratching. *CIRP Annals* 64(1):349–352.

[204] Zhang J, Shang X, He B, Zhang B (2023) Towards Understanding the Crack Suppression Mechanism in Brittle Materials With High Grinding Speed at Different Temperatures. *International Journal of Machine Tools and Manufacture* 193:104088.

[205] Zhang Q, Deng H, Song K (2025) Breaking Physics Limits: How 2nm Chip Technology Powers Next-gen 3D ICs. *Fusion of Multidisciplinary Research, An International Journal* 6(1):685–706.

[206] Zhao W, Li H, Wang S (2023) Energy Performance and Energy Conservation Technologies for High-Tech Cleanrooms: State of the Art and Future Perspectives. *Renewable & Sustainable Energy Reviews* 183:113532.

[207] Zhao B, Zhao P, Liu H, Pan J, Wu J (2023) Investigation on Surface Generation Mechanism of Single-Crystal Silicon in Grinding: Surface Crystal Orientation Effect. *Materials Today Communications* 476:203677.

[208] Zheng Z, Huang K, Lin C, Huang W, Zhang J, Chen X, Xiao J, Xu J (2024) Fundamental Investigation on Damage Evolution and Material Removal Mechanism in Scratching Anisotropic Brittle Material. *Tribology International* 197:109764.

[209] Zhou A, Zhang Y, Ding F, Lian Z, Jin R, Yang Y (2024) Research Progress of Hybrid Bonding Technology for Three-Dimensional Integration. *Microelectronics and Reliability* 155:115372.

[210] Zhu F, Jia X, Li W, Xie M, Li L, Lee J (2023) Cross-Chamber Data Transferability Evaluation for Fault Detection and Classification in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 36(1):68–77.

[211] Zhu F, Jia X, Miller M, Li X, Li F, Wang Y, Lee J (2021) Methodology for Important Sensor Screening for fault Detection and Classification in Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing* 34(1):65–73.