

Review Article

The Real Bottlenecks of the AI Semiconductor Race: Advanced Manufacturing, HBM, and Critical Mineral Supply Chains

Jenny Kim*

Deerfield Academy, United States

Accepted: July 29, 2026 **Published:** August 25, 2026

Correspondence to: Jenny Kim, Deerfield Academy, United States

Citation: Kim J. The Real Bottlenecks of the AI Semiconductor Race: Advanced Manufacturing, HBM, and Critical Mineral Supply Chains. American Research Journal of Business and Management, Vol 12, no. 1, 2026, 4-13.

Copyright © 2026 Kim J. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The global race to dominate artificial intelligence is widely portrayed as a contest of algorithms, data, and compute power. Yet the most decisive constraints lie not in software nor even chip design, but in three interdependent physical chokepoints: the extreme concentration of advanced semiconductor fabrication, the emerging scarcity of High Bandwidth Memory (HBM), and the geopolitical fragility of critical mineral supply chains. This paper examines each bottleneck in turn—its technical foundations, its current market structure, and the policy responses it has triggered—arguing that control over these physical layers will determine which nations and corporations lead the AI era. Evidence drawn from market data, IEA commodity reports, CHIPS Act filings, and industry analyses reveals that while governments are accelerating investment to diversify supply chains, the structural concentration of advanced manufacturing in Taiwan, HBM production in South Korea, and mineral refining in China represents a vulnerability that cannot be resolved within a decade, regardless of political will or capital outlay. A June 2026 update to this analysis, incorporating the latest SEMI and Boston Consulting Group market intelligence, finds that this concentration is intensifying rather than easing global fab investment is concentrating further in Korea and the United States, DRAM capacity is being reallocated toward HBM at the expense of legacy nodes, and the packaging chokepoint that defined the 2023–2024 GPU shortage has shifted upstream to wafer fabrication capacity itself.

Keywords: Artificial Intelligence; Semiconductor Supply Chain; High Bandwidth Memory; TSMC; CHIPS Act; Critical Minerals; CoWoS Packaging; Geopolitics of Technology.

INTRODUCTION

When most people think about the AI race, they think about ChatGPT, Gemini, or Grok—about parameters, training runs, and which model produces the most convincing essay. This framing is understandable but misleading. Behind every language model, every image generator, every recommendation system lies a stack of physical infrastructure so specialized and so geographically concentrated that it represents one of the most brittle supply chains in the history of industrial capitalism.

The semiconductor industry is the foundation of this stack. A modern AI accelerator—Nvidia's H100, for example—contains tens of billions of transistors manufactured at geometries measured in nanometers, stacked with memory

chips that must transfer data at terabyte-per-second speeds, assembled in packaging processes requiring equipment that exists in perhaps a dozen factories on Earth. The entire chain from mineral ore to finished GPU spans dozens of countries and companies, but its most critical segments are controlled by a tiny number of actors operating at the technological frontier.

Three bottlenecks define this frontier. First, the fabrication of leading-edge logic chips—those at 3 nm and below—is effectively monopolized by TSMC, which produces roughly 90% of the world's most advanced processors [1]. Second, the High Bandwidth Memory (HBM) chips that make AI accelerators functional are manufactured almost exclusively by two South Korean firms, SK Hynix and Samsung, whose combined dominance exceeds 90%

of the market [1]. Third, the raw and refined materials that feed both industries—gallium, germanium, rare earth elements, cobalt—are mined and processed predominantly in nations that are either geopolitical rivals of the West or deeply entangled with China’s industrial policy [2].

This paper argues that understanding these three bottlenecks is essential for any serious analysis of the AI landscape. Section II examines the physics and economics of advanced semiconductor fabrication and why replicating TSMC’s capabilities is a generational challenge [3]. Section III analyzes HBM’s role in AI systems, including a generational overview of specifications and the tight oligopoly that controls its production. Section IV surveys the critical mineral dependencies that underpin the entire chain. Section V assesses current policy responses—chiefly the U.S. CHIPS Act—and their realistic prospects. Section VI concludes with an assessment of which constraints are most likely to shape the AI race through 2030 [4].

THE FABRICATION BOTTLENECK: TSMC AND THE LIMITS OF THE POSSIBLE

The Physics of Scaling

A semiconductor chip is, at its most basic, a device that controls the flow of electricity through extremely small switches called transistors. The history of computing is largely the history of making these transistors smaller, faster, and more energy-efficient trend captured by Moore’s Law, the observation that transistor density doubles approximately every two years. But as transistors have shrunk from micrometers to nanometers, the engineering required to manufacture them has become exponentially more complex.

Today, the most advanced processors use transistors with gate lengths of 3 nm—roughly the width of ten silicon atoms. Manufacturing at this scale requires a process called Extreme Ultraviolet (EUV) lithography, in which patterns are etched onto silicon wafers using light with a wavelength of 13.5 nm. The only company in the world that builds

EUV machines is ASML, a Dutch firm headquartered in Eindhoven [3]. A single EUV machine costs approximately \$200 million and requires years of lead time to build, with components sourced from over 800 suppliers across 24 countries [3].

The sophistication of EUV lithography illustrates a broader principle: cutting-edge semiconductor manufacturing is not merely expensive; it is extraordinarily knowledge intensive. The tacit knowledge embedded in TSMC’s process engineering, yield management, and manufacturing operations has been accumulated over 35 years and cannot be replicated simply by purchasing the same equipment. As TSMC’s founder Morris Chang has noted, the gap between owning a tool and mastering it is measured in decades, not dollars [4].

TSMC’s Near-Monopoly and Its Implications

TSMC was founded in 1987 and pioneered the “pure-play foundry” model, in which the company manufactures chips designed by others rather than developing its own products. By 2024, TSMC commanded over 60% of global foundry revenue and produced approximately 90% of the world’s most advanced chips at the 5 nm node and below [5]. Advanced technologies at 7 nm and below accounted for 74% of TSMC’s wafer revenue in Q2 2025, underscoring its dominance in the segments most critical to AI [6].

For the AI semiconductor industry, this concentration has profound implications. Every major AI chip—Nvidia’s H100 and B200, AMD’s MI300X, Google’s TPU v5, Apple’s M-series—is manufactured by TSMC or relies on TSMC for its leading-edge components. When demand surges, as it did after the launch of ChatGPT in late 2022, TSMC’s capacity constraints become a binding supply limit on the entire AI industry. Nvidia’s ability to ship GPUs is not limited by its chip designs or customer demand; it is limited by TSMC’s CoWoS (Chip-on-Wafer-on-Substrate) advanced packaging capacity constraint that persisted as the primary bottleneck through the Hopper and early Blackwell GPU generations [7].

Table 1: Global semiconductor foundry market share (revenue, 2024).

Company	Country	Revenue Share	Leading Node	AI Relevance
TSMC	Taiwan	~61%	2 nm (2025)	Sole manufacturer of Nvidia, AMD, Apple, Google AI chips
Samsung Foundry	S. Korea	~11%	3 nm (GAA)	Qualcomm, Tesla; emerging AI ASIC customer base
GlobalFoundries	USA/Sing.	~6%	12 nm	Mature-node chips; defense & automotive focus
UMC	Taiwan	~6%	22 nm	Mature-node; automotive, IoT
SMIC	China	~6%	7 nm (DUV)	Domestic China AI chips; restricted from EUV access
Intel Foundry	USA	~2%	18A (2025E)	CHIPS Act recipient; ramping advanced manufacturing

Sources: TSMC Annual Report 2024 [27]; Quartr (2025) [16]; Silicon Semiconductor (2024) [23].

The geopolitical dimension amplifies these concerns. Taiwan, where virtually all TSMC’s leading-edge capacity resides, sits 160 km from the Chinese mainland. Bloomberg Economics has estimated that a conflict over Taiwan could cost the global economy \$10 trillion in its first year [7]. Even short of conflict, the concentration of global chip production on a single island creates systemic vulnerability to natural disasters, supply disruptions, or political crises.

ASML’s Monopoly Within the Monopoly

If TSMC’s dominance seems remarkable, ASML’s is even more so. Every chip manufactured at advanced nodes anywhere in the world depends on lithography equipment, and the most advanced tier—EUV—is produced exclusively by ASML. The Netherlands has been drawn into the geopolitical contest over semiconductors as a result, facing pressure from both the United States and China over export licensing for its machines [8]. In 2023, the Dutch government revoked ASML’s license to export its most advanced EUV systems to China, reinforcing the export control architecture the United States had been building since 2022.

China has responded with significant state investment in domestic lithography alternatives, led by Shanghai Semiconductor Equipment Company (SMEE). A 2023 patent filing from SMEE related to EUV indicated stepped-up domestic R&D, and Huawei has reportedly been recruiting engineers with EUV expertise from ASML’s supplier network [9]. However, analysts assess that Chinese domestic EUV development lags ASML by at least a decade—the system-level integration challenges cannot be overcome through reverse engineering alone [10].

The 2026 Fab Investment Surge: A Widening, Not Narrowing, Race

Data presented at SEMI’s Market Intelligence Forum in June 2026 confirms that concentration in advanced fabrication has not eased since 2024—it has deepened, even as the pool of capital chasing that concentration has grown dramatically. SEMI’s Q2 2026 forecast revised the global semiconductor market outlook upward from an already-aggressive \$1 trillion figure to approximately \$1.5 trillion by 2030, an upgrade driven almost entirely by AI infrastructure spending; the 2026 forecast alone was revised up by \$263 billion, a 35% increase over the prior-quarter estimate [11]. Front-end fab equipment spending, tracked by SEMI’s World Fab Forecast, was revised upward across every year from 2026 through 2030, with the industry now projected to spend approximately \$200 billion on wafer fab equipment in 2029—up from \$154

billion in the prior (4Q 2025) forecast, a 30% upward revision for that year alone [12].

This capital is not spreading evenly. Table 2 summarizes SEMI’s regional fab investment forecast for 2025 versus 2029, and it shows the same concentration dynamic documented in Section II-B repeating itself at the country level: America and Korea are projected to roughly double their fab investment by 2029, China’s investment is projected to contract slightly from its elevated 2025 base, and Taiwan—home to TSMC’s leading-edge capacity—remains a top-tier investor even as its share of total incremental spending narrows relative to Korea and the United States [13].

Table 2: regional semiconductor fab investment forecast, 2025 vs. 2029 (us\$ billion)

Region	2025	2029F	Change
America	\$21.5B	\$50B	+133%
Korea	\$27.7B	\$50B	+81%
China	\$57B	\$45B	–21%
Taiwan	\$32B	\$41B	+28%
Japan	\$11.6B	\$19.5B	+68%
Europe & Middle East	\$5.4B	\$14B	+159%
Southeast Asia	\$6B	\$8B	+33%

Source: SEMI World Fab Forecast Report, 2Q 2026 [12].

The regional shift is most striking for Korea, whose fab investment forecast rises from \$27.7 billion in 2025 to \$50 billion by 2029—propelling it, alongside the United States, into the top tier of global fab investors for the first time [12]. This is driven almost entirely by AI-related memory: SEMI data shows Korean-headquartered device makers’ global fab investment (equipment plus facility, across all their fabs worldwide) climbing from roughly \$20 billion in 2020 to a projected \$60 billion by 2029, with the investment mix shifting decisively toward DRAM and NAND at the expense of logic and foundry spending as HBM and AI-driven memory demand crowd out other priorities [13]. This dynamic is examined in detail in Section III-E below, and it reinforces—rather than undercuts—this paper’s central argument: capital is now chasing the same handful of concentrated chokepoints even more intensely than before.

THE MEMORY BOTTLENECK: HIGH BANDWIDTH MEMORY AND THE AI ACCELERATOR STACK

Why AI Systems Are Starved for Memory Bandwidth

The dominant computational task in AI—matrix multiplication—requires moving enormous quantities of data between memory and processing cores at very

high speed. A conventional personal computer’s CPU communicates with its RAM through a memory bus that delivers roughly 50 GB/s. This is adequate for general-purpose computing but catastrophically insufficient for training large neural networks. An Nvidia H100 GPU, running with six stacks of HBM3 memory, achieves approximately 3.35 TB/s of memory bandwidth—roughly 65 times the bandwidth of a desktop DDR5 system [14].

HBM achieves this performance through 3D stacking: multiple DRAM dies are stacked vertically and connected by through-silicon vias (TSVs) microscopic conductive pathways that pass through the silicon itself. The resulting stack is then mounted directly adjacent to the GPU die on a shared silicon interposer using TSMC’s CoWoS packaging technology. The memory wall—the growing gap between processor speed and memory bandwidth—has become the central engineering challenge of the AI era [15]. Table 3 summarizes HBM’s generational evolution and the AI accelerators that depend on each generation.

HBM now accounts for approximately half of an AI chip’s total production cost [1], a figure that reflects both the manufacturing complexity of the memory stack and the tight supply that follows from oligopolistic production. As GPU compute scales faster than HBM bandwidth, AI system designers face a widening memory wall that constrains effective utilization of even the most powerful accelerators [16].

The HBM Oligopoly: SK Hynix, Samsung, and Micron

As of mid-2025, three companies produce essentially

all the world’s HBM: SK Hynix of South Korea, Samsung Electronics of South Korea, and Micron Technology of the United States. Table 4 presents market share trends across the three vendors from 2024 through mid-2025, alongside projected market size.

SK Hynix’s position has been so commanding that by late 2025, the company announced that its entire production capacity for DRAM, NAND, and HBM through 2026 was fully sold out, largely committed to Nvidia [17]. Goldman Sachs has assessed that SK Hynix will maintain a total HBM market share of over 50% through at least 2026 [18]. Samsung, despite its scale, lost market share in 2024–2025 because its HBM3E products failed to meet Nvidia’s yield and reliability standards. For Nvidia’s next-generation Rubin platform, industry sources indicate that SK Hynix is expected to supply approximately two-thirds of HBM4 requirements [19].

CoWoS: The Packaging Chokepoint

The integration of HBM and GPU dies onto a common silicon interposer—TSMC’s CoWoS process—emerged as a critical bottleneck during the 2023–2024 AI GPU demand surge. CoWoS capacity is entirely separate from TSMC’s wafer fabrication capacity and follows the same multi-year expansion lead times. During the peak of Nvidia’s Hopper GPU demand, CoWoS became the binding supply constraint—not the N4 wafer process, not SK Hynix’s HBM production, but the packaging step that joins them into a shippable product [20].

TSMC has been expanding CoWoS capacity aggressively,

Table 3: HBM generation specifications and ai accelerator deployment.

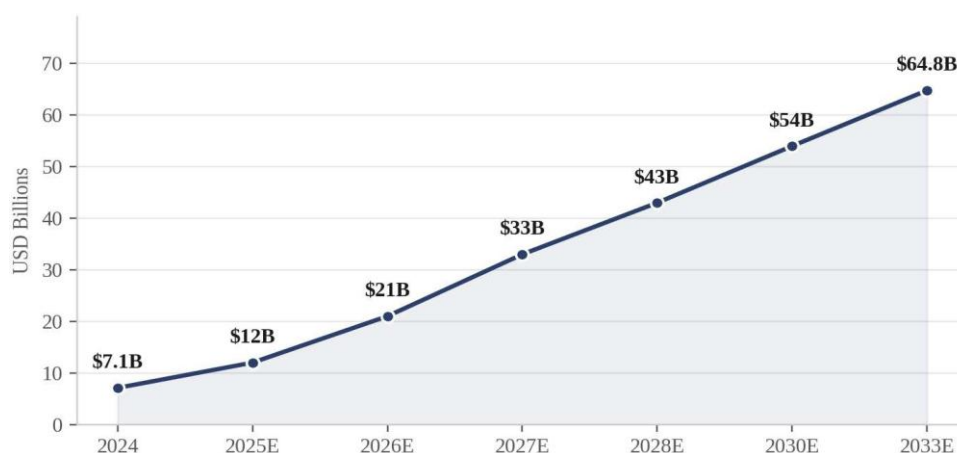
Generation	Year	Bandwidth (per stack)	Capacity (per stack)	Stack Height	AI Accelerators Using
HBM2	2016	256 GB/s	8 GB	4-Hi	NVIDIA V100, AMD MI60
HBM2E	2018	461 GB/s	16 GB	8-Hi	NVIDIA A100, AMD MI250X
HBM3	2022	819 GB/s	24 GB	12-Hi	NVIDIA H100, AMD MI300X
HBM3E	2024	1,180 GB/s	36 GB	12-Hi	NVIDIA H200, B200; Google TPU v5p
HBM4	2026E	~1,500 GB/s	48 GB	16-Hi	NVIDIA Rubin, next-gen ASICs

Sources: SemiconductorX (2024) [22]; Silicon Analysts (2026); TrendForce (2026) [26].

Table 4: HBM market share by vendor (shipment volume, %).

Vendor	Q2 2024	Q4 2024	Q2 2025	Status / Notes
SK Hynix	~50%	~55%	62%	Dominant NVIDIA supplier; HBM4 ~70% share (est.)
Samsung	~40%	~28%	17%	Failed NVIDIA HBM3E qual. tests; recovering 2026
Micron	~10%	~17%	21%	Only U.S.-based HBM maker; share rising rapidly
Market Size	\$7.1B (2024)	—	\$43B (2027E)	CAGR ~28.4% through 2033 (projected)

Sources: Counterpoint Research via Astute Group (2025) [4]; SK Hynix Newsroom (2026) [24]; Notebookcheck (2025) [14]; Citi Research (2025).



Sources: Probity Market Insights (2024); Citi Research (2025); SK Hynix Newsroom (2026). [1],[14],[24]

Figure 1: Global HBM Market Size Growth, 2024–2033 (Projected, USD Billions).

reaching 75,000 wafers per month in 2025, but even this expansion has struggled to satisfy the simultaneous ramp of Nvidia’s Blackwell architecture and the growing custom ASIC programs of Google, Amazon, and Microsoft [21]. The packaging bottleneck will persist as a structural constraint through at least 2027.

A mid-2026 assessment from Boston Consulting Group offers an important nuance to this picture: CoWoS and HBM packaging shortages that constrained the 2023–2024 GPU ramp were, in BCG’s assessment, largely resolved during 2024, and the primary bottleneck for AI chip supply has since shifted upstream to wafer fabrication capacity itself—particularly the wafer-constrained supply of leading-edge logic and DRAM dies feeding both GPU and ASIC production [22]. This does not eliminate CoWoS and HBM as constraints; it relocates the tightest link in the chain rather than removing it, and as Section III-E details, HBM’s disproportionate claim on DRAM wafer capacity has, if anything, intensified through 2026.

HBM’s Macroeconomic Effects

The scarcity of HBM and CoWoS capacity has macroeconomic ripple effects extending well beyond the AI industry. Because HBM commands a 5–6× price premium over equivalent DDR5 capacity, memory manufacturers have systematically redirected wafer capacity from commodity DRAM to HBM. Although HBM accounted for only around 8% of DRAM output in 2025, it generated over 30% of total DRAM revenue [23]. The resulting squeeze pushed DDR5 contract prices up by nearly 50% in late 2025, raising costs for PC manufacturers, server vendors, and consumers globally [24–30].

2026 Update: Korea’s Memory-Led Capacity Buildout and the DRAM/HBM Squeeze

SEMI’s June 2026 Market Outlook provides the most detailed public accounting to date of how HBM demand is reshaping memory capacity planning, and it confirms that the memory bottleneck described above is intensifying rather than resolving. DRAM wafer capacity is now projected to grow at a 7% compound annual rate from 2025 to 2030, up from the 4.8% projected only two quarters earlier, as SK Hynix, Samsung, and Micron accelerate expansion in response to AI demand [18]. Yet this capacity growth does not translate into relief for conventional memory buyers: nodes of 15 nm and below—the nodes required to serve HBM and DDR5 production—are projected to account for approximately 80% of all DRAM capacity by 2030, up sharply from their current share, while capacity at 16 nm and above is progressively converted or retired [18]. In effect, HBM’s wafer requirements are crowding out legacy DRAM capacity at the process-node level, not merely at the margin, reinforcing the memory wall dynamic and DDR5 price pressure discussed in Section III-D.

3D NAND capacity is following a parallel, if less acute, trajectory: wafer capacity is now projected to grow at a 5% CAGR from 2025 to 2030 (up from 4% in the prior forecast), driven by AI-related datacenter SSD and storage demand, with approximately 90% of NAND wafers expected to sit on greater-than-400-layer technology by 2030 as investment shifts from adding cleanroom capacity toward layer migration and yield improvement [18].

SEMI’s analysis also reframes where in the supply chain the next constraint is likely to bind. Rather than a

single chokepoint, SEMI describes a chain of sequential bottlenecks running from front-end HBM and DRAM wafer allocation, through advanced packaging steps (through-silicon-via formation and thermal compression bonding), to test and burn-in throughput, and finally to substrate, PCB, power, and cooling readiness at the system level. Constraints can and do shift between these links: as HBM allocation tightens supply for conventional DRAM, TCB and TSV bonding yield becomes the pacing item further downstream, and system-level test, burn-in, and power-delivery readiness can delay qualified output even after wafers and packaging are complete [18]. This sequential-bottleneck framing is consistent with, but more granular than, the CoWoS-centered account in Section III-C, and it underscores that Korea’s HBM leadership depends on execution across the entire chain—memory wafers, packaging yield, test capacity, and system-level readiness—not on wafer capacity alone [18].

Table 5 summarizes Korean device makers’ global fab investment step-up and its shift toward memory, drawing on the same SEMI World Fab Forecast dataset used in Table 5.

THE MATERIALS BOTTLENECK: CRITICAL MINERALS AND CHINA’S REFINING DOMINANCE

What Are Critical Minerals and Why Do They Matter for Semiconductors?

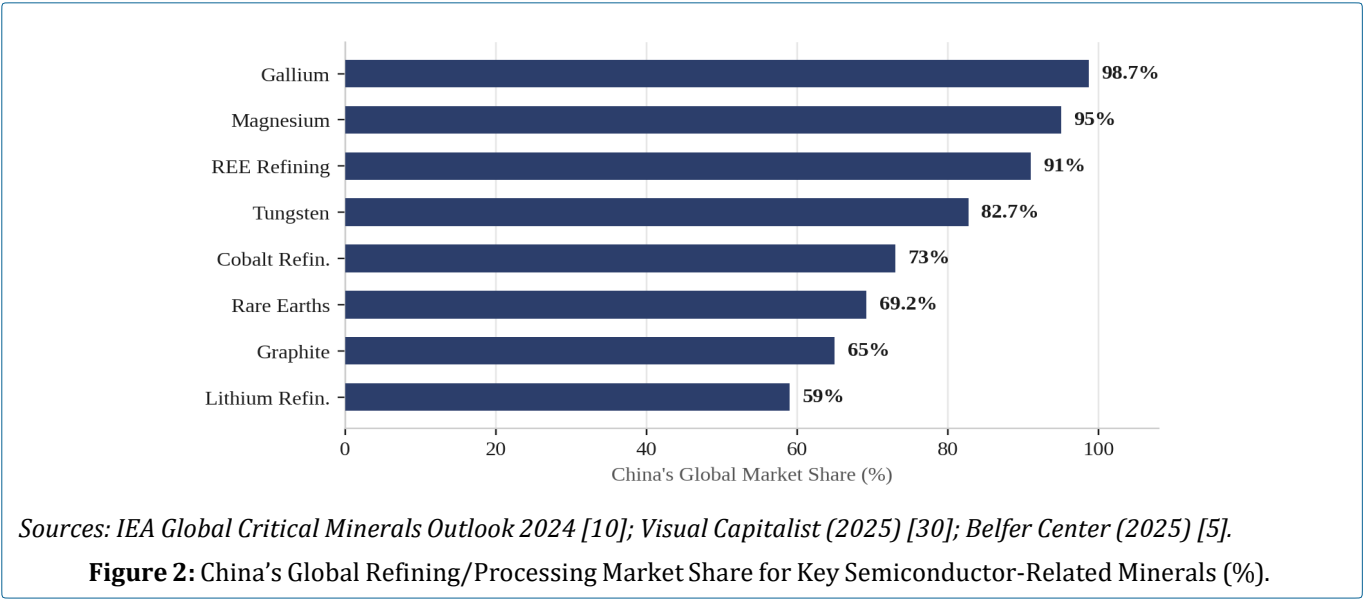
Semiconductors are made of much more than silicon. A leading-edge chip requires dozens of specialty materials: ultra-pure silicon, germanium for transistor channels, gallium for compound semiconductors, cobalt for barrier layers in interconnects, rare earth elements for chemical mechanical planarization slurries, and tungsten for via fills. Each of these materials has a supply chain that, in most cases, flows through China.

The International Energy Agency’s 2024 Critical Minerals Market Review found that China maintains leading refining positions across 19 of 20 strategic minerals, with average market share concentrations of approximately 70% in key materials including lithium, cobalt, graphite, and rare earth elements [10]. Figure 2 illustrates the extent of China’s processing dominance across the minerals most relevant to semiconductor manufacturing.

Table 5: Korean device makers’ global fab investment and dram/nand capacity outlook.

Metric	2020 / 2025	2029 / 2030F	Notes
Korean device makers’ global fab investment	\$20B (2020)	\$60B (2029F)	Equipment + facility, global basis
DRAM capacity CAGR (2025–2030)	4.8% (prior forecast)	7% (revised)	Revised upward on AI/HBM demand
≤15 nm share of DRAM capacity	n/a	~80% (2030F)	Serves HBM and DDR5; crowds out legacy nodes
3D NAND capacity CAGR (2025–2030)	4% (prior forecast)	5% (revised)	AI datacenter SSD demand
NAND wafers on >400-layer tech	n/a	~90% (2030F)	Investment shifts to layer migration, yield

Source: SEMI World Fab Forecast, 2Q 2026, as presented in SEMI Market Outlook, Jun. 22, 2026 [18].



For rare earth elements—essential for permanent magnets in data center cooling systems and fab equipment—China accounts for approximately 91% of global separation and refining capacity [11]. China’s dominance in mineral processing is, as the Belfer Center at Harvard notes, “built on scale and control across the supply chain rather than on geological abundance” [5]. Table 6 provides a detailed breakdown of export control status and semiconductor applications for the six most critical materials.

Gallium, Germanium, and the Weaponization of Supply

The strategic significance of mineral control became explicit in December 2024, when China imposed comprehensive export bans on gallium, germanium, and antimony shipments to the United States [11]. China produces approximately 98.7% of the world’s gallium and dominates germanium production as well [30]. China’s 2024 restrictions on rare earth magnet exports, covering 58,000 tonnes annually—enough to manufacture components for millions of vehicles or thousands of data centers—demonstrated the breadth of China’s leverage [11]. As the U.S.-China Economic and Security Review Commission documented in its 2025 Annual Report, Chinese authorities appear “increasingly willing to use strategic material dominance as a foreign policy tool, particularly in response to technology restrictions and trade tensions” [29].

China’s Strategy: Processing Power Over Mining Rights

A critical distinction in analyzing China’s mineral dominance is between mining and processing. For lithium, China’s share of raw extraction is only around 10–30%—Australia now leads global lithium output with 36.7% [30]. But China’s share of lithium refining reaches approximately 59% [9]. The same pattern holds for cobalt, where China

refines approximately 73% of global supply despite not being the leading mine producer. A mine can be opened in a few years; a refinery with the technical capability to produce semiconductor-grade materials from complex ores requires a decade or more to develop [5].

The Council on Foreign Relations assessed in early 2026 that China’s near-total control over many critical minerals—and its demonstrated willingness to weaponize that dominance—has “put the world on notice” [8]. Even the most optimistic projections for supply chain diversification concede that “supply chains don’t move quickly and can’t be built quickly” and that “the main bottlenecks remain minerals mined outside China are almost always sent there for refining, processing, and manufacturing” [25].

HBM Export Controls and Smuggling

The intersection of HBM scarcity and geopolitical competition has produced organized smuggling of AI memory chips. Following the December 2024 U.S. export controls on HBM to China, several Taiwanese companies reportedly exploited a loophole by loosely attaching Samsung HBM to very simple processor chips, designing these assemblies to appear compliant with export restrictions while allowing the HBM to be extracted after delivery to China [1]. The structural incentive to smuggle Western HBM stacks is high, given their far superior performance relative to domestically produced alternatives. Effective enforcement requires physical supply chain verification at a level of granularity that existing customs infrastructure struggles to provide [1].

POLICY RESPONSES: THE CHIPS ACT AND ITS LIMITS

The CHIPS and Science Act: Rationale and Scale

The United States passed the CHIPS and Science Act

Table 6: China’s dominance in critical minerals for semiconductor manufacturing.

Mineral	China Mining Share	China Refining Share	Export Control?	Semiconductor Use
Gallium	~80%	98.7%	Yes (Dec. 2024)	GaN/GaAs chips, RF semiconductors, LEDs
Germanium	~60%	~80%	Yes (Dec. 2024)	Transistor channels, fiber optics, IR optics
Rare Earths	60%	91%	Magnets (2025)	CMP slurries, magnets in fab equipment/data centers
Cobalt	~3%	73%	No	Barrier layers in chip interconnects; batteries
Graphite	~65%	~65%	Sep. 2023	Electrodes in semiconductor manufacturing equipment
Antimony	~48%	~50%	Yes (Dec. 2024)	Flame retardants, certain semiconductor processes

Sources: IEA Critical Minerals Outlook 2024 [10]; IEA Commentary (2025) [11]; Visual Capitalist (2025) [30]; USCC Annual Report (2025) [29].

in August 2022, directing approximately \$52.7 billion in direct grants and \$24 billion in manufacturing tax credits toward rebuilding domestic semiconductor manufacturing capacity. The law’s passage reflected a bipartisan recognition that the United States had allowed its share of global semiconductor manufacturing to fall from approximately 37% in 1990 to around 10% by 2022 [7]. The Peterson Institute for International Economics estimated that by 2024, the Act had catalyzed approximately \$158 billion in realized construction investment, with announced investments exceeding \$500 billion [15]. The Semiconductor Industry Association projected that U.S.-based semiconductor manufacturing capacity would increase by 203% from 2022 to 2032—though even this growth would raise the U.S. share of global fabrication only from approximately 10% to 14% [20].

TSMC Arizona: Promise and Complications

TSMC announced a non-binding memorandum for up to \$6.6 billion in direct funding in April 2024 [27] and subsequently announced total capital expenditure at the Phoenix site exceeding \$165 billion, including three fabrication plants and two advanced packaging facilities. TSMC Arizona’s first fab is on track to begin production at 4 nm in the first half of 2025, with a second fab producing 3 nm and 2 nm technologies beginning in 2028 [27].

The Arizona expansion has faced significant practical challenges. Multiple TSMC employees cited deep workplace cultural differences between Taiwanese and American engineers as a factor in construction delays. Building costs

in the United States run approximately 30% higher than in Taiwan, South Korea, or Singapore, and 37–50% higher than in China [21]. Even at full capacity, TSMC Arizona’s three fabs will contribute only about 10% of TSMC’s total silicon output [7], meaning the Arizona fabs reduce—but do not eliminate—the concentration risk associated with Taiwan’s geopolitical exposure.

CHBM and Memory: SK Hynix’s U.S. Investment

The CHIPS Act has also reached into the HBM supply chain. SK Hynix committed \$3.87 billion between August 2024 and January 2026 to build a next-generation HBM packaging facility at Purdue University Research Park in Indiana [17]. Micron Technology, the only U.S.-headquartered memory manufacturer, received a CHIPS Act grant of \$6.1 billion toward new manufacturing facilities in Clay, New York and Boise, Idaho, aimed at onshoring approximately 40% of Micron’s DRAM production over two decades [13].

Critical Minerals: The Harder Problem

Policy responses to the critical minerals bottleneck have been less coherent and face more structural obstacles. The Trump administration’s 2025 executive actions targeted the expansion of U.S. critical mineral supplies through a national energy emergency declaration and use of the Defense Production Act, while a Section 232 investigation into national security risks from critical mineral import dependence was initiated [29]. However, as the CFR concluded in February 2026, “the United States cannot out-mine and out-process China” through conventional expansion alone; instead, it should focus on

Table 7: major chips and science act awards (direct funding, finalized 2024–2025).

Recipient	Location(s)	Award Amount	Node / Type	Strategic Significance
TSMC	Arizona	\$6.6B	4 / 3 / 2 nm	First leading-edge logic fabs on U.S. soil; \$165B total investment
Intel	OH, AZ, NM, OR	\$7.9B	18A / mature	Largest award; Intel Foundry Services ramp for defense & AI
Micron	NY, ID	\$6.1B	DRAM / HBM	~40% of Micron DRAM onshored; supports domestic HBM capacity
Samsung	Texas	\$4.7B	4 nm / 2 nm	Diversifies advanced logic beyond Taiwan; R&D center included
SK Hynix*	Indiana	\$3.87B†	HBM pkg.	HBM advanced packaging at Purdue Research Park; opens 2028
GlobalFoundries	NY, VT	\$1.5B	Mature node	Essential chips for automotive, defense, aerospace

** SK Hynix investment is private capital, not a direct CHIPS grant. † Total private investment commitment, Aug. 2024–Jan. 2026.*

Sources: *Manufacturing Dive CHIPS Act Tracker (2024)* [13]; *TSMC Press Release (2024)* [27]; *Raise Summit (2026)* [17]; *PIIE (2025)* [15].

“scaling disruptive innovation, recovery, and recycling” to leapfrog China’s dominance [8]. Recycling of end-of-life semiconductors and batteries offers a pathway to recovering gallium, rare earths, and cobalt that bypasses the Chinese refining bottleneck—but the infrastructure for doing so at scale does not yet exist in the United States or Europe.

CONCLUSION

The AI semiconductor race is fundamentally a contest over physical infrastructure: who controls the nanoscale manufacturing processes that etch circuits onto silicon, who controls the memory architectures that allow those circuits to function at AI-relevant scales, and who controls the minerals and processing capacity that feed the entire chain. In each of these domains, concentration is the defining feature: TSMC in advanced logic, SK Hynix in HBM, and China in critical mineral refining.

The policy responses to this concentration—the CHIPS Act, the SK Hynix Indiana investment, the ASML export controls, the gallium and germanium bans and counter-bans—represent a recognition that the AI era’s industrial geography has become a strategic matter. But the timelines involved are sobering. A new semiconductor fabrication plant takes three to five years to build and several more to achieve competitive yields. A new mineral refinery requires a decade or more to reach full production. The knowledge embedded in TSMC’s process engineering, SK Hynix’s HBM yield management, and China’s rare earth processing has been accumulated over decades and cannot be transferred or replicated on a policy timeline.

The most immediate constraint through 2027 is likely to remain HBM and CoWoS packaging capacity. The memory wall is widening—GPU compute is scaling faster than HBM bandwidth—and the tight oligopoly of HBM producers, combined with TSMC’s exclusive control of CoWoS, means that supply cannot rapidly expand to meet demand. Over a five-to-ten-year horizon, the critical mineral supply chain poses the deepest structural challenge. China’s demonstrated willingness to impose export bans on gallium, germanium, and rare earth elements—materials with no near-term substitutes in semiconductor manufacturing—gives Beijing a source of leverage that cannot be neutralized quickly. The West’s best response is a combination of accelerated mining development in allied nations, investment in refining capacity, and technology development aimed at reducing or substituting for the most exposed mineral dependencies.

A June 2026 update to this assessment, drawing on newly available SEMI and BCG market intelligence, sharpens rather than revises this conclusion. The industry’s own forecasters have raised their outlook for AI-driven semiconductor demand even further—to roughly \$1.5 trillion by 2030—while simultaneously confirming that the packaging bottleneck of 2023–2024 has migrated upstream to wafer capacity, that DRAM capacity is being structurally reallocated toward HBM at the expense of legacy nodes, and that Korea’s role has expanded from HBM supplier to one of the two fastest-growing centers of global fab investment alongside the United States. None of this loosens the underlying concentration; it relocates and, in the case of memory, deepens it.

For students and citizens trying to make sense of AI’s geopolitical dimensions, the key insight is this: the decisive constraints on who leads the AI era are not primarily algorithmic or intellectual. They are physical, geographic, and industrial rooted in the same dynamics of geography, resource control, and manufacturing know-how that have shaped great power competition throughout history. The silicon atom, it turns out, is no more exempt from geopolitics than the oil barrel or the iron ingot.

References

1. AI Frontiers, “High-Bandwidth Memory: The Critical Gaps in US Export Controls,” AI Frontiers, Feb. 2026.
2. American Affairs Journal, “The Evolution of China’s Semiconductor Industry under U.S. Export Controls,” American Affairs Journal, vol. VIII, no. 4, Nov. 2024.
3. ASML Holding N.V., Annual Report 2024. Veldhoven, Netherlands: ASML, 2024.
4. Astute Group, “SK hynix Holds 62% of HBM, Micron Overtakes Samsung, 2026 Battle Pivots to HBM4,” Astute Group, Sep. 2025.
5. Belfer Center for Science and International Affairs, Harvard University, “Critical Minerals Explained: Why They Matter for Geopolitics, Clean Energy & Tech,” Oct. 2025.
6. Boston Consulting Group, “Restructuring of the Semiconductor Industry Driven by AI and the Opportunities It Creates,” SEMI MI Forum presentation, BCG, 2026.
7. Council on Foreign Relations, “The CHIPS Act: How U.S. Microchip Factories Could Reshape the Economy,” CFR, Oct. 2024.
8. Council on Foreign Relations, “Leapfrogging China’s Critical Minerals Dominance,” CFR, Feb. 2026.
9. CrossDock Insights, “Critical Minerals Race: China’s Dominance vs. Global Challengers,” CrossDock Insights, Aug. 2025.

10. International Energy Agency (IEA), *Global Critical Minerals Outlook 2024*. Paris: IEA, 2024.
11. International Energy Agency (IEA), "With New Export Controls on Critical Minerals, Supply Concentration Risks Become Reality," IEA Commentary, Apr. 2025.
12. Intelligent Living, "HBM Is Coming for Your PC's RAM: AI Datacenter High-Bandwidth Memory Squeezes Global DDR5 RAM Supply Chain," Intelligent Living, Dec. 2025.
13. Manufacturing Dive, "Tracking CHIPS and Science Act Awards," Manufacturing Dive, 2024.
14. Notebookcheck, "SK hynix Sells Out Its DRAM, NAND, and HBM Chip Supply to Nvidia Through 2026," Notebookcheck, Oct. 2025.
15. Peterson Institute for International Economics (PIIE), "The CHIPS Act Already Puts America First. Scrapping It Would Poison the Well for U.S. Investment," PIIE, Mar. 2025.
16. Quartr, "The Silicon Empire: TSMC's Revolution and Morris Chang's Legacy," Quartr Insights, Nov. 2025.
17. Raise Summit, "Geopolitics of Chips: How Supply Chains Are Reshaping Global Power," Raise Summit, Feb. 2026.
18. C. Tseng, "AI Demand Meets Capacity Reality: Korea's Role in the Next Memory and Manufacturing Cycle," SEMI Market Outlook, SEMI, Jun. 22, 2026.
19. SEMI, *World Fab Forecast Report, 2Q 2026*. Milpitas, CA: SEMI, Jun. 2026.
20. Semiconductor Industry Association (SIA) and Boston Consulting Group, "Strengthening the Global Semiconductor Supply Chain in an Uncertain Era," SIA/BCG, May 2024.
21. Semiconductor Industry Association (SIA), "Chip Incentives & Investments," SIA, 2025.
22. SemiconductorX, "HBM (High Bandwidth Memory) Supply Chain," SemiconductorX, 2024.
23. Silicon Semiconductor, "The Geopolitics of the Semiconductor Industry: Navigating a Global Power Struggle," Silicon Semiconductor News, 2024.
24. SK hynix, "2026 Market Outlook: SK hynix's HBM to Fuel AI Memory Boom," SK hynix Newsroom, Apr. 2026.
25. Time, "America's New Critical Minerals Playbook," Time, Jun. 2026.
26. TrendForce, "SK hynix Reportedly to Supply About Two-Thirds of NVIDIA HBM4; Samsung Targets Early Delivery," TrendForce News, Jan. 2026.
27. Taiwan Semiconductor Manufacturing Company (TSMC), "TSMC Arizona and U.S. Department of Commerce Announce up to US\$6.6 Billion in Proposed CHIPS Act Direct Funding," TSMC Press Release, Apr. 2024.
28. TSPA Semiconductor, "The Infinite AI Compute Loop: HBM Big Three + TSMC × NVIDIA × OpenAI Shaping the Next-Generation Industry Chain," TSPA Semiconductor Substack, Oct. 2025.
29. U.S.-China Economic and Security Review Commission (USCC), "Chapter 9: Chained to China—Beijing's Weaponization of Supply Chains," 2025 Annual Report to Congress. Washington, DC: USCC, Nov. 2025.
30. Visual Capitalist, "How Much Control China Has Over the World's Critical Minerals," Elements by Visual Capitalist, Oct. 2025.