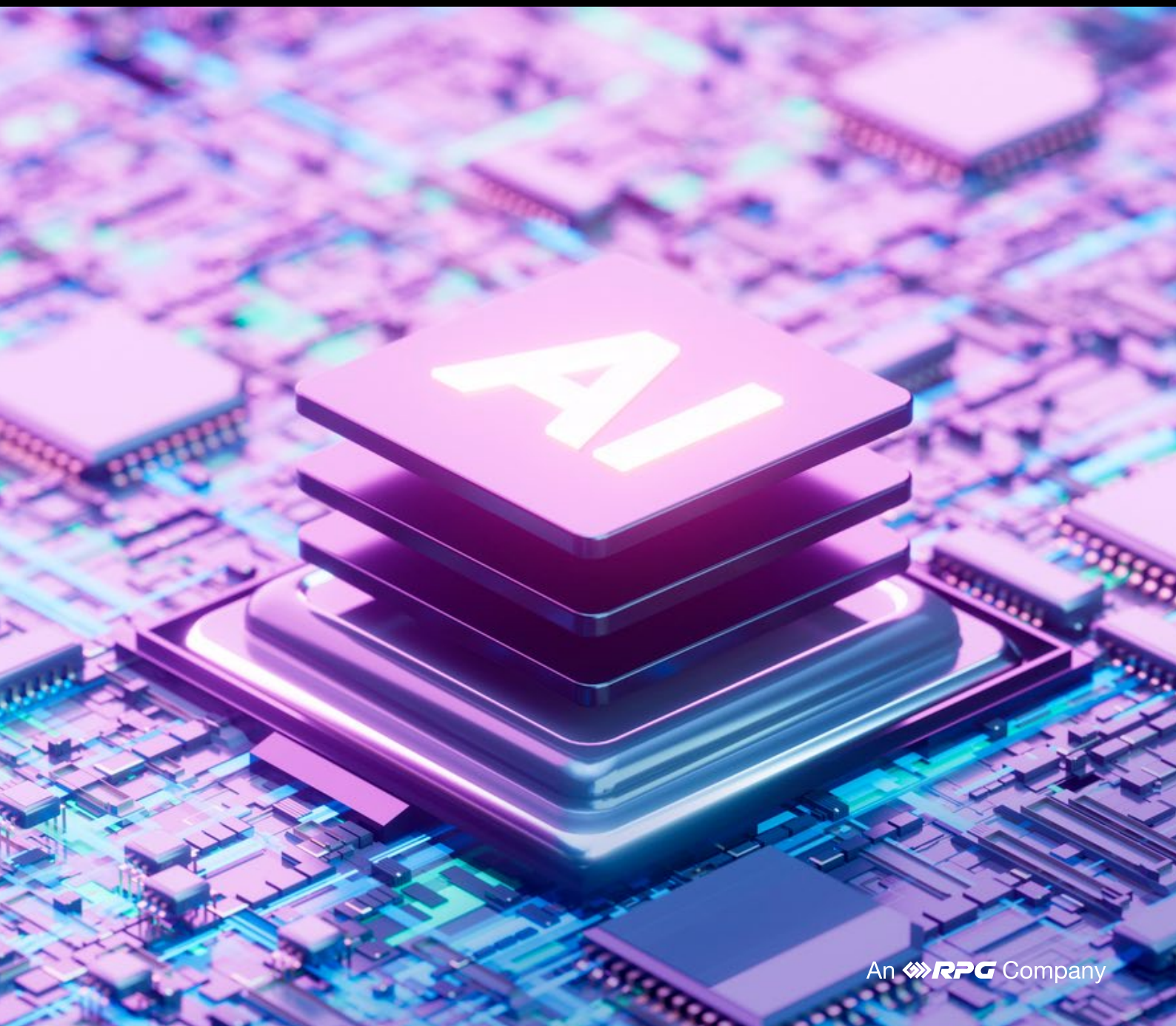
zensar

Foundries | Memory | AI Silicon | Equipment

Global Trends in Semiconductors

Whitepaper



An RPG Company

Table of Contents

Global Semiconductor Outlook

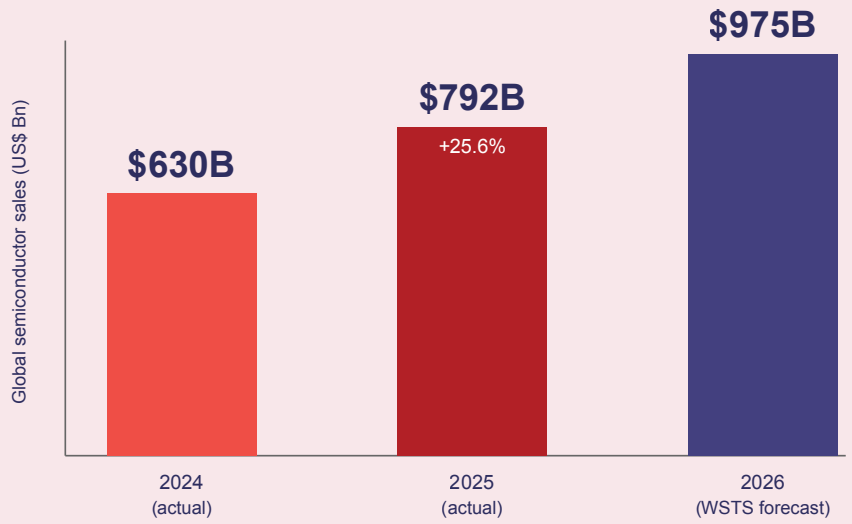
Two facts now anchor every conversation in the semiconductor industry: 2025 was the largest single-year revenue in semiconductor history, and 2026 will be the first to approach $1 trillion in revenue. AI demand has reshaped every sub-sector - logic, memory, packaging, equipment - and drawn in record levels of geopolitical scrutiny.

1.

The Market crosses $792Bn in 2025 and approaches $1Tn in 2026

Global semiconductor sales hit $791.7 billion in 2025 (SIA, February 2026), up 25.6% on 2024. WSTS forecasts 2026 sales of $975.4 billion, a further +25% growth. Q1 2026 alone reached $298.5 billion, and logic and memory sectors each grew over 30% in 2025, driven primarily by AI-related data center infrastructure.

Global semiconductor market: $630B → $792B → ~$1T

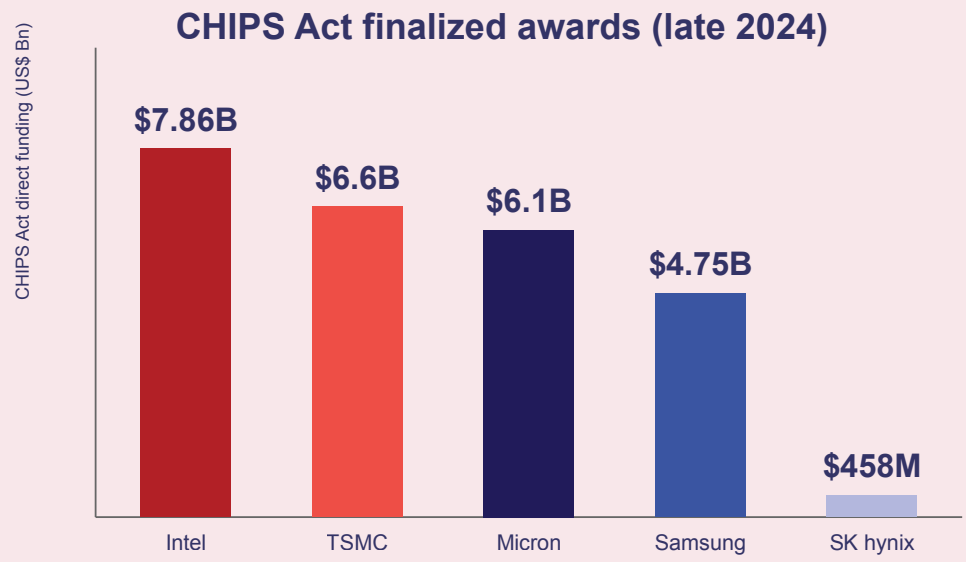


Source: SIA Annual Sales Release Feb 2026 (2024, 2025 actuals); WSTS Autumn 2025 Forecast (2026F).

2.

CHIPS Act funding is locked in; geographic competition intensifies

The US Commerce Department finalized direct CHIPS Act awards totaling roughly $25.8 billion to five firms - Intel ($7.86Bn), TSMC ($6.6Bn), Micron ($6.1Bn), Samsung ($4.75Bn), SK Hynix ($458Mn) - by late 2024. The EU Chips Act (€43Bn), Japan's Rapidus initiative, Korea's mega-cluster, and India's $19.5 billion in approved projects across 12 facilities make this the most coordinated industrial policy cycle the sector has ever seen.



Source: US Department of Commerce CHIPS Program Office final award notifications, Nov–Dec 2024.

Sources: *: SIA Annual Sales Release Feb 2026; WSTS Autumn 2025 Forecast; US Department of Commerce CHIPS Program Office award notifications Nov–Dec 2024; India Semiconductor Mission status May 2026; European Commission Chips Act.*

Global Semiconductor Sector Overview

Eight structural forces are simultaneously reshaping the industry. They explain why every major operator - foundries, IDMs, memory, fabless, and equipment - is simultaneously rebuilding its strategy around AI, geopolitics, and packaging.

AI infrastructure is the demand engine.

Logic revenue grew 37% in 2025 to $301.9 billion, and memory grew 28% to $223.1 billion (WSTS), both led by AI-related computing demand. The computer segment alone expanded by more than 60% year-on-year, and Deloitte estimates the 2026 AI chip market at roughly $500 billion.

Export controls have become industrial policy.

The US Bureau of Industry and Security added 42 PRC entities to the Entity List in March 2025 and 23 more in September 2025. In August 2025, NVIDIA's H200 and AMD's MI308 were approved for sale in China, with the US government taking 15% of revenue; H200 approval followed in December 2025.

Sovereign semiconductor fabrication and friend-shoring have become operational.

The CHIPS Act funding decisions have been finalized; TSMC has commenced N4 volume production in Arizona and has authorized the acquisition of an additional 900 acres in Phoenix for advanced packaging facilities by January 2026. India has endorsed 12 projects valued at approximately $19.5 billion across six states as part of ISM.

Advanced packaging represents the primary bottleneck.

TSMC's CoWoS capacity expanded to about 680,000 wafers in 2025, with the interposer roadmap targeting a 9.5x reticle by 2027 and a 14x reticle in 2028. TSMC captures roughly 50% of the high-end CoWoS market for AI chips; Alphabet has allegedly scaled back its TPU production goal for 2026 due to constraints in CoWoS supply.

Memory is the new margin pool.

HBM and high-capacity DRAM/NAND are reshaping memory economics. Counterpoint and Tom's Hardware report that memory prices rose by around 50% in Q4 2025, with the rally projected to continue through 2026, and IDC has warned that the PC market could shrink by up to 9% in 2026 due to memory pricing.

Hyperscalers are becoming chip designers.

Google TPU, AWS Trainium, Microsoft Maia, and Meta MTIA are now production-grade silicon programs. NVIDIA still dominates merchant AI silicon - FY26 revenue reached $215.9 billion, up 65% year-on-year - but the long-term mix of merchant vs custom silicon is shifting.

Mature nodes are a geopolitical asymmetry.

While the West restricts China's access to advanced nodes, China is consolidating dominance in 28nm-and-above legacy nodes that power automotive, industrial, and consumer electronics. This creates a new form of supply-chain dependency that policy is only beginning to address.

ASML's monopoly on EUV is the most genuine chokepoint.

EUV and High-NA EUV (each costing over €350 million) remain ASML's exclusive franchise. TSMC has chosen to extend its current low-NA EUV via Design-Technology Co-Optimization for A12 and A13 (1.2nm, 1.3nm), targeting 2029 mass production - a deliberate divergence from Intel's High-NA strategy.

Sources: *WSTS March 2026 Press Release; SIA Global Semiconductor Sales February 2026; US Congressional Research Service R48642 (Export Controls); Deloitte 2026 Semiconductor Industry Outlook; Trendforce and DigiTimes industry reports; TSMC 2026 North America Technology Symposium.*

Key Market Movers and Shakers

Across foundries, memory, AI silicon, hyperscaler custom chips, and equipment, these organizations are defining the 2026 semiconductor cycle.

Foundries and IDMs

TSMC: Received $6.6 billion CHIPS Act grant plus $5 billion in loans, finalized November 2024, for a third Arizona fab; approved purchase of 900 acres in Phoenix in January 2026; planning a CoWoS and 3D-IC packaging facility in Arizona before 2029. Unveiled A12 (1.2nm) and A13 (1.3nm) processes at the 2026 North America Technology Symposium, targeting mass production in 2029.

Intel: Received $7.86 billion CHIPS Act direct funding (finalized late November 2024, reduced from $8.5 billion originally announced). The US government has subsequently taken a 10% equity stake in Intel under the Trump administration.
Samsung Foundry: Received $4.75 billion CHIPS Act funding (finalized December 20, 2024; reduced from proposed $6.4 billion). Focus on Taylor, Texas, leading-edge logic fabs, and Austin expansion.

Memory and HBM leaders

SK Hynix: Received $458 million CHIPS Act funding for Indiana HBM packaging facility ($3.87 billion total investment); leading HBM supplier with HBM4 16-Hi volume production targeted for first half of 2026.

Micron: Received $6.1 billion CHIPS Act funding (finalized December 10, 2024) for New York and Idaho fabs; scaling HBM, DDR5, and enterprise SSD capacity to capture AI-driven memory pricing tailwinds.

AI silicon and hyperscaler custom compute

NVIDIA: FY26 revenue of $215.9 billion, up 65% year-on-year, with data center compute up 59% and networking up 142%. Blackwell Ultra delivers 50× higher throughput and 35× lower token cost versus Hopper.

AMD: MI355X and MI325X positioning to close the gap with NVIDIA; CDNA roadmap and ROCm software stack push to address application-layer competitiveness.

Hyperscaler ASICs: Google TPU, AWS Trainium, Microsoft Maia, and Meta MTIA are reshaping demand, moving hyperscalers from chip buyers to chip designers, in partnership with Broadcom and Marvell.

Equipment and materials

ASML: EUV and High-NA EUV monopoly remains the most genuine chokepoint in the supply chain. Shipped its first advanced-packaging lithography system in late 2025; ASML projects $71 billion in revenue by 2030, driven by EUV demand for AI.

Applied Materials, Lam Research, Tokyo Electron:
Advanced-packaging, deposition, and etch leaders; sales of chip production equipment projected to reach $156 billion by 2027 (SEMI forecast), driven by AI and DRAM demand.

Sources: *US Department of Commerce CHIPS Program Office; NVIDIA FY2026 Proxy Statement (Form DEF 14A); TSMC 2026 North America Technology Symposium; ASML investor disclosures; SEMI Equipment Forecast.*

Eight Trends Defining 2026

This white paper examines each trend in turn - its drivers, leading players, and the executive imperatives it creates. These eight have been selected based on strategic significance, breadth of impact across the semiconductor value chain, and the depth of verified data available to support the analysis.

01 The AI capex supercycle and the $1 trillion market

02 Geopolitics and the new export-control regime

03 Sovereign fabs and friend-shoring

04 Advanced packaging as the real chokepoint

05 Memory and HBM as the new margin pool

06 AI custom silicon reshapes the hyperscaler stack

07 Mature-node asymmetry and China dominance

08 Sustainability and energy intensity of AI silicon

01. The AI Capex Supercycle and the $1 Trillion Market

At a glance

- The semiconductor industry posted record revenue of $791.7 billion in 2025 - up 25.6% year-on-year -and WSTS forecasts 2026 will approach $1 trillion with sales of $975.4 billion (+25%). Logic grew 37% and memory 28% in 2025, both driven by AI-related computing demand.
- The growth is concentrated. The computer end-market segment alone expanded by over 60% in 2025, reflecting unprecedented data center buildouts. Deloitte estimates the 2026 AI chip market at roughly $500 billion - more than half the total industry revenue.

Key market trends

1 Hyperscaler capex multi-year visibility

- AI infrastructure programs are pre-committed through 2027 in HBM, CoWoS, and advanced logic capacity.
- NVIDIA and AMD GPU platforms, along with a growing set of hyperscaler custom silicon programs, account for most of the incremental demand.

2 Q1 2026 momentum confirms the cycle

- Q1 2026 global sales reached $298.5 billion, a 25% sequential increase from Q4 2025.
- January and February 2026 averaged over $85 billion per month, materially above the 2025 monthly run rate of approximately $66 billion.

3 Regional growth concentration

- In 2025, the Asia Pacific region (excluding China and Japan) experienced a growth rate of 45.4%; the Americas saw an increase of 31.4%; and China saw a growth rate of 17.9%.
- Europe experienced modest growth of 6.7%; Japan recorded a 4.3% decline for the full year, reflecting changes in end-market dynamics and structural transitions.

4 Hyperscaler capex multi-year visibility

- TSMC executives have publicly noted that they are “very nervous” about concerns over an AI bubble, despite record quarters.
- Roughly $1 trillion in big-tech stock value tumbled at one point as investors weighed projected AI spending against the realization of revenue.

Initiatives by major players

TSMC: Setting the pace as the gating supplier for AI accelerators; CoWoS capacity reached approximately 680,000 wafers in 2025; the interposer roadmap extends to 9.5× reticle in 2027 and 14× in 2028.
NVIDIA: FY26 revenue of $215.9 billion, up 65% year-on-year; data center compute up 59%; networking up 142%. Original $190 billion stretch revenue plan adjusted to $160 billion due to H20 export controls in H1 FY26.
Memory makers (SK Hynix, Micron, Samsung): Memory makers expected to earn the bulk of AI-cycle profits - Tom's Hardware reports memory pricing rose roughly 50% in Q4 2025, with the rally projected to continue through 2026.
industry revenue.

Key imperatives

1 Treat the supercycle as multi-year, not one quarter.

- Hyperscaler capex commitments and HBM pre-allocations point to sustained demand through 2027.
- Capacity planning, supplier contracting, and workforce hiring should be sized to a multi-year horizon, not the next earnings cycle.

2 Diversify across the AI value chain.

- Logic, memory, advanced packaging, equipment, and substrates are all benefiting unevenly.
- Portfolios concentrated in any single layer face higher cyclical risk than diversified positions across the value chain.

3 Watch the bubble signals, but don't underinvest.

- AI bubble concerns are real and acknowledged by TSMC itself.
- The asymmetric risk of underinvesting in capacity that takes years to build remains higher than the risk of overinvesting at current price levels.

4 Plan for non-AI end-market softness.

- PC and smartphone demand may decline in 2026 -IDC has warned of a PC market contraction of up to 9% due to memory pricing.
- Companies dependent on consumer end-markets should plan for compressed near-term demand even as AI continues to lead.

Sources: *SIA Annual Sales Release Feb 2026; WSTS Autumn 2025 Forecast; WSTS March 2026 Press Release; Deloitte 2026 Semiconductor Industry Outlook; NVIDIA FY2026 Form DEF 14A.*

02. Geopolitics and the New Export-Control Regime

Upon initial observation

- Export controls have evolved from a means of targeted restriction to a vital aspect of industrial policy. In March 2025, the US Bureau of Industry and Security added 42 entities from the People's Republic of China to the Entity List, and in September 2025, it included an additional 23; during this time, BIS also removed Samsung and SK Hynix's designated facilities in China from the list.

- In August 2025, NVIDIA's H20 and AMD's MI308 were approved for sale in China, with the US government receiving 15% of the revenue from these sales. H200 approval to China followed in December 2025 - a substantial expansion versus the spring 2025 freeze.

Key market trends

1 The 15% revenue arrangement

- An August 11, 2025 arrangement allows Nvidia and AMD to sell H20 and MI308 to China, with the US government collecting 15% of revenue.
- This is a novel industrial-policy instrument with no clear precedent - and not yet widely understood by buyers.

2 Performance-tier licensing

- Under the revised export rules, "green zone" chips such as H20 flow more freely while more advanced GPUs remain restricted.
- Chinese firms have begun optimizing midrange offerings and ramping their own semiconductor research toward AI hardware self-sufficiency by 2027.

3 Entity-List dynamics

- BIS has expanded and, on multiple occasions, selectively reduced the Entity List through 2025–2026.
- Notably, Samsung and SK Hynix's named PRC facilities were removed in September 2025, demonstrating that listings are not permanent.

4 The Huawei catch-up question

- BIS assessed in May 2025 that Huawei had developed its Ascend chips in violation of US controls.
- CFR analysis estimates Huawei will not be able to produce a chip matching H200 performance until Q4 2027 at the earliest - and even then, constrained by manufacturing bottlenecks.

Initiatives by major players

NVIDIA: Original FY26 stretch revenue goal of $190 billion was automatically adjusted to $160 billion due to additional H20 export controls in H1. Actual FY2026 revenue reached $215.9 billion.
AMD: MI308 approval for China sales restored a material revenue line in late 2025; broader MI355X positioning continues to develop.

Huawei continues to develop the Ascend chip line, bound by manufacturing constraints and the inability to source the most advanced equipment. Production scale estimated at 1 – 2% of US AI chip production in 2026.

Key imperatives

1 Plan for licensing volatility.

- Export controls have changed direction multiple times through 2025 – 2026 across administrations.
- Build product roadmaps and supply chains that can accommodate licensing shifts with three-to-six months' notice.

2 Track the 15% revenue arrangement as a precedent.

- The H20 / MI308 revenue-sharing model may be extended to other chip families or geographies.
- Treasury and pricing strategies should be built to accommodate similar arrangements without disrupting margins.

3 Distinguish performance tiers in product planning.

- The "green-zone"/restricted-tier framework appears to be a durable feature of US policy.
- Product roadmaps should explicitly target each performance tier rather than design only at the frontier.

4 Diversify customer geography deliberately.

- Revenue concentration in China is now a regulatory risk as much as a commercial one.
- Treat geographic diversification as an explicit business-continuity discipline, not just a growth strategy.

Sources: *US Congressional Research Service R48642; Built In coverage Apr 2026; Council on Foreign Relations Dec 2025; Data Center Knowledge Feb 2026; NVIDIA FY2026 Form DEF 14A.*

03. Sovereign Fabs and Friend-Shoring

At a glance

- The US CHIPS Act secured around $25.8 billion in confirmed direct funding for five companies -
Intel ($7.86Bn), TSMC ($6.6Bn), Micron ($6.1Bn), Samsung ($4.75Bn), and SK Hynix ($458Mn) - by the end of 2024, in addition to $52.7 billion in overall program authority and related loan facilities.
- Parallel programs are now operating in the EU (€43 billion CHIPS Act), Japan (Rapidus and TSMC Kumamoto), Korea (mega-cluster), and India (12 approved projects worth roughly $19.5 billion, including the $10 billion Tata Electronics Dholera fab and operational Micron and Kaynes facilities at Sanand).

Key market trends

1 CHIPS Act fully finalized

- All five major awards were finalized between November and December 2024 under the Biden administration.
- The Trump administration subsequently took a 10% equity stake in Intel in exchange for CHIPS Act funds (announced in August 2025).

2 Performance-tier licensing

- TSMC's first Arizona fab will enter N4 volume production in the first half of 2025.
- TSMC's board approved the purchase of an additional 900 acres in Phoenix on January 7, 2026, for advanced packaging capability - building CoWoS and 3D-IC capacity in Arizona before 2029.

3 India Semiconductor Mission scales

- 12 projects approved under ISM as of May 5, 2026, with a cumulative investment of approximately Rs 1.64 lakh crore (~$19.5 billion) across six states.
- Micron's $2.75 billion ATMP at Sanand was inaugurated on February 28, 2026; Kaynes Semicon's $347.9 million OSAT facility opened on March 31, 2026. Budget 2026 – 27 allocated Rs 8,000 crore - the largest single-year outlay since the program launched.

4 Cost asymmetry remains

- Construction costs in the US run materially higher than in Asia, and labor shortages have delayed multiple flagship projects.
- Companies have indicated willingness to accept 20 –30% higher production costs for geographic certainty - the new political economy of friend-shoring.

Initiatives by major players

TSMC (US): First Arizona fab in volume production; second fab scheduled for 2028; third planned by the end of the decade. Plans 12 fabs in Arizona long-term and a CoWoS /3D-IC packaging facility before 2029. Has committed $100 billion in additional US investment for AI chip production.

Intel: $7.86 billion CHIPS Act funding plus 10% US government equity stake; Ohio fab delayed into 2026 due to labor and grant timing; Panther Lake launch on 18A is the credibility test for advanced-node execution.

Rapidus (Japan): Secured $1.7 billion combined government and private funding for 2nm chip production -Japan's strategic re-entry into leading-edge logic.

India Semiconductor Mission (ISM): Twelve approved projects across Gujarat, Uttar Pradesh, Assam, Andhra Pradesh, Odisha, and Punjab. Tata Electronics Dholera $10Bn fab; Micron Sanand ATMP $2.75Bn; Kaynes Sanand OSAT $347.9Mn; CG Power G1 OSAT operational since August 2025.

Key imperatives

1 Sequence regulatory and capital decisions deliberately.

- Geographic capital allocation is now contingent on policy, equity, and licensing arrangements that vary widely by host country.
- Treat the policy framework as a primary input to siting decisions, not a constraint to work around afterward.

2 Plan for construction-cost premiums

- US and EU construction is structurally more expensive than in Asia, and the gap is unlikely to close quickly.
- Build a 20 – 30% cost premium into Western fab project economics, and explicitly factor in the subsidies.

3 Engage early with sovereign programs.

- India's ISM, Japan's Rapidus, and Korea's mega-cluster all offer significant fiscal incentives - but for committed early movers.
- First-mover advantages in sovereign-fab programs include preferred sites, anchor-tenant status, and ecosystem partner networks.

4 Diversify the manufacturing footprint.

- Concentration in any single jurisdiction is now a board-level risk.
- Strategic-grade firms are building presence in at least three of the four major geographies - the US, EU, Asia, and India.

Sources: *US Department of Commerce CHIPS Program Office award notifications; Manufacturing Dive CHIPS Act tracker; India Semiconductor Mission (ism.gov.in); India Briefing May 2026; PIB India Feb 2026 ISM 2.0 release.*

04. Advanced Packaging as the Key Bottleneck

At a glance

TSMC holds around 50% of the high-end CoWoS market for AI chips and nearly 18 – 20% of total advanced packaging revenue. By 2025, CoWoS capacity will have grown to about 680,000 wafers, with the interposer roadmap reaching 9.5× reticle by 2027 and 14× by 2028.

The demand surpasses the available supply. It has been reported that Alphabet plans to cut its 2026 TPU production target because of restricted access to TSMC's CoWoS advanced packaging capacity, especially after NVIDIA was given priority allocations. Advanced packaging is increasingly becoming the limiting factor, rather than lithography.

Key market trends

1 CoWoS and its successors

- TSMC's roadmap pairs CoWoS with SoIC, COUPE (co-packaged optics), CoPoS, and InFO for chips beyond 2nm.
- AMD already uses an SoIC + CoWoS structure; the system architecture, not just node geometry, is now the differentiator.

2 Performance-tier licensing

- Interposer size has become a primary capacity constraint for AI accelerators.
- TSMC will extend CoWoS interposer size to 9.5× reticle in 2027 and 14× reticle in 2028, addressing the scale-up needs of next-generation training systems.

3 India Semiconductor Mission scales

- TSMC has flagged System-on-Wafer (SOW) as the architectural direction beyond traditional interposers.
- A 300mm wafer is roughly 40× the reticle size when measured using CoWoS/interposer technology. Cerebras is among the first to use SOW technology for inferencing silicon.

4 ASML enters the packaging market

- ASML shipped its first advanced-packaging lithography system in late 2025 - its first material expansion beyond EUV.
- TSMC's grip on high-end CoWoS limits short-term displacement risk, but the equipment supply chain is broadening.

Initiatives by major players

TSMC: Approximately 50% market share in high-end CoWoS for AI; CoWoS capacity at ~680,000 wafers in 2025; building CoWoS and 3D-IC capacity in Arizona before 2029. AP7 packaging technology targets 2026 output; advanced packaging growing at ~80% CAGR.

NVIDIA: Has secured priority CoWoS allocation at TSMC, displacing some Alphabet TPU production from 2026 plans. NVIDIA's volume drives the leading edge of the packaging roadmap.

ASML: Shipped first advanced-packaging lithography system in late 2025; expanding beyond EUV monopoly into the packaging value chain.

Applied Materials, Lam, Tokyo Electron: Primary suppliers of deposition and etch tools that enable advanced packaging; also primary targets of proposed export controls on shipments to China.

Key imperatives

1 Treat packaging capacity as a primary planning constraint

- Packaging - not logic node - now gates AI accelerator output for most major customers.
- Allocate packaging capacity early and explicitly in capacity-planning conversations.

2 Invest in system-architecture optimization.

- Performance gains increasingly come from packaging integration - SoIC, CoWoS, COUPE - not lithographic node alone.
- Design teams need expertise spanning logic, memory, and packaging from the start.

3 Build supplier relationships across multiple packaging tiers

- OSAT providers, foundries with internal packaging, and emerging equipment vendors form a more complex supply web than in past cycles.
- Single-source dependence on any packaging supplier is now a board-level concentration risk.

4 Track the SoW transition

- System-on-Wafer represents a multi-year architectural shift beyond reticle-limited interposers.
- Companies whose roadmaps extend into the late 2020s should evaluate SoW implications now, not later.

Sources: *TSMC 2026 North America Technology Symposium; EE Times TSMC roadmap coverage; Data Center Dynamics April 2026; DigiTimes packaging capacity reports; 247WallSt March 2026 ASML coverage.*

05. Memory and HBM as the New Margin Pool

At a glance

Memory revenue grew 28% in 2025 (WSTS) - exceeding logic's growth rate in absolute revenue terms for the first time in years. WSTS forecasts that memory will grow by another 30% in 2026.

Pricing has reset higher. Counterpoint reported memory prices rose roughly 50% in Q4 2025, with the rally projected to continue through 2026. Industry analysis indicates data centers will consume approximately 70% of all memory chips made in 2026.

Key market trends

1 HBM is the strategic margin pool

- HBM3E 12-Hi is in volume shipment to major AI customers; HBM4 16-Hi volume production is targeted by SK Hynix for the first half of 2026.
- HBM pricing per unit is materially higher than commodity DRAM, with capacity largely pre-allocated to NVIDIA and AMD.

2 Memory shortage shapes consumer markets

- IDC has warned that the PC market could shrink by up to 9% in 2026 due to skyrocketing RAM prices.
- TrendForce has revised 2026 smartphone and notebook outlooks downward - consumer end markets are the bill-payers of the AI memory boom.

3 Hyperscaler memory pre-allocation

- AI data centers are reportedly absorbing the bulk of leading-edge memory output.
- Tom's Hardware has described "a pricing apocalypse that could last a decade" if AI demand growth continues at the current pace.

4 Enterprise SSD and DDR5 ride the same wave

- Beyond HBM, DDR5, and enterprise SSD demand, average selling prices have reset materially higher.
- Micron, Samsung, and SK Hynix are scaling all three product lines simultaneously to capture the cycle.

Initiatives by major players

SK Hynix is a top supplier of HBM3E and aims to start mass production of HBM4 16-Hi in the first half of 2026, with the majority of its capacity already allocated to NVIDIA and AMD. The company has secured $458 million in CHIPS Act funding for its HBM packaging facility in Indiana, bringing the total to $3.87 billion.

Micron has been granted $6.1 billion in funding from the CHIPS Act (the agreement was finalized on December 10, 2024) for its fabs in New York and Idaho. They are currently shipping HBM3E 12-Hi production-capable products to industry partners while also increasing their DRAM, NAND, and enterprise SSD production capacity.

Samsung Memory is a major DRAM and NAND producer; its HBM roadmap pursues volume parity with SK Hynix; and memory business profitability has rebounded materially with pricing recovery.

Key imperatives

1 Consider memory as a key strategic asset rather than just a commodity.

- Fluctuations in memory prices significantly affect the financial dynamics of AI infrastructure and consumer electronics.
- Approach memory acquisition strategy as an important issue for the board, rather than a standard buying decision.

2 Plan for multi-year HBM supply tightness.

- HBM3E and HBM4 supply is pre-allocated through 2026 and into 2027.
- Companies dependent on HBM should secure multi-year contracts now, not negotiate on the spot market.

3 Stress-test downstream end-market exposure.

- PC, smartphone, and consumer electronics demand may contract in 2026 because of memory pricing.
- Build demand scenarios that explicitly capture the downside path.

4 Track the SoW transition

- Memory cycles have historically been brutal in both directions.
- Operators should use the current up-cycle to fund capacity expansion and HBM transition rather than maximize near-term distributions.

Sources: *WSTS Autumn 2025 Forecast; Counterpoint memory pricing Jan 2026; Tom's Hardware coverage Oct 2025; IDC PC market warning Dec 2025; TrendForce Nov 2025 revision; SEMI silicon wafer shipment forecast.*

06. AI Custom Silicon Reshapes the Hyperscaler Stack

At a glance

NVIDIA still dominates merchant AI silicon - FY26 revenue reached $215.9 billion (+65% year-on-year) and Blackwell Ultra delivers 50× higher throughput and 35× lower token cost versus Hopper.

But the long-term mix is shifting. Google TPU, AWS Trainium, Microsoft Maia, and Meta MTIA are now in production-grade deployment, with hyperscalers partnering with Broadcom and Marvell to design their own AI ASICs at scale.

Key market trends

1 Blackwell Ultra at production scale

- Blackwell Ultra delivers 50× the throughput and 35× lower token cost than the prior Hopper generation.
- GB200 NVL72 rack-scale systems are defining the AI compute reference architecture for hyperscaler deployments through 2026.

2 Hyperscaler ASICs reach production grade

- Google TPU, AWS Trainium, Microsoft Maia, and Meta MTIA are no longer experimental.
- Each represents a different bet on the architecture-vs-cost trade-off, and each pulls on TSMC's capacity and on the HBM suppliers.

3 AMD pursues a closing strategy

- MI355X and MI325X are positioning aggressively against NVIDIA on price-performance.
- CDNA roadmap and ROCm software stack maturation are the key drivers - the application-software gap, not the silicon gap, is now the constraint.

4 Broadcom and Marvell ascend

- Custom-silicon design partners are now strategic suppliers to every major hyperscaler.
- These firms are absorbing engineering effort that hyperscalers cannot scale internally fast enough.

Initiatives by major players

NVIDIA: FY26 revenue $215.9 billion (+65% YoY); data center compute +59%; networking +142%. Original $190 billion stretch revenue plan adjusted to $160 billion due to H20 controls in H1; actual delivery materially exceeded the stretch goal.

AMD: MI325X/MI355X positioning aggressively against NVIDIA; the CDNA roadmap and ROCm software stack are critical investments to close the application-layer gap.

Hyperscaler ASICs (Google, AWS, Microsoft, Meta): TPU v7, Trainium 3, Maia, and MTIA are now production-grade silicon programs, partnered with Broadcom and Marvell, moving hyperscalers from chip buyers to chip designers.

Key imperatives

1 Reframe AI silicon as a portfolio question

- Most large customers will mix merchant GPUs and custom ASICs over the next three years.
- Procurement, capacity planning, and software stack strategy should be built around that hybrid model, not single-vendor lock-in.

2 Treat the software gap as the moat

- CUDA's lead is the most durable element of NVIDIA's position.
- Closing it - for AMD, for custom-silicon ecosystems - requires sustained software investment, not just better silicon.

3 Plan for hyperscaler custom-silicon expansion

- TPU, Trainium, Maia, and MTIA volumes will grow materially through 2027.
- Suppliers of memory, packaging, and tooling should treat hyperscalers as direct customers, not downstream consumers.

4 Watch the export-control overhang on China-bound silicon

- Custom silicon programs face the same licensing dynamics as merchant GPUs.
- Architecture decisions should consider the addressable geographic market, not just technical performance.

Sources: *NVIDIA FY2026 Form DEF 14A; AMD investor disclosures and ISSCC 2026 papers; Reuters AMD November 2025; company press releases for hyperscaler custom silicon.*

07. Mature-Node Asymmetry and China's Dominance

At a glance

While export controls restrict China's access to advanced nodes, China has consolidated dominance in 28nm-and-above legacy nodes that power automotive, industrial, power management, and consumer electronics.

This creates a new form of supply-chain dependency for Western OEMs - one that current trade policy is only beginning to address. Production cost and scale economics favor the incumbent.

Key market trends

1 The legacy-node share shift

- Mature-node fab capacity has expanded faster in China than anywhere else through 2024–2026.
- Power management chips, microcontrollers, analog, and consumer logic are increasingly supplied by China across global value chains.

2 Decoupling difficulty

- Re-shoring mature-node production faces unfavorable economics - Western manufacturers cannot match the Chinese unit-cost structure on legacy nodes.
- Subsidies sized for advanced-node fabs are an order of magnitude too small for comprehensive mature-node decoupling.

3 Automotive and industrial exposure

- Automotive supply chains in particular are heavily exposed to Chinese-made power and analog chips.
- Recall the 2021 chip crisis - most of the missing chips were mature-node parts, not advanced logic.

4 Policy response is nascent

- Western governments are only beginning to consider tools to address mature-node asymmetry.
- The political economy of subsidizing low-margin commodity chips is difficult, and the policy toolkit remains underdeveloped.

Initiatives by major players

Chinese foundries (SMIC and peers): Aggressive mature-node capacity expansion across 28nm and above, supported by state-backed financing and equipment sourced where export controls permit.

Western OEMs (automotive, industrial): Major automotive OEMs and industrial-electronics firms remain dependent on mature-node supply chains that touch Chinese capacity at multiple points.

US and EU policy: Both jurisdictions are actively reviewing tools - tariffs, sourcing requirements, additional CHIPS-style support - to address mature-node dependency, though no comprehensive policy is yet in force.

Key imperatives

1 Map mature-node dependency in your supply chain.

- Most operators do not have a clear view of how much of their product stack depends on the Chinese mature-node supply.
- A clean supply-chain map is a precondition for any meaningful diversification strategy.

2 Diversify before being mandated to.

- Policy mandates on mature-node sourcing may emerge in the next 12 – 24 months.
- Operators that diversify proactively will face lower switching costs than those forced to react under tight timelines.

3 Engage with the policy debate. a

- The policy framework for mature-node decoupling is being shaped now.
- Industry voices on what is and is not achievable will materially affect the eventual rules.

4 Plan for cost premiums on diversified sourcing.

- Non-Chinese mature-node supply is structurally more expensive at current capacity utilization.
- Pricing strategies and contract terms should explicitly incorporate this premium.

Sources: *US Congressional Research Service R48642; industry analyst commentary on mature-node competition; Council on Foreign Relations export-control analysis, Dec 2025.*

08. Sustainability and Energy Intensity of AI Silicon

At a glance

AI data centers are pushing power, water, and emissions limits across every major geography. Fab construction is energy-, water-, and emissions-intensive, and the chips themselves drive incremental electricity demand of multi-gigawatt scale wherever they are deployed.

Sustainability disclosures, water permitting, and emissions reporting are now material competitive variables - not corporate social responsibility line items. The same is true for fabs operating in water-stressed regions and for the data centers that consume their output.

Key market trends

1 Energy intensity at the chip level

- Blackwell Ultra delivers 50× higher throughput and 35× lower token cost versus Hopper - efficiency improvements are real and material.
- But the demand for absolute power from AI clusters continues to grow faster than per-chip efficiency gains.

2 Water and fab siting

- Leading-edge fabs consume large volumes of ultra-pure water; siting decisions are increasingly shaped by water availability.
- Arizona, Taiwan, and several Indian sites have all faced water-permitting scrutiny - a constraint the industry has historically under-weighted.

3 Co-packaged optics for power reduction

- Co-packaged optics (COUPE) can cut networking power consumption by up to 70%.
- TSMC's COUPE platform applies SoIC-X chip-stacking to bring optics close to compute - material for the next generation of AI clusters.

4 Scope 3 disclosure becomes competitive

- Scope 3 emissions disclosures - covering supply chains and end use - are becoming material to customer procurement and capital-market access.
- Major hyperscaler customers are increasingly asking for supply-chain emissions data as part of vendor selection.

Initiatives by major players

TSMC: Public commitments to sustainability include sourcing renewable energy for its fabs in Taiwan; enhancements to the CoWoS/COUPE platform achieve significant reductions in power consumption per AI accelerator.

NVIDIA: The 35× reduction in token costs for Blackwell Ultra compared to Hopper signifies the most significant efficiency improvement per unit in the AI age; however, total power demand continues to rise because the volume of deployed compute resources increases at a faster rate.

Sovereign fab programs: Most CHIPS Act, EU Chips Act, and ISM agreements include sustainability conditions -water reuse, renewable sourcing, emissions targets - as part of funding terms.

Key imperatives

1 Treat Scope 3 disclosure as a competitive variable.

- Customer procurement and capital markets are converging on the need for supply-chain emissions transparency.
- Firms that lead on Scope 3 reporting will access cheaper capital and preferred customer access -those that lag will be

2 Diversify before being mandated to.

- Throughput-per-watt and tokens-per-joule are becoming customer-visible specifications.
- Engineering organizations should treat efficiency targets as primary, not secondary, design constraints.

3 Plan fab siting around water and power realistically.

- Water-stressed sites face permitting delays and growing community opposition.
- Site selection should weigh water and power availability on par with capital incentives.

4 Invest in advanced packaging for power gains

- Co-packaged optics and 3D integration deliver power reductions that no node shrink can match.
- These technologies are the most economically attractive sustainability lever in

Sources: *TSMC sustainability disclosures; NVIDIA Blackwell Ultra technical announcements; sovereign-fab program funding conditions (CHIPS Act, EU Chips Act, ISM); industry analyst coverage of COUPE and 3D integration.*

Looking Ahead: A 2030 Perspective

By 2030, the semiconductor industry will look fundamentally different. Five shifts will be visible across foundries, memory, AI silicon, equipment, and manufacturing geographies.

1.

The trillion-dollar industry will be the norm, not the milestone. If WSTS's 2026 forecast holds, sustained AI-driven growth would compound to a $1.4 – 1.6 trillion industry by 2030. The capacity built to support that scale is already in motion.

2.

Packaging will be the new node. Performance gains will come increasingly from packaging integration —CoWoS successors, SoIC, COUPE, System-on-Wafer — rather than from lithographic node shrinks alone. The boundary between fabrication and packaging is dissolving.

3.

Sovereign manufacturing capacity will be measurable, not aspirational. The US, EU, Japan, Korea, and India fab capacity built under CHIPS-style programs will be producing leading-edge chips at material volumes by 2030. The geographic concentration of advanced semiconductor capacity will have meaningfully eased.

4.

Custom silicon will rival merchant GPUs in volume. Hyperscaler ASIC programs are scaling quickly enough to materially shift the mix away from merchant GPUs. NVIDIA's dominant position is likely to remain intact, but the share of total AI compute on custom silicon will have grown materially.

5.

The mature-node policy will catch up to the advanced-node policy. Policy attention to mature-node asymmetry is just beginning. Western trade policy and industrial policy tools will likely have evolved to address legacy-node dependency before 2030.

The supercycle is real, the geopolitics are durable, and the technology roadmap is more visible than at any previous point in the industry's history. Operators that move decisively on the eight trends in this paper will define the next generation of the semiconductor industry.

Glossary

SIA – Semiconductor Industry Association
WSTS – World Semiconductor Trade Statistics
CAGR – Compound Annual Growth Rate
CHIPS Act – Creating Helpful Incentives to Produce Semiconductors Act
EU – European Union
ISM – India Semiconductor Mission
IDM – Integrated Device Manufacturer
OSAT – Outsourced Semiconductor Assembly and Test
ATMP – Assembly, Test, Marking and Packaging
CoWoS – Chip-on-Wafer-on-Substrate (TSMC advanced packaging)
SoIC – System on Integrated Chips
InFO – Integrated Fan-Out (packaging technology)
CoPoS – Chip-on-Panel-on-Substrate
COUPE – Compact Universal Photonic Engine (TSMC co-packaged optics platform)
SoW – System-on-Wafer
HBM – High Bandwidth Memory
DRAM – Dynamic Random Access Memory
NAND – NAND Flash Memory
SSD – Solid-State Drive
GPU – Graphics Processing Unit
TPU – Tensor Processing Unit
ASIC – Application-Specific Integrated Circuit
CDNA – Compute DNA (AMD GPU compute architecture)
ROCm – Radeon Open Compute (AMD software stack)
CUDA – Compute Unified Device Architecture (NVIDIA)
EUV – Extreme Ultraviolet Lithography
High-NA EUV – High Numerical Aperture Extreme Ultraviolet Lithography
DTCO – Design-Technology Co-Optimization
nm – Nanometer
GW – Gigawatt
BIS – Bureau of Industry and Security (US Department of Commerce)
PRC – People's Republic of China
CFR – Council on Foreign Relations
CRS – Congressional Research Service (US)
SEMI – Semiconductor Equipment and Materials International
SEC – Securities and Exchange Commission (US)
DEF 14A – Definitive Proxy Statement (SEC filing)
GHG – Greenhouse Gas
Scope 3 – Value-chain (indirect) emissions category, covering supply chain and end-use
PIB – Press Information Bureau (India)
TMT – Technology, Media, and Telecommunications
ISG – Industry Solutions Group

References

Every figure cited in this paper is sourced from the references listed below. All URLs were verified at the time of paper preparation. The charts in this paper are Zensar visualizations based on verified data points from these sources.

Global market sizing and forecasts

• Semiconductor Industry Association (SIA). Global Annual Semiconductor Sales Increase 25.6% to $791.7 Billion in 2025 (February 2026). *https://www.semiconductors.org/global-annual-semiconductor-sales-increase-25-6-to-791-7-billion-in-2025/*

• SIA. Global Semiconductor Sales Increase 25% from Q4 2025 to Q1 2026. *https://www.semiconductors.org/global-semiconductor-sales-increase-25-from-q4-2025-to-q1-2026/*

• WSTS. Global Semiconductor Market grows 26% in 2025 to $796 billion (March 2026). *https://www.wsts.org/76/Global-Semiconductor-Market-grows-26-in-2025-to-796B*

• WSTS. Global Semiconductor Market Approaches $1 Trillion in 2026 -Autumn 2025 Forecast. *https://www.wsts.org/76/Recent-News-Release*

• Deloitte. 2026 Global Semiconductor Industry Outlook. *https://www.deloitte.com/us/en/insights/industry/technology/technology-media-telecom-outlooks/semiconductor-industry-outlook.html*

• Tom's Hardware. Semiconductor industry on track to hit $1 trillion in sales in 2026 (Feb 2026). *https://www.tomshardware.com/tech-industry/semiconductors/semiconductor-industry-on-track-to-hit-usd1-trillion-in-sales-in-2026-sia-predicts-bumper-forecast-follows-usd791-7-billion-haul-for-2025*

US CHIPS Act and US industrial policy

• US Department of Commerce CHIPS Program Office -finalized award notifications, Nov–Dec 2024 (covered by multiple secondary sources below).

Every figure cited in this paper is sourced from the references listed below. All URLs were verified at the time of paper preparation. The charts in this paper are Zensar visualizations based on verified data points from these sources.

Global market sizing and forecasts

• Semiconductor Industry Association (SIA). Global Annual Semiconductor Sales Increase 25.6% to $791.7 Billion in 2025 (February 2026). *https://www.semiconductors.org/global-annual-semiconductor-sales-increase-25-6-to-791-7-billion-in-2025/*

• SIA. Global Semiconductor Sales Increase 25% from Q4 2025 to Q1 2026. *https://www.semiconductors.org/global-semiconductor-sales-increase-25-from-q4-2025-to-q1-2026/*

• WSTS. Global Semiconductor Market grows 26% in 2025 to $796 billion (March 2026). *https://www.wsts.org/76/Global-Semiconductor-Market-grows-26-in-2025-to-796B*

• WSTS. Global Semiconductor Market Approaches $1 Trillion in 2026 -Autumn 2025 Forecast. *https://www.wsts.org/76/Recent-News-Release*

• Deloitte. 2026 Global Semiconductor Industry Outlook. *https://www.deloitte.com/us/en/insights/industry/technology/technology-media-telecom-outlooks/semiconductor-industry-outlook.html*

• Tom's Hardware. Semiconductor industry on track to hit $1 trillion in sales in 2026 (Feb 2026). *https://www.tomshardware.com/tech-industry/semiconductors/semiconductor-industry-on-track-to-hit-usd1-trillion-in-sales-in-2026-sia-predicts-bumper-forecast-follows-usd791-7-billion-haul-for-2025*

US CHIPS Act and US industrial policy

• US Department of Commerce CHIPS Program Office -finalized award notifications, Nov–Dec 2024 (covered by multiple secondary sources below).

• AIP. Huge CHIPS grants awarded to TSMC,

Samsung, and Micron.
https://www.aip.org/fyi/huge-chips-grants-awarded-to-tsmc-samsung-and-micron

• Manufacturing Dive. Tracking CHIPS and Science Act awards (updated through 2026).
https://www.manufacturingdive.com/news/chips-and-science-act-tracker-semiconductor-manufacturing/734039/

• Trendforce. CHIPS Act Funding Highlights before Trump Takes Office (Dec 2024).
https://www.trendforce.com/news/2024/12/23/news-chips-act-funding-highlights-before-trump-takes-office-tsmc-intel-samsung-and-more/

• Wikipedia. CHIPS and Science Act (continuously updated reference).
https://en.wikipedia.org/wiki/CHIPS_and_Science_Act

• KED Global. US government eyes stakes in Samsung and TSMC in return for CHIPS Act funds (Aug 2025).
https://www.kedglobal.com/business-politics/newsView/ked202508200001

Export controls and US-China dynamics

• US Congressional Research Service. R48642 -Export Controls and China: Advanced Semiconductors.
https://www.congress.gov/crs-product/R48642

• Built In. Trump Lifted the AI Chip Ban on China, Clearing the Way for Nvidia and AMD to Resume Sales.
https://builtin.com/articles/trump-lifts-ai-chip-ban-china-nvidia

• Council on Foreign Relations. China's AI Chip Deficit: Why Huawei Can't Catch Nvidia (Dec 2025).
https://www.cfr.org/articles/chinas-ai-chip-deficit-why-huawei-cant-catch-nvidia-and-us-export-controls-should-remain

• Data Center Knowledge. AI Chip Export Controls: A New Challenge for Data Center Operators.
https://www.datacenterknowledge.com/data-center-chips/ai-chip-export-controls-a-new-challenge-for-data-center-operators

India Semiconductor Mission

• India Semiconductor Mission. Official portal.
https://ism.gov.in/

• Press Information Bureau India. India Semiconductor Mission 2.0 (Feb 2026).
https://www.pib.gov.in/PressReleasePage.aspx?PRID=2224839

• IMARC Engineering. Semiconductor Manufacturing in India 2026 Expansion.
https://www.imarcengineering.com/news/semiconductor-manufacturing-in-india-expansion

• India Briefing. Semiconductor Industry in India: Incentives and Key Players (May 2026).
https://www.india-briefing.com/news/setting-up-a-semiconductor-fabrication-plant-in-india-what-foreign-investors-should-know-22009.html/

• India's World. India's Semiconductor Mission 2.0: From Fabs to Ecosystems.
https://indiasworld.in/indias-semiconductor-mission-2-0-from-fabs-to-ecosystems/

TSMC, ASML, and the technology roadmap

• Data Center Dynamics. TSMC updates its roadmap with three new process nodes and breaks ground on an Arizona packaging facility.
https://www.datacenterdynamics.com/en/news/tsmc-updates-roadmap-with-three-new-process-nodes-breaks-ground-on-arizona-packaging-facility/

• EE Times. TSMC Unfolds Map for Process, Packaging Tech (2026 NA Tech Symposium).
https://www.eetimes.com/tsmc-unfolds-map-for-process-packaging-tech/

• TrendForce. TSMC Speeds Advanced Packaging: AP7 Targets 2026 Output; Arizona P6 Eyed for US

About Us

This paper, published by the TMT Industry Solutions Group (ISG) at Zensar, presents our perspective on the global trends shaping semiconductors companies in 2026 and beyond.

Who is Zensar?

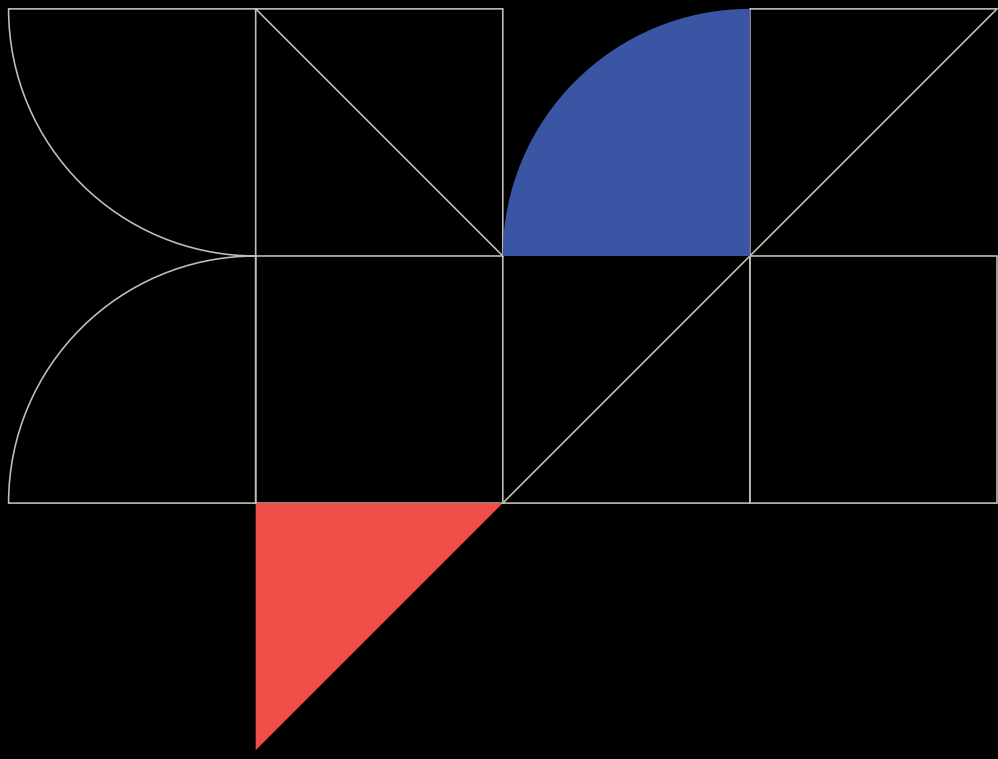
Zensar stands out as a premier technology consulting and services company, embracing an experience-led everything philosophy. We are creators, thinkers, and problem solvers, passionate about designing digital experiences that we engineer into scale-ready solutions to deliver superior engagement for high-growth companies.

10,000+ employees
30+ locations worldwide
130+ global clients
$624.5Mn FY25 revenue

Who is ISG?

ISG (Industry Solutions Group) is the cornerstone of our engagements at Zensar, differentiating with "domain-driven everything," innovation, and empowering client engagement.

- **Rituraj Patil**, Director- TMT Industry Solutions Group
- **Chanchalakshi Jatariya**, Sr Business Consultant TMT ISG
- **Mukul Vashist**, GMR TMT ISG
- **Pranjali Bhople**, Business Analyst TMT ISG



zensar

An RPG Company

At Zensar, we're 'experience-led everything.' We are committed to conceptualizing, designing, engineering, marketing, and managing digital solutions and experiences for over 145 leading enterprises. Using our 3Es of experience, engineering, and engagement, we harness the power of technology, creativity, and insight to deliver impact.

Part of the $4.8 billion RPG Group, we are headquartered in Pune, India. Our 10,000+ employees work across 30+ locations worldwide, including Milpitas, Seattle, Princeton, Cape Town, London, Zurich, Singapore, and Mexico City.

For more information, please contact: info@zensar.com | www.zensar.com