IT Hardware

AI leaving other applications far behind

Key message

1. Foundry – Cycle to peak in 2025F; advanced node supply tight; oversupply in mature nodes.
2. Fabless – Non-AI fabless sales growth in 2025F dependent on spec upgrades.
3. Testing interface – Strong AI demand the key driver.
4. IP – Long-term growth secured on China demand.

Impact

Foundry – Cycle to peak in 2025F; advanced node supply tight; oversupply in mature nodes. In 2025F, we expect the global semiconductor sector to see another strong year, with sales estimated at US$731.6bn (up 18% YoY), thanks to strong AI demand and moderate recovery of non-AI applications. We believe the cycle will likely peak next year, in line with historical patterns. In addition, due to strong AI demand, mobile flagship SoC 3nm migration and fewer competitors in N3, N4 and N5, we expect N3, N4 and N5 utilization rates and also for CoWoS will remain high throughout the year, whereas mature nodes may see greater pricing pressure due to aggressive capacity expansion at Chinese foundries.

Fabless – Non-AI fabless sales growth in 2025F dependent on spec upgrades. Most of Taiwan's fabless plays believe that end-market demand for consumer electronics will see healthy recovery next year. Without the strong inventory restocking seen in 1H24, fabless sales growth next year will rely heavily on process and specification upgrades. However, China's self-sufficiency drive, particularly in areas like DDI, optical sensor, Wi-Fi 5/ 6, and low-speed Ethernet, could weigh on fabless earnings growth. We recommend a strategic focus on firms benefiting from spec upgrades, or those aggressively seeking out new customers and expanding markets.

Testing interface – Strong AI demand the key driver. Due to the increasing complexity of AI and HPC chip design, requirements for the testing environment and the time required for testing AI and HPC-related chips are increasing. This is further driving demand for wafer sorting (chip probing), final testing, burn-in testing, and system-level testing. We also note that the current product cycle of Nvidia's (US) AI chips has been shortened to 1.0-1.5 years, versus 2-3 years in the past. We attribute this to: (1) major CSPs requiring GPUs with stronger computing power, as they are still engaged in an AI arms race; and (2) Nvidia ramping up product releases as the AI market continues to expand. AMD (US) is also developing its MI-series of GPUs, while major cloud service providers continue to develop AI ASIC projects in a bid to undercut Nvidia's market share. We believe resilient AI demand and structural trends such as higher testing requirements for AI and HPC chips will be key long-term growth drivers for testing interface vendors.

IP – Long-term growth secured on China demand. We are positive on the long-term growth prospects of the IP industry, driven by the shift toward advanced process nodes, which is significantly increasing design complexity and R&D costs. Additionally, China's push for technological self-reliance is fueling local demand for non-western third-party IP suppliers. Given the underdeveloped state of China's IP ecosystem, we believe Taiwanese IP suppliers stand out as key beneficiaries. Looking ahead to 2025F, with 2nm capacity coming on line, related IP licensing underway, and continued localization and wafer fab expansion in China, we are particularly positive on third-party IP suppliers with strong advanced-node offerings and high revenue contribution from China.

Stocks for Action

TSMC (2330 TT, NT$1,000, OP) is our top pick in the semiconductor sector due to its leadership position in technology and earnings certainty. Moreover, as capex continues to be raised (particularly in advanced packaging and 2nm), we see all equipment vendors, IP suppliers and testing inference plays benefit from TAM growth. Our top picks are Winway (6515 TT, NT$1,170, OP), MPI (6223 TT, NT$760, OP) and M31 (6643 TT, NT$680, OP). In view of AI boom, we are also positive on AI ASIC development, with both MediaTek (2454 TT, NT$1,250, OP) and Alchip (3661 TT, NT$2,160, OP) benefiting from the take-off of AI inference demand. We also believe Andes (6533 TT, NT$361.5, OP) will benefit from RISC-V penetration growth. We also like Faraday (3035 TT, NT$217, OP) due to its long-term business transformation increasing RFQ demand and to its design-wins.

Risks

Slowdown of AI spending and demand; economic recession.

2025F semiconductor sector outlook

Summary

1. We expect the semiconductor cycle to peak in 2025 and maintain the cadence of two years of expansion, followed by two years of contraction.

2. Differing from previous cycles, semiconductor sales growth will exceed capex growth in the current cycle, which indicates semiconductor companies may have room to increase cash payout ratios.

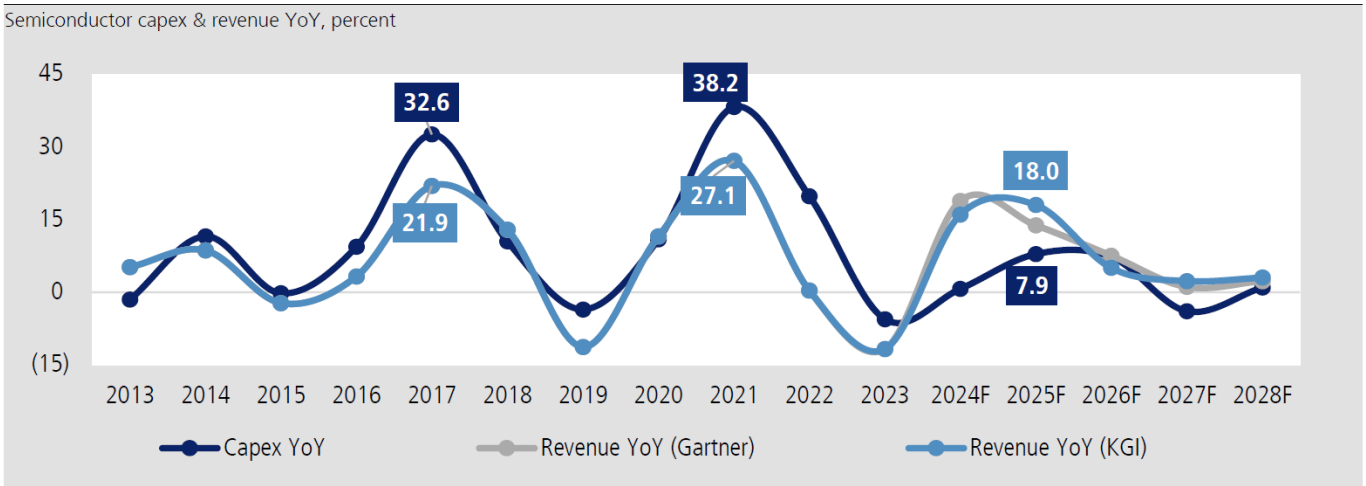
3. Historically, the cycle peak turning point is caused by a decline in demand, whereas the turning point of the trough is the result of supply or production cuts.

The big picture – Semiconductor cycle likely to peak in 2025F

In the past two decades, semiconductor cycles have maintained the cadence of two years of expansion, followed by two years of contraction. During cycle troughs, most semiconductor companies over-cut capex due to margin contraction or insufficient cash flow. When demand began to recover, restocking resulted in capacity shortages in the following year. Moreover, semiconductor firms will use higher-than-normal demand in year one as a base to plan year two's capex. As a result, capacity over-expansion forces the semiconductor sector into a downcycle when demand normalizes or even weakens more than previously expected. In general, from historical patterns, normally the semiconductor cycle peak inflection point is the result of declining demand, whereas the turning point at the trough is usually driven by production cuts or capex discipline.

Looking into the 2025F or current cycle, the key difference in the current cycle from previous cycles is semiconductor revenue growth will surpass capex growth. We attribute this to fewer competitors for advanced nodes heading into the 3nm (or lower) era. TSMC's pure foundry model, leading technological position, and positive venture into advanced packaging ahead of the AI boom have helped it widen its lead over competitors, such as Samsung (KR) and Intel (US). As the competition landscape has largely settled, competing firms will likely scale down investment unless they have significant breakthroughs, either on the client or technology side. In addition, with the Sino-US tech-war escalating, China's foundries may find difficulties purchasing equipment from western suppliers. ASML's (NL) downward revision of its 2025 outlook reflects the slowing of Samsung and Intel's investment as TSMC dominates advanced node business.

Figure 1: Semiconductor revenue & capex YoY since 2013



Source: Gartner; KGI Research estimates

On the other hand, heading into the 3nm and 2nm era, the cost of IC design and fabrication is getting higher for migration at each node. This is the other reason why semiconductor revenue growth will be higher than capex growth in the current cycle.

Breaking down by application, we expect strong AI demand to continue to drive cycle growth. Recovery of demand for non-AI (PC, smartphone, and consumer) applications will remain slow, but on-device AI will drive silicon content growth of 10% YoY, according to TSMC, in 2025F. Recovery of demand for non-AI (PC, smartphone, and consumer) applications will remain slow, but on-device AI will drive silicon content growth by 10% YoY according to TSMC. Due to a lack of killer applications and that most AI PCs and smartphones carry premium pricing, we don't expect on-device AI to result in a 50-60% penetration rate as OEMs expect to happen within two years. However, with an extremely low base in 2024, strong YoY growth in 2025 can still be expected. In general, we expect AI to maintain strong growth, moderate recovery for non-AI applications, and silicon content growth to underpin sector growth in 2025.

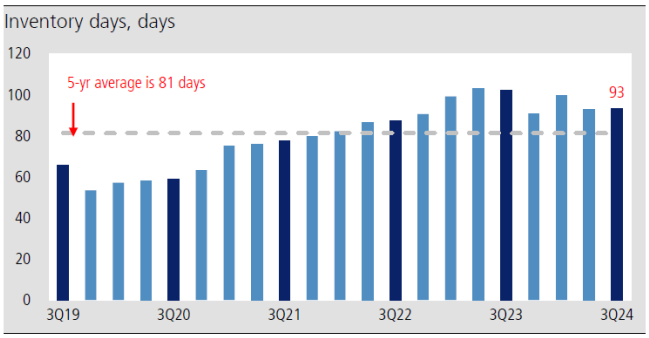
According to Gartner, automotive and data processing (AI) semiconductor sales will outperform in the next few years while communication and consumer electronics semiconductor business will underperform. We attribute outperformance of the auto business to: (1) a low base in 2024F due to inventory adjustments; (2) the megatrend of the digitalization of automobiles (including EV); and (3) global interest rate cuts spurring an auto demand recovery and inventory adjustments. However, AI still leads the pack among all semiconductor applications, boasting a much stronger sales CAGR through 2028F (according to Gartner), in line with our view.

Figure 2: Semiconductor market revenue breakdown by application

End market forecast (US$mn)	2023	2024F	2025F	2026F	2027F	2028F	
Automotive Electronics	78,007	83,116	96,039	104,669	110,596	115,048	
Communication Electronics	158,718	191,618	213,898	223,614	217,119	215,736	
Consumer Electronics	55,738	61,290	66,046	68,104	69,552	72,440	
Data Processing Electronics	167,527	226,238	263,879	288,058	287,115	290,302	
Industrial and Military/Civil Aerospcae Electronics	69,975	67,533	76,860	86,335	94,738	103,474	
Total	**529,964**	**629,796**	**716,723**	**770,779**	**779,120**	**797,000**	
YoY growth rate	**2023**	**2024F**	**2025F**	**2026F**	**2027F**	**2028F**	**2023-28 CAGR**
Automotive Electronics	*14.0%*	*6.6%*	*15.5%*	*9.0%*	*5.7%*	*4.0%*	*8.1%*
Communication Electronics	*-17.7%*	*20.7%*	*11.6%*	*4.5%*	*-2.9%*	*-0.6%*	*6.3%*
Consumer Electronics	*-16.6%*	*10.0%*	*7.8%*	*3.1%*	*2.1%*	*4.2%*	*5.4%*
Data Processing Electronics	*-15.6%*	*35.0%*	*16.6%*	*9.2%*	*-0.3%*	*1.1%*	*11.6%*
Industrial and Military/Civil Aerospcae Electronics	*-5.0%*	*-3.5%*	*13.8%*	*12.3%*	*9.7%*	*9.2%*	*8.1%*
Total	***-11.7%***	***18.8%***	***13.8%***	***7.5%***	***1.1%***	***2.3%***	***8.5%***

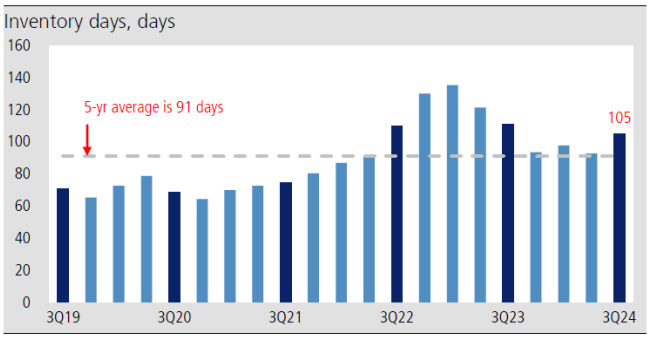
Source: Gartner; KGI Research

Figure 3: Global foundry inventory days



Source: Bloomberg; TEJ; KGI Research

Figure 4: Global fabless inventory days



Source: Bloomberg; TEJ; KGI Research

2025F outlook by sub-sector

Summary

1. TSMC's positive outlook and CSP solid capex is due to strong AI demand. However, strong AI demand has overshadowed non-AI application demand due to weak consumption recovery.

2. For non-AI applications, we expect only a moderate recovery and see no signs of strong restocking demand in 1H25F (as seen in 1H24). Edge AI and spec upgrades are key drivers for edge devices and fabless.

3. AI demand and TSMC's market share expansion will keep N3, N4, N5 and CoWoS utilization rates at higher levels through 2025F. Continual upward revisions of capex should benefit related equipment and testing interface supply chains.

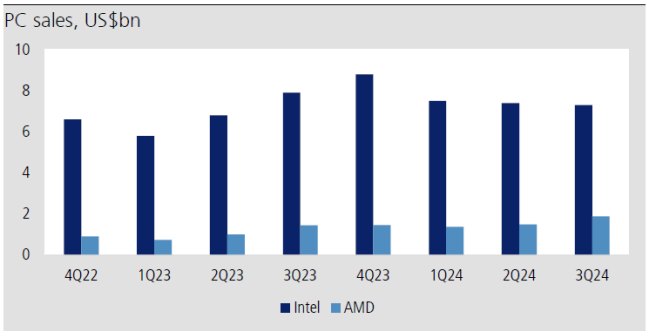
PC semiconductor – Slow recovery; Intel is losing market share to AMD & ARM

Overall PC recovery has been slow in 2024, and we have yet to see any signs of strong growth for 2025, despite PC OEMs maintaining hope for AI PC growth and a corporate replacement cycle due to Windows 10 EOL. In 2025, despite limited shipment growth (we forecast 4% YoY growth for downstream firms), we expect spec upgrades and rising silicon content in AI PC to drive blended ASP increase. However, capacity oversupply for mature nodes, due to aggressive capacity expansion by China foundries and slow demand recovery, may help PC semiconductor firms achieve margin expansion. That said, leading PC semiconductor companies, such as Realtek (2379 TT, NT$477.5, N) note pricing competition will likely offset the benefit of lower foundry pricing.

On the CPU front, due to a lack of killer applications and compatibility issues for ARM (UK)-based CPUs, Qualcomm's (US) X Elite (AI PC) sales are below expectations. In addition, a lawsuit between ARM and Qualcomm has shaken the confidence of the ARM-based camp. ARM claims Qualcomm infringed its patents on ARM architecture in AI PC solutions, while Qualcomm believes its acquisition of Nuvia (US) allows it to waive licensing fees to ARM by using Nuvia's architecture. The x86 camp may view this as a victory for x86 CPU in the AI PC space.

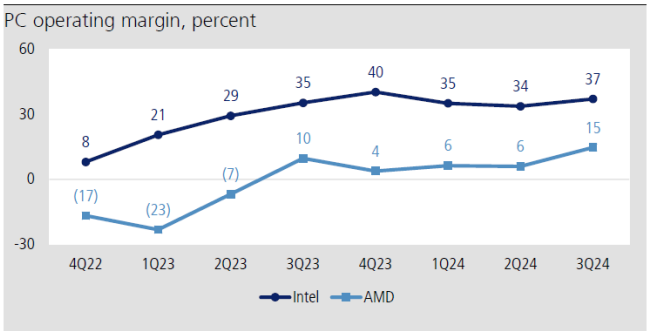
Lastly, with the help of TSMC, AMD has continued to narrow the gap between itself and Intel for both PC and server CPU. However, financial difficulties may force Intel to dispose of its non-core business, restructure, and increase outsourcing to TSMC. Following Lunar Lake (3nm), Intel will continue to increase outsourcing of Arrow Lake to TSMC. As Intel cancelled its 20A process for Arrow Lake, we expect TSMC to receive all Arrow Lake business in 2025. We believe TSMC will help Intel slow market share contraction within the PC space.

Figure 5: Intel vs. AMD PC sales



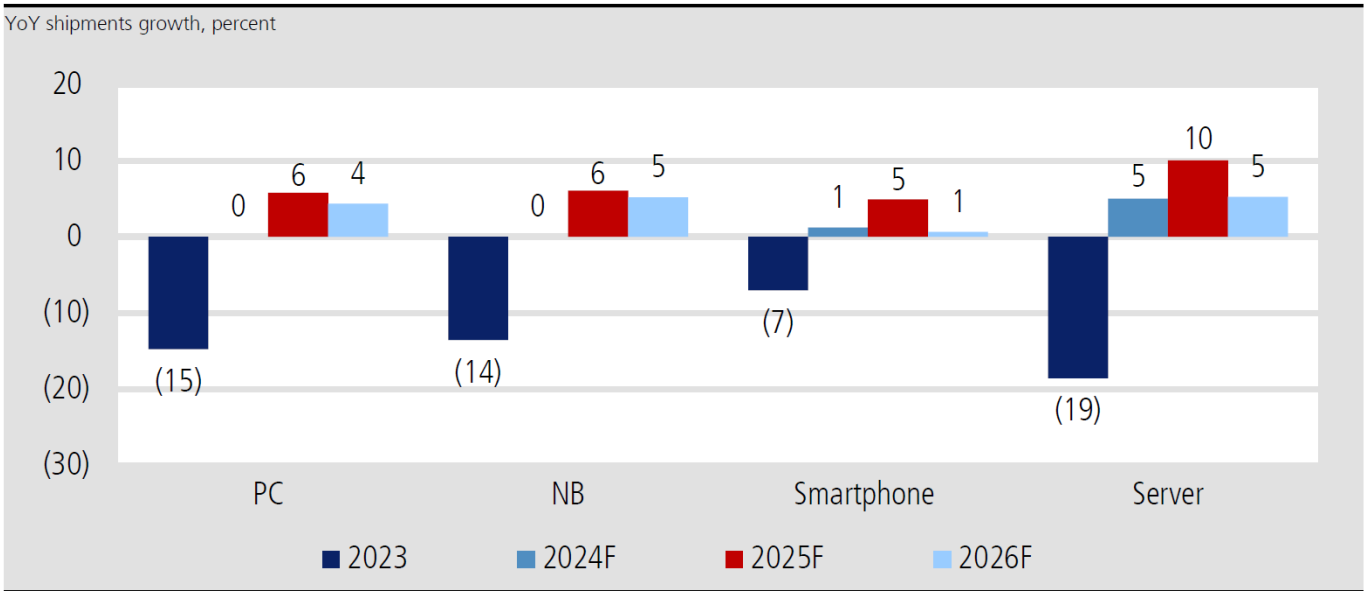
Source: Company data; KGI Research

Figure 6: Intel vs. AMD PC operating margin



Source: Company data; KGI Research

Figure 7: IT hardware devices – NB, PC, smartphone & server shipments to all grow in 2025F

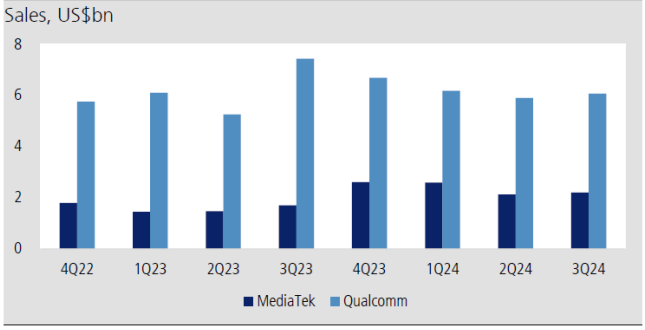


Source: Gartner; KGI Research estimates

Smartphone SoC/ AP – No sign of strong restocking demand

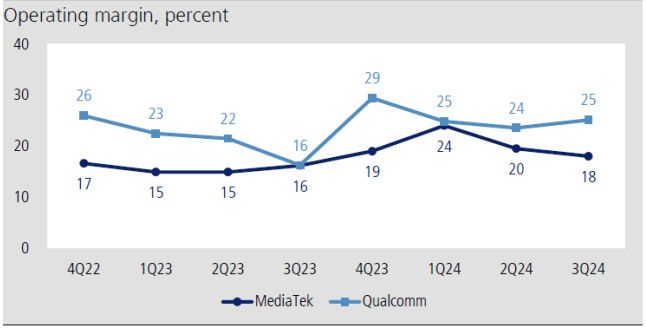
We forecast total 2025 smartphone SoC/ AP shipments of 1.24bn units, up 4% YoY, which indicates SoC/ AP shipments look to surpass sell-through of 1.16bn units, meaning this phenomenon will have occurred for two years in a row. Thus, despite current inventory levels being manageable, the chance of strong restocking demand, like we saw in 4Q23 or 1Q24, is low. As for flagship SoC, both MediaTek (D9400) and Qualcomm (SDM8750) launched new 3nm SoC in 4Q24. We expect MediaTek to continue to expand its market share within the flagship segment due to: (1) better performance-to-cost ratio; and (2) its clients' product mixes are skewed toward high-end products due to higher margins.

Figure 8: MediaTek mobile vs. Qualcomm handset sales



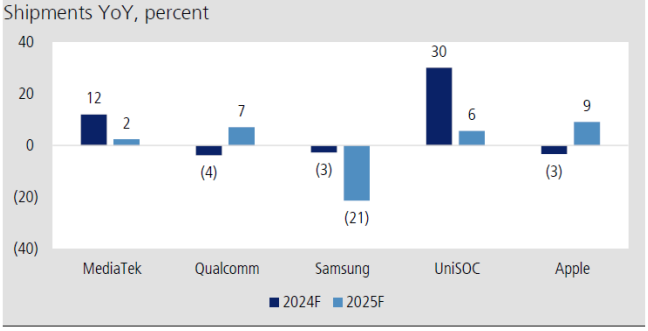
Source: Company data; KGI Research

Figure 9: MediaTek vs. Qualcomm operating margin



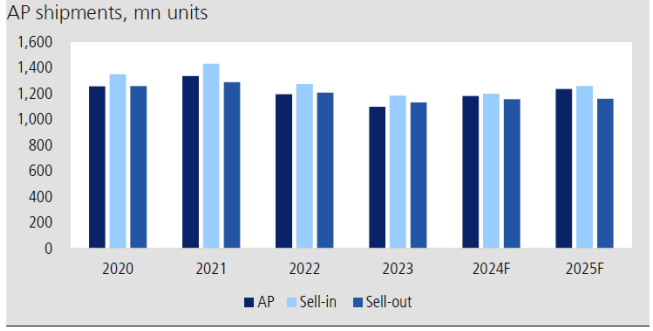
Source: Company data; KGI Research

Figure 10: Mobile SoC/ AP shipments growth



Source: KGI Research estimates

Figure 11: AP shipments vs. smartphone (set) shipments

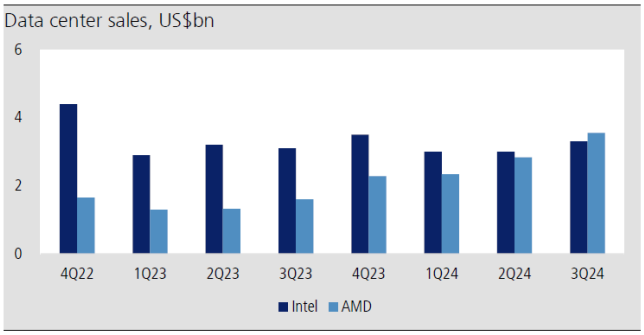


Source: KGI Research estimates

Data center – General server starting to recover; AMD & Intel gap narrowing

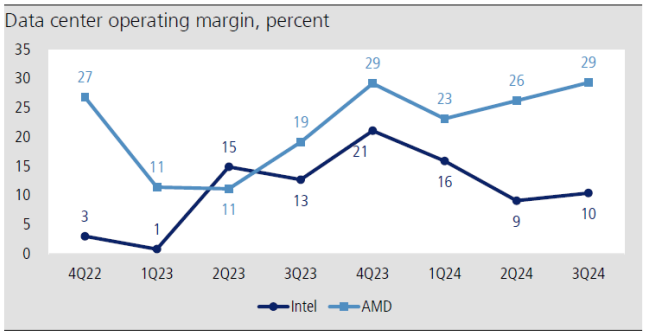
In 3Q24, AMD's data center revenue came in at US$3.5bn, up 25% QoQ and 122% YoY, surpassing Intel's for the first time thanks to better performance-to-cost ratio and much stronger AI GPU sales. Differing from PC applications, most of Intel's server CPU are still produced at its foundry. Therefore, with TSMC's help and better performance-to-cost ratio, AMD's Milan and Genoa solutions continue to gain the market share from Intel. Due to strong AI demand, skyrocketing costs have cannibalized the budget for conventional servers, forcing CSPs to adopt AMD solutions instead of those from Intel. As a result, we forecast AMD's market share for server CPU will expand from 30% in 2023 to 40% by end-2024. In addition, AMD revised up its 2024 AI chip sales guidance from US$4.5bn to US$5.0bn. Meanwhile, Intel has had difficulties achieving its target for Gaudi sales of US$500mn.

Figure 12: Intel vs. AMD data center sales



Source: Company data; KGI Research

Figure 13: Intel vs. AMD data center operating margin

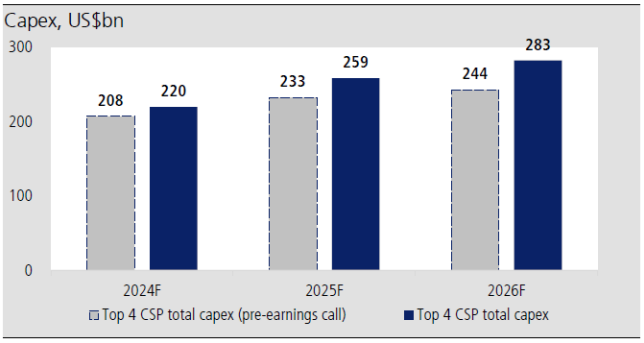


Source: Company data; KGI Research

AI – No signs of slowing down

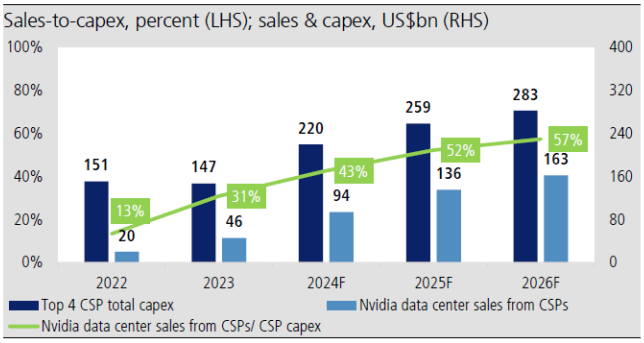
The launch of ChatGPT is driving an AI boom, which will start to expand from the cloud to the edge business in 2025F. Amid concerns about training server and GPU demand, we are yet to see any sign of the AI boom slowing down. In 3Q24, we continued to see upward revisions of the top four CSPs' capex in 2024F and 2025F. These solid upward revisions are key to supporting strong AI demand growth. Indeed, while it is not rational to assume unlimited growth for CSPs' capex going forward, we are yet to see signs that AI investments will cool down in 2025F, and CSPs' cloud business growth remains strong. Due to the delay of GB200, we expect to see very strong rack (GB200) sales at Nvidia in 1H25F. Moreover, Nvidia has pulled in B300 and B300A schedules a bit, and we expect both to enter mass production in June next year. With better GPU performance (30-40% higher) and higher HBM density (8xHBM3e12Hi), CoWoS supply may remain tight through next year.

Figure 14: Top-four CSP capex revisions post-3Q24 earnings



Source: Bloomberg; KGI Research estimates

Figure 15: Nvidia's data center sales-to-CSP capex



Source: Bloomberg; KGI Research estimates

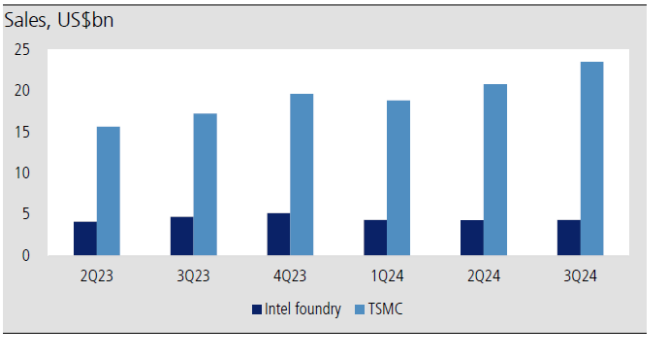
Foundry - TSMC dominates advanced nodes; pricing pressure on mature nodes

With growth slowing for both Samsung and Intel's foundry businesses, TSMC has widened its lead over competitors for advanced nodes, and we don't think the gap will narrow in the foreseeable future due to TSMC's leading technological position and unique foundry model. Meanwhile, the strength of its advanced packaging business, along with strong AI demand, has also increased TSMC's lead. Skyrocketing AI demand has propelled both front-end business and back-end CoWoS growth. With several rounds of price hikes due to capacity shortages, the gross margin of CoWoS is close to TSMC's corporate average. In 2025, we expect TSMC's CoWoS revenue to grow more than 100% YoY again, and Nvidia remains the key contributor, not only for volume, but also as the primary supporter of TSMC's rising price quotes.

In addition, on the back of mobile flagship SoC 3nm migration (Qualcomm and MediaTek), strong Apple (US) demand and outsourcing orders from Intel, we don't expect to see a slow season for N3 utilization. Meanwhile, strong AI demand is driving up N4 and N5 utilization rates. Together with higher costs from overseas expansion and inflated costs (electricity), we expect TSMC will increase prices of N3, N4 and N5 by 3-7% in 2025F, while N6, N7 and mature node prices (12-inch) will be flattish. In mature nodes (8-inch), due to low utilization and oversupply, we forecast TSMC to cut prices by 5-10% and believe it will further pressure tier-2 foundries' margins.

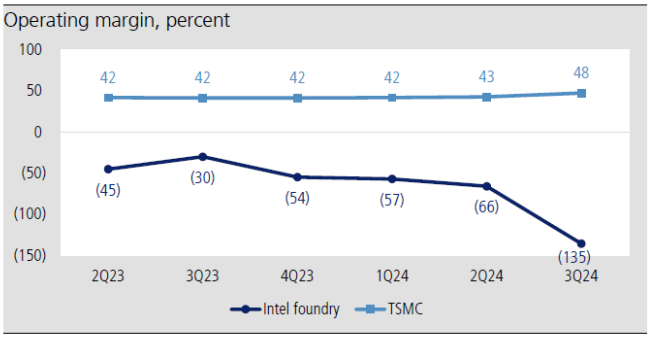
The only concern we have about TSMC is the potential for an escalating Sino-US tech war on Trump's return to the White House. While we don't believe political noise will structurally change TSMC's leading industry position, more restrictive export controls on China may still negatively affect the firm's margins.

Figure 16: Intel's foundry vs. TSMC's sales



Source: Company data; KGI Research

Figure 17: Intel's foundry vs. TSMC's operating margin



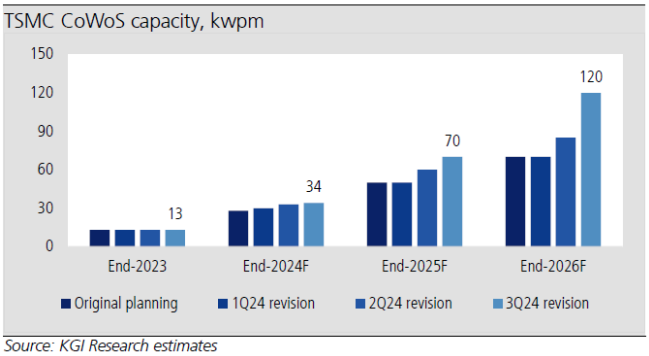
Source: Company data; KGI Research

OSAT – Advanced packaging the only focus

CoWoS has been attracting much market attention in 2023-25F thanks to the AI frenzy. However, CoWoS was only a niche technology at TSMC, while InFO represented the bulk of advanced packaging before 2023. With skyrocketing demand for AI GPU and aggressive CoWoS expansion, all next-generation packaging methods are also now getting market attention (e.g. SoIC/ 3D, WMCM, FOPLP). In 2025F, CoWoS will continue to be the focus. With the acquisition of Innolux's (3481 TT, NT$15.1, N) Tainan fab, we forecast TSMC's CoWoS capacity will grow from 34 thousand wafers per month (kwpm) by end-2024F to 70-75kwpm by end-2025F. Annual supply could grow from 310k wafers in 2024F to 660k wafers in 2025F (up 112% YoY). Among clients, Nvidia is the largest, absorbing 60-61% of CoWoS supply, followed by AMD, Broadcom (US) and Marvell (US).

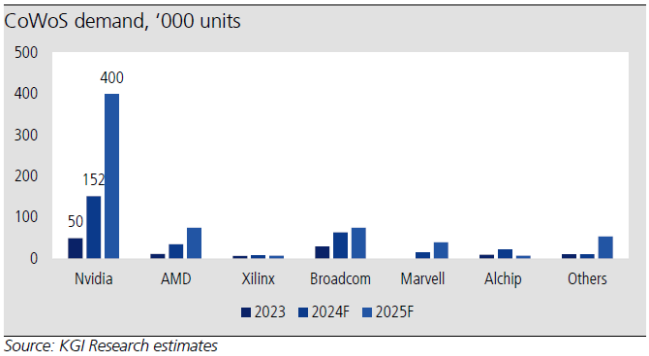
For next-generation CoWoS-L (used in Blackwell), TSMC's capacity could grow from 30-40k wafers in 2024F to 300k in 2025F. Based on our calculations, we forecast every 10kwpm of CoWoS capacity will contribute 1.0-1.2% of TSMC's sales, depending on prices and respective weightings of CoW-S and CoW-L.

Figure 18: TSMC's CoWoS capacity forecast revisions



Source: KGI Research estimates

Figure 19: 2023-25F CoWoS demand forecasts



Source: KGI Research estimates

Figure 20: Overview of TSMC's advanced packaging fabs

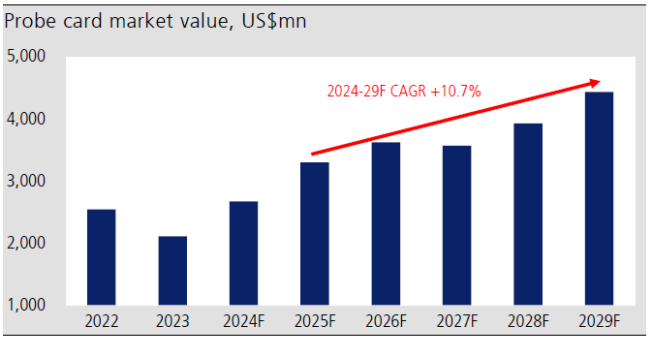
TSMC AP fab	Location	Bumping	InFO/WMCM	CoW-S	CoW-L	oS	SoIC	Status
AP1 (RD)	Hsinchu	V		V	V	V	V	Online
AP2	Tainan	V						Online
AP3	Longtan		V	V	V			Online
AP5	Taichung		V	V	V	V		Join from 4Q24
AP6	Zhunam		V	V	V		V	Online
AP7 (P1)	Chiayi			V	V			P1 potentially delays, due to unearthed histrocial site
AP7 (P2)	Chiayi		V					P2 on track, Contruction done in 2026
AP8	Tainan (Innolux)			V	V			Initially 10Kwpm for CoW (L) (3Q25)

Source: KGI Research estimates

Testing interface – AI & HPC testing a long-term growth driver

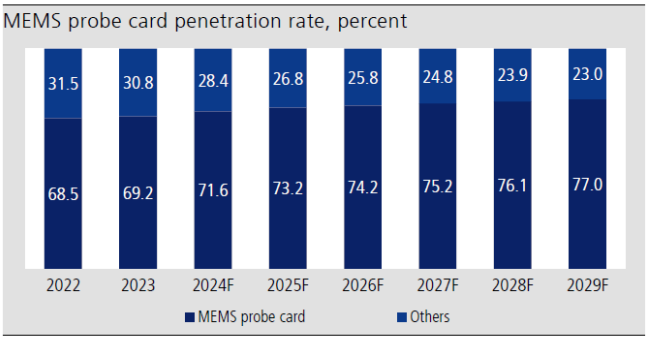
In 2024, the testing interface market is set to benefit from a moderate recovery in end-market demand, alongside strong and sustained demand for AI and HPC, driving the need for probe card and IC test socket. According to research firm TechInsights, 2024F global revenues of probe card and IC test/ burn-in socket will grow by 26.2% and 30.2%, reaching US$2.67bn and US$2.13bn, respectively. Looking ahead to 2025F, as end-market demand continues to recover and AI and HPC remain in a state of undersupply, we expect the testing interface market to expand further. TechInsights projects global revenues of probe card and IC test/ burn-in socket to grow by 23.6% and 16.7%, reaching US$3.3bn and US$2.48bn, respectively, in 2025F, with a CAGR of 10.7% and 8.0% from 2024-29F. Additionally, TechInsights forecasts that the global MEMS probe card market will achieve a CAGR of 12.3% from 2024-29F, outpacing overall probe card market growth. We believe this is mainly due to: (1) the use of MEMS probe card in AI chips; and (2) as the manufacturing process advances to 3nm or even more advanced nodes, Cobra probes are likely to damage the device under test, while MEMS probe card offers an advantage in terms of force.

Figure 21: 2024-29F global probe card market value CAGR will reach 10.7%



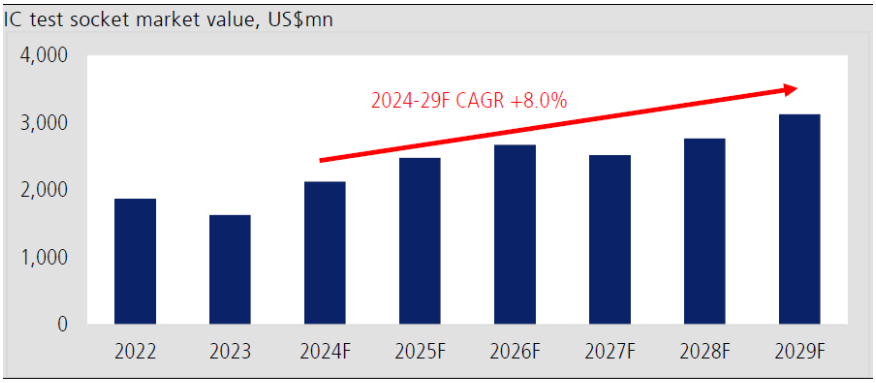
Source: TechInsights; KGI Research estimates

Figure 22: Penetration rate of MEMS probe card will increase with process advances



Source: TechInsights; KGI Research estimates

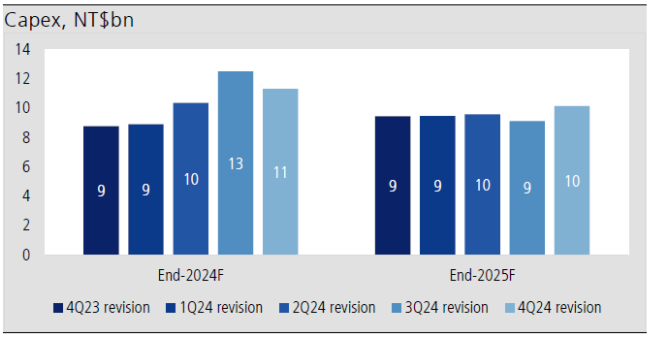
Figure 23: Global IC test socket market value CAGR will reach 8.0% 2024-29F



Source: TechInsights; KGI Research estimates

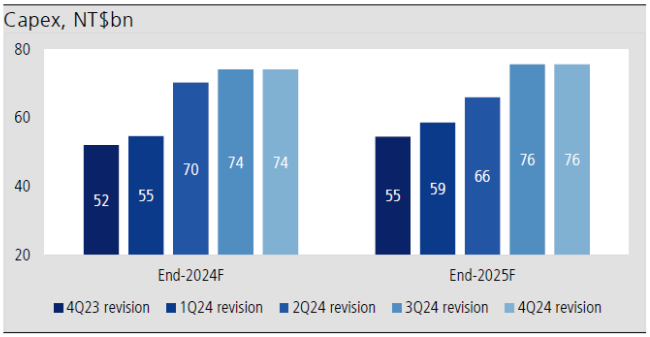
Notably, due to the increasing complexity of AI and HPC chip design, the requirements for the testing environment and testing time for AI and HPC-related chips are also rising. This has further driven demand for wafer testing (chip probing), final test, burn-in test, and system-level test. Advanced test equipment manufacturer Advantest (JP) recently said at its earnings call that the increased complexity of AI and HPC testing will stimulate demand for test equipment, highlighting especially strong demand for testing equipment from TSMC and Taiwanese packaging and testing companies. More testing equipment will be needed to meet robust demand for AI and HPC. We believe this structural trend will be the main growth driver for test interface providers over the next 3–5 years. Packaging and testing companies, including ASEH (3711 TT, NT$150.5, N) and King Yuan (2449 TT, NT$123, NR), have both revised up 2024F capital expenditure estimates for the second time this year. ASE's 2024F capex estimate was revised up from NT$55bn at the beginning of the year to NT$74bn, up over 35% YoY, primarily to expand advanced packaging and testing capacity. King Yuan Electronics raised its capex from NT$9bn to NT$11bn, up over 40% YoY, indicating that current testing capacity remains tight in light of increased testing demand and testing time for AI and HPC. Looking ahead to 2025–26F, Nvidia's next-generation AI GPU, the Rubin series, is expected to enter mass production by the end of 2025F or early 2026F. Major cloud service providers such as Meta (US), Microsoft (US), AWS (US), and Alphabet (US) all have AI ASIC projects slated for mass production between 2025 and 2026. We believe this trend will benefit test interface providers with leading positions and a high revenue exposure to the AI supply chain.

Figure 24: King Yuan raises 2024-25F capex significantly



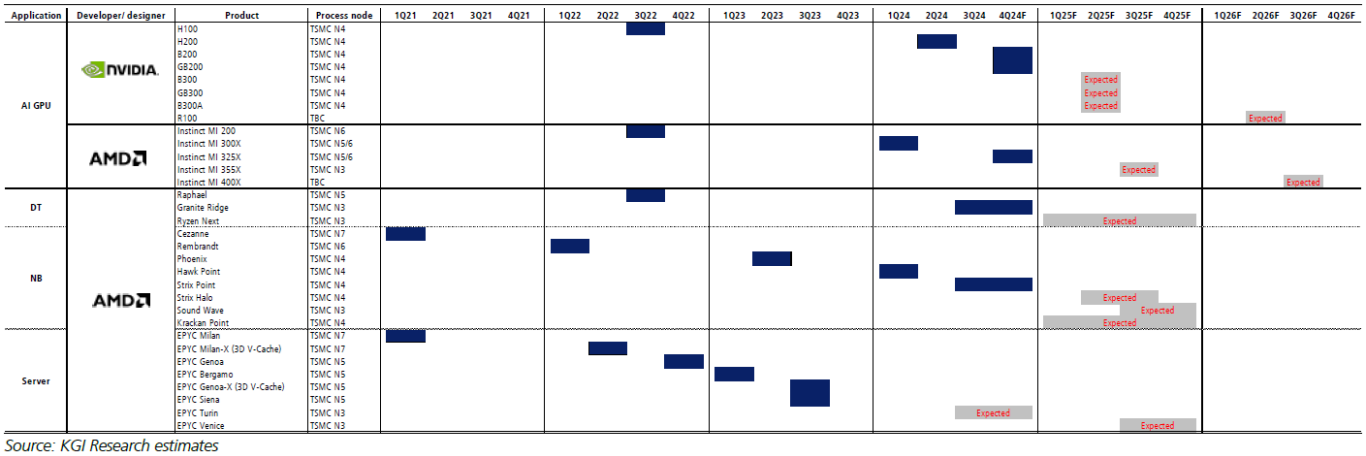
Source: Bloomberg; KGI Research estimates

Figure 25: ASEH's capex up significantly in 2024-25F



Source: Bloomberg; KGI Research estimates

Figure 26: Nvidia & AMD product roadmap



Source: KGI Research estimates

Figure 27: AI ASIC product roadmaps of CSPs

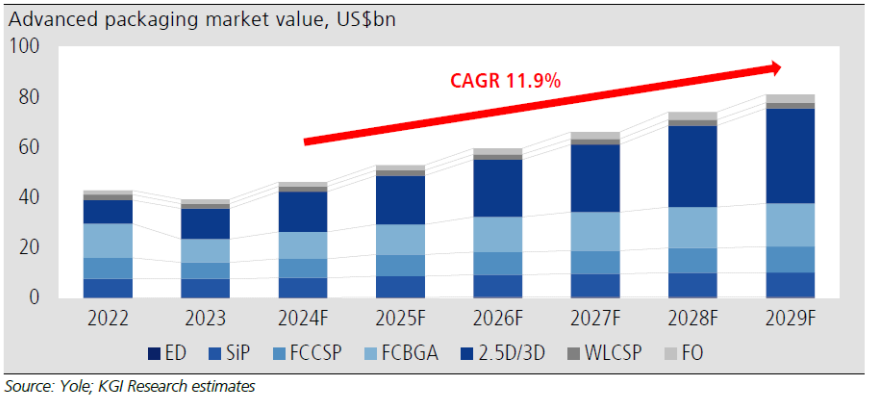
Hyperscaler	Chip	Workload	Process node	ASIC partner	Before 2021	2021 1Q 2Q 3Q 4Q	2022 1Q 2Q 3Q 4Q	2023 1Q 2Q 3Q 4Q	2024F 1Q 2Q 3Q 4Q	2025F 1Q 2Q 3Q 4Q	2026F 1Q 2Q 3Q 4Q
aws	Inferentia	Inference	TSMC 16nm	Alchip	Launched in 4Q19						
	Tranium	Training	TSMC 7nm	Alchip	Launched in 4Q20						
	Inferentia 2	Inference	TSMC 7nm	Alchip				Launched			
	Tranium 2	Training	TSMC 5nm	Marvell					Launched		
	Inferentia 2.5	Inference	TSMC 5nm	Marvell						Expected	
	Tranium 3	Training	TSMC 3nm	Marvell or Alchip (?)							Expected
	Inferentia 3	Inference	TSMC 3nm	Alchip (?)							Expected
	Graviton	Data center CPU	TSMC 16nm	AWS in-house (Annapurna Labs)	Launched in 4Q18						
	Graviton 2	Data center CPU	TSMC 7nm	AWS in-house (Annapurna Labs)	Launched in 4Q19						
	Graviton 3	Data center CPU	TSMC 5nm	AWS in-house (Annapurna Labs)			Launched				
	Graviton 3E	Data center CPU	TSMC 5nm	AWS in-house (Annapurna Labs)				Launched			
	Graviton 4	Data center CPU	TSMC 4nm	AWS in-house (Annapurna Labs)					Launched		
	Graviton 5	Data center CPU	TSMC 4nm or 3nm (?)	AWS in-house (Annapurna Labs)						Expected	
	Graviton 6	Data center CPU	TSMC 3nm or 2nm (?)	AWS in-house (Annapurna Labs)							Expected
Microsoft	Azure Maia 100	Training/inference	TSMC 5nm	GUC (turnkey-3 type)					Announced		
	Azure Maia 200	Training/inference	TSMC 3nm	GUC (turnkey-3 type)							Expected
	Azure Cobalt 100	Data center CPU	TSMC 5nm	GUC (turnkey-3 type)					Announced		
	Azure Cobalt 200	Data center CPU	TSMC 3nm	GUC (turnkey-3 type)							Expected
Meta	MTIA v1	Inference	TSMC 7nm	Broadcom				Announced			
	MTIA v2	Training/inference	TSMC 5nm	Broadcom					Announced		
	MTIA v3	Training/inference	TSMC 3nm	Broadcom						Expected	
	ARM-based CPU Gen-1	CPU	TSMC 3nm (?)	Socionext							
Google	TPU v1	Inference	TSMC 28nm	Broadcom	Introduced in 2Q15						
	TPU v2	Training/inference	TSMC 16nm	Broadcom	Introduced in 2Q17						
	TPU v3	Training/inference	TSMC 16nm	Broadcom	Introduced in 2Q18						
	TPU v4i	Inference	TSMC 7nm	Broadcom		Introduced					
	TPU v4	Training/inference	TSMC 7nm	Broadcom		Introduced					
	TPU v5e	Inference	TSMC 5nm	Broadcom				Introduced			
	TPU v5p	Training/inference	TSMC 5nm	Broadcom				Introduced			
	TPU v6	Training/inference	TSMC 3nm	Broadcom					Introduced		
	TPU v7 Training	Training	TSMC 3nm	Broadcom						Expected	
	TPU v7 Inference	Inference	TSMC 3nm	MediaTek						Expected	
	Axion	Data center CPU	TSMC 5nm	Google in-house					Introduced		
	Maple	Data center CPU	TSMC 5nm	Marvell						Expected	
	Axion 2	Data center CPU	TSMC 3nm	GUC (turnkey-3 type)						Expected	

Source: KGI Research estimates

Semiconductor equipment – Aggressive advanced packaging expansion a long-term growth driver

As semiconductor manufacturing processes become more complex, the cost of chip production rises exponentially. Therefore, we believe advanced packaging and heterogeneous integration will become the main solutions to satisfy HPC needs and extend Moore's Law. Global semiconductor giants, such as TSMC, Intel, and Samsung, along with packaging & testing companies, such as ASEH and Amkor (US), are investing heavily in R&D to develop advanced packaging solutions in-house. Yole forecasts advanced packaging market value will register a CAGR of 9.6%, rising from US$42.8bn in 2022 to US$81.1bn in 2029F, amid the 5G, AI, HPC, and autonomous vehicle trends, with a 2.5D/ 3D packaging CAGR of 22.0% outpacing overall advanced packaging market growth, rising to a market value of US$37.8bn in 2029F.

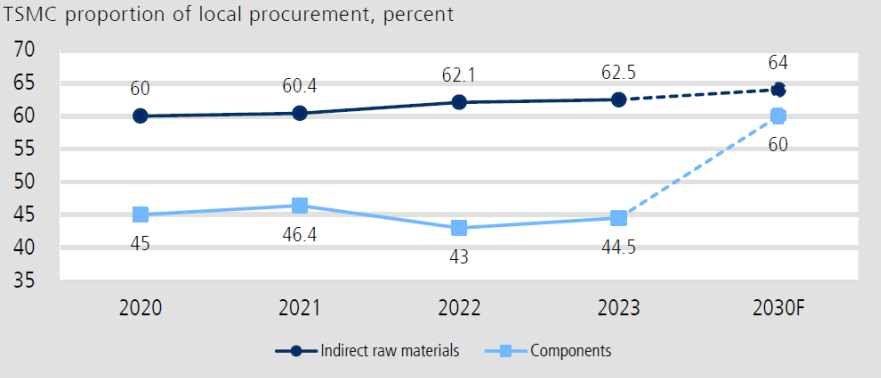
Figure 28: Yole forecasts advanced packaging market value CAGR of 11.9% in 2024-29F



Source: Yole; KGI Research estimates

Additionally, we believe that within the semiconductor equipment supply chain, Taiwanese semiconductor equipment manufacturers primarily play roles in component manufacturing, component development, and maintenance services for front-end process equipment. This is mainly due to the following reasons: (1) patent protections for front-end equipment; (2) a much higher degree of difficulty in production technology than that for back-end equipment; and (3) massive R&D resources required to enter the front-end equipment business. However, the competitive landscape for advanced packaging differs from that of the front-end manufacturing process. We believe global leading foundry vendors intend to support Taiwanese semiconductor equipment manufacturers to diversify supply chains, and the back-end equipment market is easier to enter. According to TSMC's 2022 ESG report, the proportion of indirect raw materials and components procured in Taiwan in 2020 was 60% and 45%, respectively. However, TSMC expects that by 2030F, these proportions will increase to 64% and 60%, respectively. We thus believe this trend will bring structural changes to Taiwanese semiconductor equipment vendors. In addition, advanced packaging is more customized, making it necessary to provide customers with tailored equipment and timely services. Taiwanese semiconductor equipment makers have a strong edge in local services. Furthermore, according to forecasts by SEMI, packaging equipment accounts for only a mid-single digit share of the entire semiconductor equipment capex (including front-end and packaging & testing equipment). We therefore believe the aforementioned overseas giants will put greater focus on more advanced equipment for the front-end manufacturing process, rather than for the back-end process.

Figure 29: TSMC will continue to increase the proportion of indirect raw materials & components procured in Taiwan

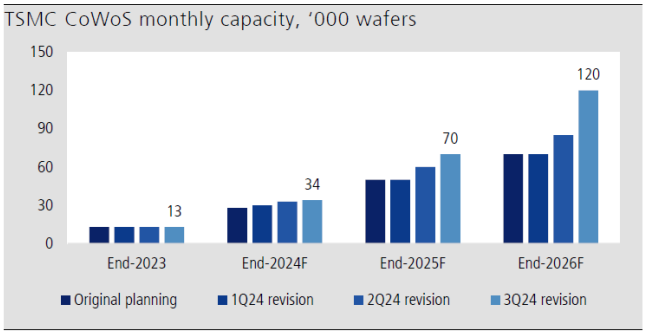


Source: TSMC; KGI Research estimates

At the moment, TSMC's CoWoS capacity mainly services AI applications, including AI server GPU, AI accelerator and AI ASIC, with clients including Nvidia, AMD, Broadcom, Marvell and Taiwanese and Japanese IC design houses. In light of growing demand from AI server applications, we believe the ongoing tightness in CoWoS capacity will persist, on the back of higher requirements for the computing power of AI server, boosting consumption of interposers in terms of area. We expect TSMC's CoWoS capacity to rise to 34kwpm and 70-75kwpm by the end of 2024-25F, respectively. AMD is the only client adopting SoIC as a solution. However, our supply chain checks suggest Apple may also use SoIC capacity to produce M5 CPU (previous generation is produced via InFO_LSI solutions). TSMC's monthly SoIC capacity is 10k wafers, which we estimate will expand to 15k by end-2024F and further to 40k by end-2025F. In a previous tech forum, TSMC also guided that the firm's CoWoS and SoIC capacities would grow at respective 2022-26F CAGRs of over 60% and 100%. Therefore, we believe that over the next 3-5 years, the trend of expanding advanced packaging capacity, along with TSMC's intention to support the local supply chain, will continue to benefit Taiwanese semiconductor equipment companies such as C Sun (2467 TT, NT$204, NR), Grand Process (3131 TT, NT$1,645, NR), Scientech

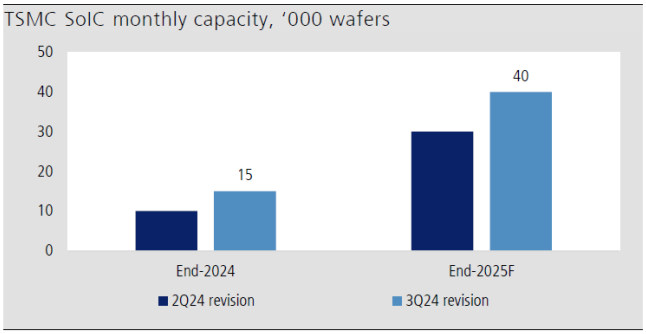
(3583 TT, NT$414.5, NR), Allring (6187 TT, NT$441, NR), and GMM (6640 TT, NT$775, NR).

Figure 30: We forecast TSMC's CoWoS monthly capacity will reach 120k wafers by end-2026F



Source: KGI Research estimates

Figure 31: We estimate TSMC's SoIC monthly capacity will reach 40k wafers by end-2025F

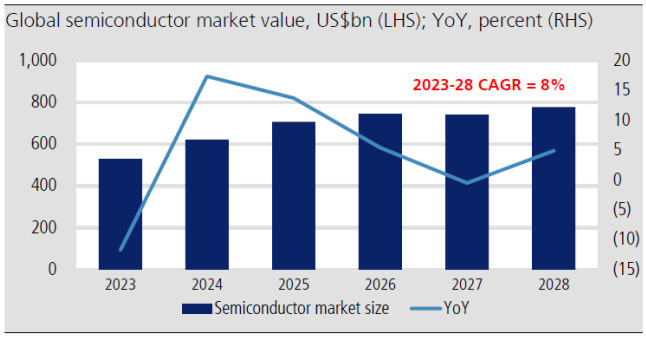


Source: KGI Research estimates

IP – Long-term growth secured, China demand fuels the surge

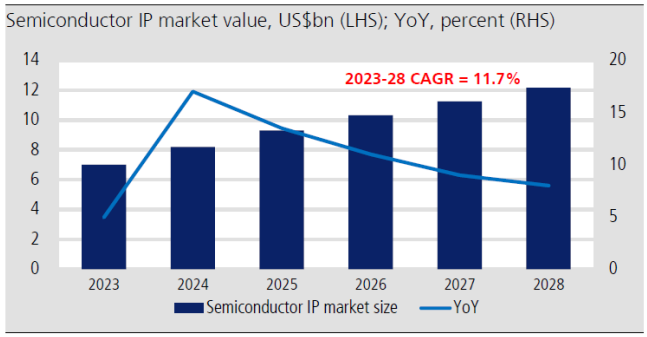
The semiconductor industry's demand for third-party IP providers has significantly increased. According to IPnest, the IP market will grow at a CAGR of 11.7% from 2023 to 2028F, outpacing the overall semiconductor market's growth rate of 8%. This strong growth is driven by several key factors, including: (1) increased design complexity, with a substantial rise in the number of IPs required for more advanced process nodes; (2) cost considerations, as upgrading to more advanced process nodes requires significant investment in research, development, and manufacturing, with R&D costs growing exponentially; and (3) the need for timely product launches and reliability concerns. Developing semiconductor IP in-house is time-consuming, often involving lengthy design, verification, and certification cycles. In contrast, third-party IP companies specialize in developing the latest IPs, providing ongoing improvements and upgrades, and offering higher reliability than in-house development. Given these factors, we hold a positive outlook on the long-term growth prospects of the IP industry.

Figure 32: Gartner estimates global semiconductor market will grow at 2023-28F CAGR of 8%



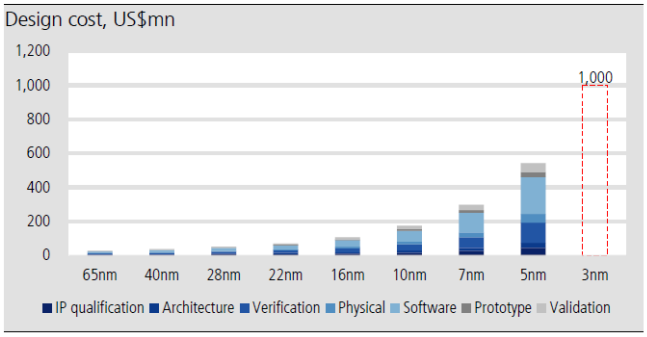
Source: Gartner; KGI Research estimates

Figure 33: IPnest estimates IP market will grow at 2023-28F CAGR of 11.7%, above total semiconductor industry growth



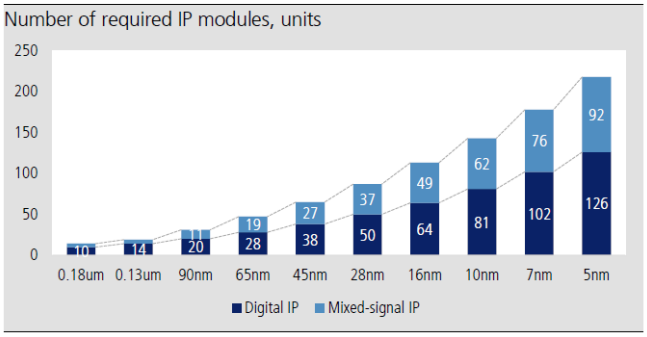
Source: IPnest; KGI Research estimates

Figure 34: Cost of chip design soars when migrating to next-generation nodes



Source: IBS; KGI Research estimates

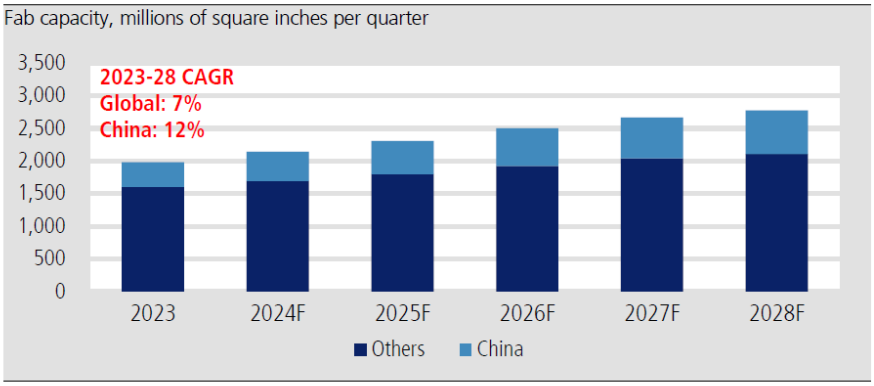
Figure 35: Greater need for logic & analog IC expertise for more advanced nodes



Source: IBS; KGI Research estimates

In addition, China stands out as a major region where non-western third-party IP companies can thrive. While we believe the trend of semiconductor localization in China will bolster local players and may have a negative impact on some IC design companies, China's local IP sector remains relatively underdeveloped. Rising domestic demand and the push for decoupling from western technology are likely to benefit Taiwanese companies with advanced IP technologies. Gartner estimates that global wafer manufacturing capacity (excluding memory) will grow from 1,981mn square inches per quarter (MSI/Q) in 2023 to 2,775 MSI/Q in 2028F, for a CAGR of 7%. During the same period, China's capacity is expected to grow at a CAGR of 12%, increasing its share of global wafer capacity from 19% in 2023 to 24% by 2028F. This rapid growth in China's wafer capacity is driven by a strategic focus on strengthening its domestic semiconductor capabilities. As new capacity comes online, demand for IP will also increase.

Figure 36: Gartner projects China's fab capacity (excluding memory) will post a 2023-28F CAGR of 12%, higher than the global 7%



Source: Gartner; KGI Research estimates

ASIC design service – AI inference to drive mass customization

AI chip development is becoming increasingly complex due to the rise of diverse neural networks and the high computational demands of AI, especially with models growing larger and more varied. Traditionally, we believe CPUs have been unable to meet the power and efficiency needs of AI workloads. Therefore, AI accelerators (e.g. AI GPU and AI ASIC) have emerged to handle these tasks more efficiently, offering high performance and reduced power consumption for certain types of AI models. However, we expect this specialization still comes with trade-offs, particularly a lack of flexibility for future AI model adaptations, especially moving into the AI inference era.

AI training and inference differ significantly. Training involves intensive back-propagation and weight adjustments, which require substantial computing power. AI inference, which is the application of a trained AI model, is less complex but still computationally demanding. Increasingly, AI models require more floating-point operations, and these

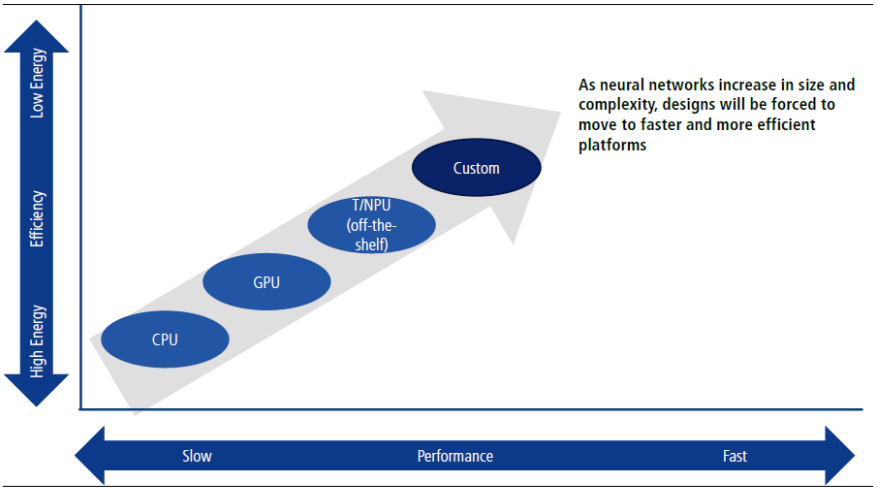
computations are challenging to handle on CPUs due to their energy consumption. GPU, TPU and AI ASIC, which are faster but also power-hungry, are used as alternatives, yet the demand for even faster processing continues to grow.

We have seen that AI is moving from initial applications, like recommendation engines and voice assistance, to more complex tasks in fields such as large language models (LLM) and generative AI. These advancements aim to create human-like responses and adaptable systems. To achieve this, model sizes have increased exponentially, with some models now exceeding a trillion parameters. Initially, AI models focused on a few basic image recognition networks. Today, a range of networks, from vision transformers to generative models, are required, pushing companies toward more programmable solutions to support diverse use cases.

Meanwhile, AI is expanding from data centers to edge devices, where it powers a variety of applications. As the complexity of AI algorithms grows, there is a push toward higher levels of acceleration to meet these demands, yet flexibility remains a priority to keep up with fast-evolving models.

Increasingly, chip designers are considering customized AI engines to address unique model requirements. However, customization introduces challenges, such as a steep learning curve for chip designers and potential delays. On the upside, these customized solutions can provide significant performance and power benefits, as seen in a company that achieved a significant improvement in processing speed and reduced power consumption by half using a tailored AI engine, or even developing their own ASICs. For chip designers, key considerations will remain data storage, data movement, and computation. This focus on efficiency is critical as AI models continue to grow and evolve.

Figure 37: Inferencing is increasing in complexity



Source: Siemens Digital Industries Software; KGI Research

Generative AI, particularly the training LLMs like GPT-4 with 1.76 trillion parameters, requires massive computing resources, traditionally powered by Nvidia GPUs. However, it's not just the GPUs that are needed; power, networking, capital, and data centers with optimal cooling are also essential for running these systems efficiently. We expect that as AI applications grow, especially those requiring fast, local responses, inferencing (the process of making predictions or decisions based on a trained model) demands computing power to be closer to the end-user for low-latency, cost-effective, and energy-efficient results.

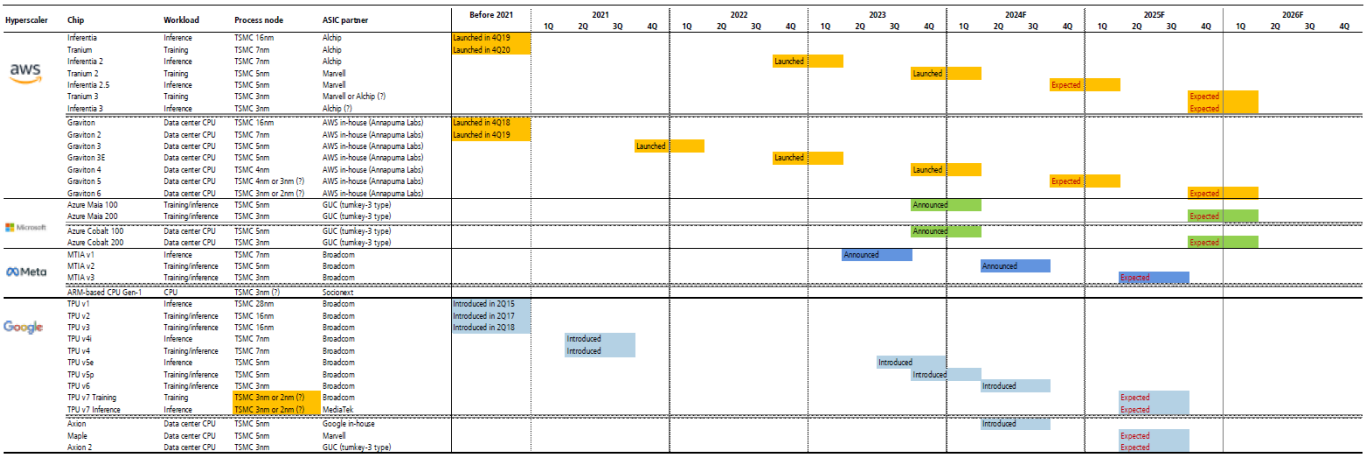
Though GPUs are capable of both training and inferencing, their high demand has led to elevated prices; for instance, a single Nvidia H100 GPU costs around US$24,000, with installations often costing millions due to the quantity of GPUs required. This high expense

has pushed US tech leaders like Amazon, Google, Intel, Microsoft, Meta, and Tesla to develop their own silicon chips, typically in the form of ASICs, which are optimized for specific AI tasks.

Unlike versatile GPUs, which handle a range of computational needs, ASICs are custom-built for specific tasks like AI inferencing, making them more efficient and powerful for such operations. ASICs achieve higher performance and power efficiency by eliminating unnecessary functions and streamlining their architecture for particular AI algorithms. However, ASICs lack flexibility, excelling only in the specific tasks they're designed for, unlike GPUs that support a broad range of functions. Developing ASICs also requires significant upfront resources and time due to their custom design and manufacturing needs, whereas GPUs are mass-produced for broader markets, making them more readily available and often less costly individually. Despite these upfront investments, ASICs are critical for efficient AI inferencing, especially in real-time applications where quick decision-making is essential. The evolution of these specialized chips highlights the need for efficient hardware solutions tailored to the growing demands of AI inferencing.

Therefore, into 2025F, we believe long-term ASIC TAM growth is still consensus. As clients continue to actively adopt AI ASIC, aside from general GPU solutions, we believe this indicates that sales generated from the AI business are set to grow rapidly, and even accelerate entering the AI inference era. Therefore, we believe higher exposure to CSPs' and hyperscalers' ASIC business will help sustain resilient ASIC TAM growth in 2024-25F. Overall, we believe ASIC TAM will continue to grow long-term, whether in AI and HPC or automotive applications.

Figure 38: US hyperscalers' data center ASIC roadmap



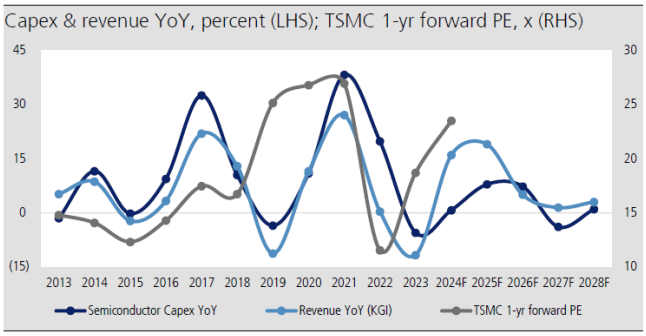
Source: KGI Research estimates

Stocks for action

Foundry

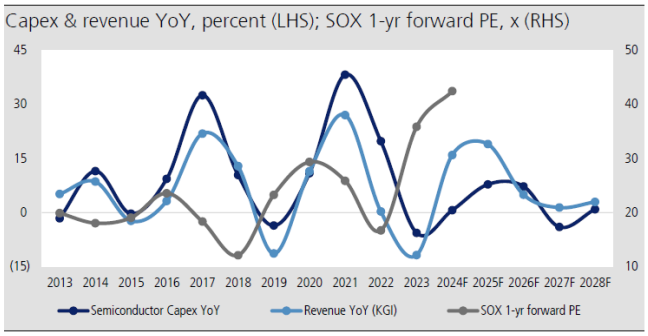
TSMC's share valuation historically leads the fundamentals cycle trough 2-4 quarters in advance but peaks out at the same year during the cycle peak. In view of structural changes in the competitive landscape and continual upward earnings revisions, we believe TSMC's valuation will reach a PE cycle peak of 25x again in 2025F (likely 1H25F).

Figure 39: Semiconductor cycle vs. TSMC's 1-yr forward PE



Source: Bloomberg; Gartner; KGI Research estimates

Figure 40: Semiconductor cycle vs. SOX 1-yr forward PE



Source: Bloomberg; Gartner; KGI Research estimates

Fabless

In the fabless sector, we like MediaTek in view of new non-smartphone products (particularly AI) driving 2026F sales growth. We expect: (1) ARM-based AI PC solutions in cooperation with Nvidia to kick in in 4Q25F; and (2) project wins in Google's inferencing TPUv7 SerDes to drive sales growth in 2026F back to 20% YoY.

Testing interface

In the test interface industry, we are positive on MPI's prospects for 2025F due to: (1) major customers continuing to show strong demand for networking IC and AI ASIC; and (2) the company securing multiple new projects. We believe that MPI's leading position in the AI ASIC supply chain will continue to be a boon for business. We are also positive on Winway in 2025F, primarily due to: (1) according to the product roadmaps of leading global GPU customers, the product cycle for AI GPU has shortened to 1.0-1.5 years, with multiple specifications of GPU under the same architecture; (2) the packaging size of AI GPU continues to increase, which will enhance the pin count and ASP of test sockets, benefiting suppliers of final test (FT) and system level test (SLT) test socket. Next-generation AI GPU products are expected to enter mass production between the end of 2025F and early 2026F, continuing to boost Winway's revenue from coaxial test socket. We believe that as a major supplier of test socket to leading global GPU manufacturers, Winway's leading position will continue to drive the company's growth.

IP

In the IP industry, we are positive on M31 due to its authorization of 2nm IP, with licensing revenue expected to contribute 20-30% of 2025F revenue. Additionally, 35-40% of the company's revenue comes from China, which is set to benefit from the active expansion of China's wafer fab capacity, driving an increase in licensing revenue. We are also positive on Andes Technology, as the penetration of RISC-V continues to rise, and the company's technology is 1-2 years ahead of Chinese peers, with 25-30% of revenue coming from the China market. Notably, both companies are expected to see a moderation in operating expense growth in 2025F, and we believe operating leverage is going to play out.

ASIC

In the ASIC design service sector, Alchip Technologies is our top pick on resilient ASIC outsourcing demand. We are positive on its leading position in the custom AI, networking and auto ASIC design realm. Although demand from US IDM customer projects is unlikely to support 2025F revenue growth, we reiterate a sales hypergrowth outlook in 2026F for

Alchip, driven by: (1) a significant ramp-up in a US hyperscaler's next-gen AI ASIC; (2) full-year contribution from the China automotive project; and (3) potential TAM expansion into networking-related applications. We believe the adverse impact of weakened demand for the US IDM's 5nm AI accelerator will be gradually reflected in Alchip's share price, while reaffirming that we expect strong revenue and earnings growth to resume in 2026F.

For Faraday, we see MP sales benefiting from strong YoY growth in advanced packaging projects, while the existing mature process MP business will recover from a low base. The NRE business will benefit from increased request for quote (RFQ) demand, expanding the design-win base, and accumulating NRE sales from multiple existing design-wins. We believe Faraday's long-term business transformation remains on track. We are still positive on Faraday's increasing RFQ demand and growing number of design-wins driving long-term sales and earnings growth.

All the above named KGI analyst(s) is SFC licensed person accredited to KGI Asia Ltd to carry on the relevant regulated activities. Each of them and/or his/her associate(s) does not have any financial interest in the respectively covered stock, issuer and/or new listing applicant.

Disclaimer

All the information contained in this report is not intended for use by persons or entities located in or residing in jurisdictions which restrict the distribution of this information by KGI Asia Limited ("KGI") or an affiliate of KGI. Such information shall not constitute investment advice, or an offer to sell, or an invitation, solicitation or recommendation to subscribe for or invest in any securities or investment products or services nor a distribution of information for any such purpose in any jurisdiction. In particular, the information herein is not for distribution and does not constitute an offer to sell or the solicitation of any offer to buy any securities in the United States of America, or to or for the benefit of United States persons (being residents of the United States of America or partnerships or corporations organised under the laws of the United States of America or any state, territory or possession thereof). All the information contained in this report is for general information and reference purpose only without taking into account of any particular investor's objectives, financial situation or needs. Such information is not intended to provide professional advice and should not be relied upon in that regard.

Some of KGI equity research and earnings estimates are available electronically on www.kgi.com.hk. Please contact your KGI representative for information. The information and opinions in this report are those of KGI internal research activity. KGI does not make any representation or warranty, express or implied, as to the fairness, accuracy, completeness or correctness of the information and opinions contained in this report. The information and opinions contained in this report are subject to change without any notice. No person accepts any liability whatsoever for any loss however arising from any use of this report or its contents. This report is not to be construed as an invitation or offer to buy or sell securities and/or to participate in any investment activity. This report is being supplied solely for informational purposes and may not be redistributed, reproduced or published (in whole or in part) by any means for any purpose without the prior written consent of KGI. Members of the KGI group and their affiliates may provide services to any companies and affiliates of such companies mentioned herein. Members of the KGI group, their affiliates and their directors, officers and employees may from time to time have a position in any securities mentioned herein.