NASDAQ:PENG Q2 2026 Earnings Call Transcript

Generated on 11/25/2026

Operator | Conference Operator:

Ladies and gentlemen, thank you for joining us and welcome to Penguin Solutions' second quarter fiscal 2026 earnings call. After today's prepared remarks, we will host a question and answer session. If you would like to ask a question, please press star one to raise your hand. To withdraw your question, press star one again. I will now hand the conference over to Suzanne Schmidt, Investor Relations. Suzanne, please go ahead.

Suzanne Schmidt | Investor Relations:

Thank you, operator. Good afternoon and thank you for joining us on today's earnings conference call and webcast to discuss Penguin Solutions' second quarter fiscal 2026 results. On the call today are Kash Shaikh, Chief Executive Officer, and Nate Olmstead, Chief Financial Officer. You can find the accompanying slide presentation and press release for this call on the investor relations section of our website. We encourage you to go to the site throughout the quarter for the most current information on the company. I would also like to remind everyone to read the note on the use of forward-looking statements that is included in the press release and the earnings call presentation. Please note that during this conference call, the company will make projections and forward-looking statements, including but not limited to statements about the market demand, technology shifts, industry trends, and the company's growth trajectory and financial outlook, business plans and strategy, including investment plans, product development and roadmap, anticipated sales, orders, revenue, and customer growth and diversification, and existing and potential strategic agreements and collaborations. Forward-looking statements are based on current beliefs and assumptions and are not guarantees of future performance and are subject to risks and uncertainties, including without limitation, the risks and uncertainties reflected in the press release and the earnings call presentation filed today, as well as in the company's most recent annual and quarterly reports. The forward-looking statements are representative only as of the date they are made, and except as required by applicable law, we assume no responsibility to publicly update or revise any forward-looking statements. We will also discuss both gap and non gap financial measures. Non gap measures should not be considered in isolation from as a substitute for or superior to our gap results. We encourage you to consider all measures when analyzing our performance. A reconciliation of the gap to non gap measures is included in today's press release and accompanying slide presentation. And with that, let me now turn the call over to Kash Shaikh, CEO. Kash?

Kash Shaikh | Chief Executive Officer:

Good afternoon. Thank you for joining our second quarter F526 earnings call. This is my first earnings call as CEO of Penguin Solutions, and I'm excited to step into this role. I want to start by thanking Mark Adams for his leadership and for the strong foundation he built. Since joining in early February, I've spent significant time with customers, partners, and our teams around the world. I've witnessed the strength of the company, both in our technology and our customer relationships. What is clear is this, AI is moving from experimentation to production, with workloads increasingly shifting towards real-time inference. We are already seeing this translate into customer demand beyond hyperscale across enterprise, neocloud, and sovereign AI markets. We expect this transition to expand our addressable market and drive increased demand for integrated AI infrastructure, where Penguin is already winning. We see this firsthand in the breadth of our deployments, from a sovereign AI factory, Hain in South Korea, to enterprise voice AI with Deepgram, to large scale research systems with Georgia Tech, along with a growing pipeline across all three market segments. What makes this opportunity so significant is that the architecture of AI is also changing. Model training was largely

compute bound. Inference powering agentic AI is memory bound and latency sensitive. We believe this is driving a re-architecture of the data center across compute, memory, interconnect, and software. We also see AI driving memory demand, not only for the high bandwidth memory or HPM used with GPUs or other accelerators, but also for general purpose memory General purpose compute wraps around every GPU build out. And whether it's reinforcement learning pipeline or inference survey, that workload runs on processors backed by significant memory content across the entire system. So while memory markets are cyclical, we believe AI is adding a more durable layer of demand for memory. As AI factories scale, I expect customers to increasingly prioritize partners that deliver with speed and precision, along with full stack AI factory platform capabilities, including compute, scalable memory systems, cluster management software, end-to-end services, and a partner ecosystem to deliver a differentiated solution. Time to deployment is now directly tied to time to force token. Against this backdrop, we are building Penguin into an AI factory platform company. Our AI factory platform is built around six core elements. First, Penguin Clusterware, our AI infrastructure management software. Second, our new Penguin Memory AI line of systems designed specifically for AI inference workloads. Third, Penguin Advanced Computing Systems optimized for AI workloads. Fourth, Penguin Origin AI factory architectures, our reference designs for AI factories. And fifth and sixth, end-to-end services and our partner ecosystem. Production-grade AI factories require full-stack design across compute, memory, storage, networking, and software. We partner with leading AI companies, including Nvidia and SK Telecom, and partners like Dell. We also offer complete end-to-end services spanning design, build, deploy, and manage services. We are strategically positioned at the intersection of AI infrastructure and memory, with a long track record in both. Few, if any, companies combine these capabilities at scale. We believe that together our AI infrastructure and memory expertise position us to meet the evolving requirements of AI infrastructure as it shifts towards inference workloads. This supports our ability to develop differentiated solution. Given the momentum we are seeing in our AI infrastructure business, and the significant market opportunity ahead of us. We are very focused in this area. We plan to invest more in our AI factory platform to accelerate our AI business growth, specifically in product innovation, go-to-market, and customer engagement. In March at NVIDIA GTC conference, we announced two AI inference-centric solutions aligned with this strategy. First, the Penguin Memory AI Server. Building upon our Compute Express Link, or CXL-based memory expansion capabilities, we introduced a new line of scalable memory systems called Memory AI. CXL is a high-speed interconnect that enables scalable shared memory across GPUs and CPUs. We also announced the immediate availability of our new memory AI KV cache server. Here KV or key value cache stores inference context to accelerate large language model responses. Second, the expansion of our origin AI factory architecture portfolio, which now includes blueprints that address the larger workloads and the low latency demands of AI inference. We also continue to expand capabilities of clusterware toward a unified control plane for AI factory infrastructure. integrating the open ecosystem to deliver repeatable production scale deployments. To accelerate the innovation and strengthen our leadership team, we recently appointed Ian Coley as Senior Vice President and Chief Product Officer. Ian brings more than two decades of experience building AI infrastructure platforms and scaling high performance computing most recently at Amazon Web Services. He was recently named by HPC Wire to its People to Watch 2026 list, reflecting his reputation in the industry. Now, let me briefly address our second quarter performance. In Q2, we delivered net sales of $343 million. Non-GAAP gross margin was 31.2%. Non-GAAP diluted earnings per share were $0.52. These results reflect strong demand and execution in memory and continued progress in our AI HPC business. Before turning to the segments, I would like to address our updated outlook As Nate will describe in further detail, following our solid Q2 net sales and EPS performance, we are raising the midpoint of our full-year net sales and EPS outlook. We are raising our outlook for our integrated memory business fueled by AI-driven demand, strong execution by our team, and favorable pricing dynamics. While our second half advanced computing net sales outlook is lower than our prior expectations, we are encouraged by strong year-over-year Q2 bookings growth for non-hyperscalar AIHPC business, which included five new AIHPC customer wins, that brings our first half total this year to seven new AI HPC logos compared to three in the first half of last year. With that context, let me turn a closer look at each of the segments. Starting with advanced computing, net sales for the quarter were $116 million. representing 34% of total company net sales and declined year over year. Advanced computing net sales for the second quarter reflect both the timing of large deployments and our transition away from hyperscaler concentration. They also reflect the

previously disclosed wind down of our Penguin Edge business. We believe diversification of net sale and wind down of Penguin Edge will strengthen the long-term quality of the business. As I mentioned, we are transitioning our AI infrastructure business from hyperscaler concentration toward a more diversified customer base across enterprise, neocloud, and sovereign AI. This transition is showing very encouraging progress but we still have more work to do. Non-hyperscale AIHPC net sales grew 50% year over year for the first half of the year, representing over 40% of first half segment net sales. Supported by strong non-hyperscale year over year booking growth in the quarter, including five new AI HPC logos across financial services, biomedical research, and energy. We expect further diversification in the second half of the fiscal year. Our AI HPC pipeline continues to strengthen with opportunities to acquire additional logos in the second half of the fiscal year across enterprise, neocloud, sovereign AI customers. As previously discussed, these engagements typically progress over many months from prospecting to design to award, followed by contracting and ultimately system build and deployment. While this sales cycle can be long, often 12 to 18 months and can introduce quarterly net sales variability. It also supports deeper customer relationships, repeat business, and a more durable long-term growth. I'm encouraged by the trajectory of the business and the signals we are seeing in the market. Beyond the numbers, we are also seeing increased activity in specific enterprise verticals. For example, we recently announced our collaboration with Deepgram and Dell to support enterprise voice AI deployments. This win highlights the growing demand for low latency production scale inference infrastructure in real time applications. In this engagement, Penguin designed and deployed an optimized inference environment built on Dell PowerEdge servers and NVIDIA RTX Pro 6000 Blackwell GPUs. This solution facilitates deep-gram speech-to-text, text-to-speech, and voice agent functionalities for applications within healthcare and retail sectors. This case study also demonstrate how design and integration expertise delivers differentiated value. As inference workload scale, we expect these types of deployments to become an increasingly important driver of AI infrastructure demand. Georgia Tech's AI Makerspace developed in partnership with NVIDIA is a strong example. Our relationship with Georgia Tech continues to grow and validates Penguin's ability to help organizations move efficiently from concept to production grade AI infrastructure. Now turning to integrated memory, net sales for the quarter were 172 million, representing 50% of total company net sales and grew 63% year over year. AI-driven demand remains strong across networking, telecommunications, and computing market segments. Pricing dynamics were favorable, and although supply remained tight, We continue to manage constraints effectively through our supplier relationships and disciplined procurement. Stepping back, our AI HPC and memory segments taken together enable us to integrate compute and memory architecture in ways that meet the requirements of production AI environments. Memory architecture is becoming increasingly central to AI performance, particularly as inference workloads scale. Our early investments in CXL position us well as customers evaluate more dynamic memory architectures. Furthermore, we are beginning to see this demand translate into customer deployments including a recent substantial order for CXL cards from a generative AI company building solutions for inference workloads. This reinforces our strategic position at the intersection of memory and AI infrastructure to capitalize on the next phase of AI. focused on inference powering agentic AI workloads. These solutions are sold to enterprise AI infrastructure buyers, the same customers we serve in our AI HPC business. For example, we sold our CXL powered KV cash servers to a tier one financial institution for their on-premise AI factory. In parallel, we continue to advance development of our photonic memory appliance, or PMA, formerly referred to as OMA, which is designed to extend memory capacity and bandwidth for large-scale AI environments. We were an early investor in a photonic memory company, Celestial AI, reflecting our longstanding focus in memory architecture innovation and our early conviction in the importance of optical interconnects for next-generation AI systems. Celestial AI was decently acquired by Marvell in a multi-billion dollar deal. Beyond the portion of proceeds we received from the acquisition as an investor, we are positioning ourselves for future growth in this market. As inference workloads expand, technologies like PMA can help address key memory scaling challenges in the next generation AI systems. Last but not least, LED. Net sales for the quarter were 56 million, representing 16% of total company net sales, and were down 7% year over year. The business continues to operate with focused leadership and dedicated operational discipline. While market conditions remain mixed, we are maintaining a disciplined approach to investment and capital allocation. We are focused on optimizing portfolio value while concentrating resources on areas where we see the strongest long-term returns. In close, the demand for

data center AI infrastructure and memory is expanding rapidly. AI factories are becoming infrastructure that powers artificial intelligence across a range of industries. As AI shifts toward inference and agentic systems and scales across large enterprise, new cloud and sovereign AI environments, we expect demand to accelerate. At the same time, memory is becoming a defining constraint. and a defining opportunity. Penguin sits at the intersection of AI infrastructure and memory innovation. And we believe that is a powerful position to be in. Our focus is clear. We are prioritizing four areas. First, to invest in product innovation across our AI factory platform. particularly at the intersection of AI infrastructure and memory to drive profitable growth. Second, to execute with speed and precision. Third, to deepen customer engagement and our ecosystem to support long-term growth. And fourth, to continue diversifying our customer base while building toward more consistent and predictable growth. We believe this focus positions us well to execute in a rapidly evolving market while continuing to build a durable and scalable business. With that, I'll turn it over to Nate.

Nate Olmstead | Chief Financial Officer:

Thanks, Cash. I will focus my remarks on our non-GAAP results, which are reconciled to GAAP in our earnings release tables and in the investor materials available on our website. With that, let me now turn to our second quarter results. In the quarter, Total Penguin Solutions net sales were $343 million, down 6% year over year. Non-GAAP gross margin came in at 31.2%, which was up 0.4 percentage points versus Q2 last year. Non-GAAP operating margin was 13.2%, down 0.2 percentage points versus last year, and non-GAAP diluted earnings per share were 52 cents, flat year over year. In the second quarter of fiscal 2026, our overall services net sales totaled $64 million, up 1% versus the prior year. Product net sales were $279 million in the quarter, down 8% versus the prior year. net sales by business segment were as follows. In advanced computing, Q2 net sales were $116 million, which was 34% of total company net sales and down 42% year over year. This sales decline reflects both the ongoing wind down of our Penguin Edge business and hyperscale hardware sales in Q2 last year, which did not recur in Q2 this year. Drilling down deeper into our advanced computing results, our non-hyperscale AI HPC net sales were down 35% year over year in the quarter, but up 50% for the first half of the year. Given the project nature of the business, where sales can be lumpy from one quarter to the next, we believe looking at the multi-quarter trend is a helpful way to evaluate the growth in this portion of our business. In addition to solid first half growth in our non-hyperscale AI HPC business, we continue to make good progress on diversifying our net sales to new customer segments. For the first half of the year, the non-hyperscale AI HPC business represented more than 40% of total advanced computing net sales versus approximately 20% in the first half of last year. We expect to see our mix of net sales from enterprises, neoclouds, and sovereign AI customers increase further in the second half of this fiscal year. In integrated memory, Q2 net sales were $172 million, which was 50% of total company net sales and up strongly with 63% growth year over year. And in optimized LED, Q2 net sales were $56 million, which was 16% of total company net sales and down 7% versus the same quarter last year. Non-GAAP gross margin for Penguin Solutions in the second quarter was 31.2%, up 0.4 percentage points year over year and up 1.2 percentage points sequentially, with strong margin performance in each business driven primarily by product mix and advanced computing favorable pricing in memory, and tariff recovery in LED. We currently project lower gross margins in the second half, driven by a higher mix of lower margin AI hardware and memory sales, rising memory costs in our AI factory solutions, and less tariff cost recovery in LED. Non-GAAP operating expenses for the second quarter were $62 million, down 3% year over year, and relatively flat sequentially. We expect a modest sequential increase in operating expenses in the second half, reflecting normal seasonality and increased investments in R&D, including for our clusterware software and memory AI solutions. Q2 non-GAAP operating income was $45 million, down 8% year-over-year and up 9% versus last quarter. Operating margins were down 0.2 percentage points versus the prior year, but up 1.1 points sequentially, driven by higher sequential gross margins in both memory and advanced computing. Non-GAAP diluted earnings per share for the second quarter were 52 cents, flat versus Q2 last year, and up 7% versus the prior quarter. Adjusted EBITDA for the second quarter was $50 million, down 6% year-over-year and up 11% versus the prior quarter. Turning to the balance sheet, for working capital, our net accounts receivable totaled $371 million compared to $330 million a year ago, with the increase driven by

higher memory sales volumes and variations in sales linearity across the quarters. Day sales outstanding were healthy at 50 days, consistent with the prior year and down one day versus last quarter. Inventory totaled $322 million at the end of the second quarter, up from $200 million a year ago, reflecting increased memory costs, growth in our memory business, and strategic purchases to fulfill memory and AI demand in the second half of the year. Days of inventory was 51 days, up from 37 days a year ago and 38 days last quarter, primarily due to our strategic memory purchases and the timing of receipts and shipments. Accounts payable were $401 million at the end of the quarter, up from $238 million a year ago, due primarily to higher memory costs, growth in our memory business, and the timing of purchases and payments. Days payable outstanding was 63 days compared to 44 days last year and 55 days last quarter. The year-over-year and quarter-over-quarter movements were due to the timing of purchases and payments. Our cash conversion cycle was 38 days, an improvement of five days compared to Q2 last year and up three days versus last quarter due to the timing of purchases and payments. Consistent with past practice, days sales outstanding, days payables outstanding, and inventory days are calculated on a gross sales and a gross cost of goods sold basis, which were 672 million and 578 million, respectively, in the second quarter. As a reminder, the difference between gross and net sales is primarily related to our memory business's logistics services, which are accounted for on an agent basis. meaning that we only recognize the net profit on logistic services as net sales. Cash, cash equivalents, and short-term investments totaled $489 million at the end of the second quarter, down 158 million versus Q2 last year, and up 28 million sequentially. The year-over-year fluctuation was primarily due to proceeds from the issuance of preferred shares in Q2 of last year, offset by debt repayments for our term loan in Q4 of last year. Sequentially, the cash increase was due to cash generated from operating activities, as well as approximately $32 million received from proceeds from the disposition of our investment in Celestial AI in connection with its sale to Marvell Technology. These sources of cash were partially offset by our share repurchase activity in the quarter. We ended the quarter with $450 million of debt, down $20 million versus last quarter due to the retirement of our 2026 convertible notes. In total, we closed the quarter in a net cash position, and based on our current debt maturity schedule, have no further scheduled debt payments due until 2029. Second quarter cash flows provided by operating activities totaled $55 million compared to $73 million provided by operating activities in the prior year quarter. The decrease in cash flow in the quarter versus last year was due primarily to investments in networking capital to support growth for the second half of this fiscal year. For those of you tracking capital expenditures and depreciation, capital expenditures were $2 million in the second quarter and depreciation was $5 million for the quarter. Wrapping up our cash flow activities, we spent $32 million to repurchase approximately 1.7 million shares in the second quarter under our stock repurchase program. As of February 27, 2026, an aggregate of $64.5 million remained available for the repurchase of our common stock under the current authorizations. And now, turning to our outlook. Given our solid half one performance and an improved half two outlook for our memory business, we are raising our full company net sales and non-GAAP diluted EPS outlook for the year, which at the midpoint now calls for 12% net sales growth and $2.15 of non-GAAP diluted EPS, up from our previous outlook of 6% net sales growth and $2 of non-GAAP diluted EPS. As a reminder, our full year outlook assumes that we will continue to diversify our customer sales mix and does not include any advanced computing AI hardware sales to hyperscale customers. And also consistent with our assumptions from last quarter, our FY26 financial outlook reflects the ongoing wind down of our high margin Penguin Edge business. We expect sales from this business to essentially cease by the end of fiscal 2026. The combined effect of these two assumptions in our FY26 outlook remains approximately a 14 percentage point unfavorable year over year impact to our total company net sales growth and approximately a 30 percentage point unfavorable impact to advanced computing. With that said, our full year net sales outlook reflects the following full year growth ranges by segment. For advanced computing, we now expect full year net sales to change between minus 25 and minus 15% year over year. While our advanced computing net sales outlook for this fiscal year is lower than our previous forecast, we are encouraged by our AI HPC bookings, including several new logos and pipeline growth. As it has previously, this outlook reflects the Penguin Edge and hyperscale hardware sales impacts mentioned earlier. For memory, we now expect net sales to grow between 65% and 75% year over year, driven by strong demand and a favorable pricing environment. And for LED, we continue to expect net sales to decline between minus 15% and minus 5% year over year. Our non-GAAP gross margin outlook for the full year is now 28%, plus or minus 0.5 percentage points. We

adjusted our gross margin outlook down by one percentage point to account for a higher mix of memory sales, which have a lower gross margin than our company average and higher memory costs in our AI hardware business. Our full year expectation for total non gap operating expenses remains $250 million and we have narrowed that range to plus or minus $5 million. For FY26, we now expect a non-GAAP diluted share count of approximately 53 million shares, down from our prior outlook, primarily reflecting the impact of our recent share repurchases. Our non-GAAP full-year diluted earnings per share is now expected to be approximately $2.15, plus or minus 15 cents. Our forecasted FY26 non-GAAP tax rate remains at 22%. And while we expect to use this normalized non-GAAP tax rate throughout FY26 and beyond, the long-term non-GAAP tax rate may be subject to changes for a variety of reasons, including the rapidly evolving global and US tax environment, significant changes in our geographic earnings mix, or changes to our strategy or business operations. Our outlook for fiscal year 2026 is based on the current environment, which contemplates, among other things, the global macroeconomic environment and ongoing supply chain constraints, especially as they relate to our advanced computing and integrated memory businesses. This includes extended lead times for certain components that are incorporated into our overall solutions, impacting how quickly we can ramp existing and new customer projects and fulfill customer orders. Our outlook also contemplates the industry-wide higher costs for memory, which may slow customer demand for our products and solutions and may lower our gross margins in our advanced computing and memory businesses. Overall, we believe our focused execution, disciplined expense management, and balance sheet strength provide a strong foundation for sustained profitable growth. We expect these qualities to support our continued progress as we pursue opportunities to enhance long-term shareholder value. Please refer to the non-GAAP financial information section and the reconciliation of GAAP to non-GAAP measures tables in our earnings release and the investor materials on our website for further details. With that, operator, we are ready for Q&A.

Operator | Conference Operator:

We will now begin the question and answer session. Please limit yourself to one question and one follow up. If you would like to ask a question, please press star one to raise your hand. To withdraw your question, press star one again. We ask that you pick up your headset when asking a question to allow for optimum sound quality. If you are muted locally, please remember to unmute your device. Please stand by while we compile the Q&A roster.

Catherine Murphy | Analyst, Goldman Sachs:

your first question comes from the line of catherine murphy from goldman sachs your line is open please go ahead ask about the raised memory segment outlook for 65 to 75 growth for how much of this is from increased favorable pricing versus demand for new product categories and as a follow-up How should we think about the impacts to the operating margin outlook for this segment and the investments that need to be made into new technologies like CXL and photonic memory appliances? Thank you very much.

Nate Olmstead | Chief Financial Officer:

Hey, Kat. It's Nate. So on the memory outlook, listen, we're really pleased with the demand that we're seeing as well as the favorability that we see in the pricing environment. I would say for the increase that we're seeing in the second half, that's majority pricing, but demand is also very strong across markets. Telco, networking, AI-driven demand is just very strong. In fact, you know, to get to the high end of that outlook really just refers to our ability to secure materials, which is really the only inhibitor we see right now to raising that outlook here in the second half. So we're chasing materials. We're using the balance sheet to strategically purchase ahead where we can. But the demand is very strong in memory. In terms of the investments, we've reflected it in the outlook, so I've kept the OPEX for the year at $250 million, plus or minus

$5 million. We're balancing the portfolio, as we always do, to look for opportunities to accelerate our investments in innovation, in AI, or in the memory solutions that we've been talking about. But that's all included in the outlook. I expect the operating margins for memory to remain pretty healthy in the back half of the year. I do expect some pressure on gross margins in AI as we see a higher mix of new hardware shipments in the second half, as well as factoring in some of the higher memory input costs that we have in that business.

Catherine Murphy | Analyst, Goldman Sachs:

Thank you very much, Nate.

Nate Olmstead | Chief Financial Officer:

You bet.

Operator | Conference Operator:

sorry your next question comes from the line of brian chin we're experiencing some mild technical difficulties my apologies your next question comes from the line of brian chin from stifle your line is open please go ahead hi hi uh great thank you for letting us ask a few questions and good afternoon um maybe first question i guess in advanced computing what change that caused you to lower the midpoint

Operator | Conference Operator:

of your prior guidance to the new range you've communicated? And can you describe how booked you are to that midpoint of that new range?

Kash Shaikh | Chief Executive Officer:

So one of the main factor is the lag between our bookings and the revenue. Our revenue lags about three to six months from from the time of the bookings. And this is primarily driven by the timing of the deployments, in some cases, the material availability and so on and so forth. And given where we are in terms of our fiscal year, we have five months remaining. So going forward, most of the bookings that we are expecting may not materialize into the revenue for the second half of this fiscal, but we believe that it will have A positive impact, obviously, going into the first half of the next fiscal. So that's one of the reasons that we are lowering the guidance for advanced computing driven by the deployments. But we are seeing strong momentum in our pipeline as well as bookings. Bookings grew very significantly in Q2 for non-hyperscale AI SPC business. which is very strategic for us, and we are encouraged to see the progress. We closed five new logos with AISPC in Q2, and in first half, that takes a total to seven new logos as compared to three new logos last year. So we are very confident in our ability to execute. The main issue at this point is timing.

Operator | Conference Operator:

Okay. Yeah, I appreciate that, Cash. And it sounds like you're pretty well booked into the fiscal second half, lowered outlook, and that some of these new bookings are more kind of beyond a six-month window. Also thinking about growth in the business, obviously there's that sort of headwind that you helped to clarify in terms of reduction in hardware revenue to the new hyperscaler, the wind down of Penguin Edge. And so 30 percentage points, impact if we kind of net that against the guidance and maybe 10 growth for this year um

none of that in that segment so moving forward you know as you survey the business and you haven't been in the role that long and you think about what that sort of apples to apples growth rate was or is tracking to for this fiscal year how are you thinking about sort of target growth rates for the advanced computing business moving forward

Kash Shaikh | Chief Executive Officer:

So overall, let me give you a data point. So the first half of this fiscal, our net sales grew about 50% year over year for non-hyperscale AI SPC business, representing 40% of the overall mix of advanced computing, which is almost 2x of what we closed last fiscal. So the growth is substantial in terms of the bookings as well as the revenue that we see. And we expect that to continue. And as we continue to close the bookings, converting the pipeline, we see strong pipeline across all three segments that I mentioned between enterprises, on-prem AI deployments, significant activity with sovereign AI customers, as well as new cloud customers.

Anata Baro | Analyst, Loop Capital:

Thank you. Thanks, Brian.

Operator | Conference Operator:

Your next question comes from the line of Matthew Calatree from Needham and Company. Please go ahead.

Matt Calatree | Analyst, Needham & Company:

Hey, guys. Matt Calatree here from Needham. Thanks for taking our question. Do the new memory launches mark a shift in strategy on that front? Just curious, because in the past, the company has talked kind of more about the the niche parts of the integrated memory business and noted it's early on things like the CXL front, but now it sounds like memory is expected to be a larger driver as part of this AI factory platform. So just wondering if anything has changed there and what gives you confidence there's durable demand here?

Kash Shaikh | Chief Executive Officer:

Yeah, so it is a part of our strategy. The memory AI appliances that we launched about a month ago is starting with GPC, is a part of us investing more in our AI factory platform strategy. So there are six elements to this strategy, and memory AI is one of the strategic elements where it is very timely if you look at how AI is transitioning from model training to inference. And in the workloads where you are focused on inference, memory becomes an increased requirement because of lower latency as well as larger context size for inference, powering the agentic AI. So this is very strategic for our business. In fact, we are leading the market in this area, taking advantage of our unique position at the intersection of memory and AI infrastructure. And combining the deep understanding and architecture, we introduced this memory AI KVCAS server as one of the products in the line of memory AI. We are working on other products and we will continue to invest and in fact, invest more in this area to take advantage of the market opportunity because the timing is perfect and our leadership in the memory AI line of products. To give you a proof point, one of the new logos we acquired, Tier 1 Financial Institution, not only are we deploying the AI infrastructure, AI factory deployment for them, they also purchased our CXL-based KVCash server, which is a proof point of as customers are transitioning from training and bringing AI on-premise in their factories, deploying on-premise, focusing on inference and powering agentic AI. It is very strategic for us, and the timing is just right. So we expect to see this demand, and we plan to continue to invest in this area.

Matt Calatree | Analyst, Needham & Company:

Awesome. That's great to hear. And then, Nate, with a new CEO in the seat and some moving pieces around sales cycles and supply change, did you change the guidance philosophy at all or embed any additional conservatism? Any color on the puts and takes there would be helpful.

Nate Olmstead | Chief Financial Officer:

Yeah. Hey, Matt. No, no change in the philosophy. You know, we... Cash and I very quickly aligned, I think, on how we think about tracking the business and looking at things. And, in fact, I think with our new CRO, who came in a couple quarters ago, he's done a nice job of adding some more rigor to the planning process in our AI business and just improving the visibility there a little bit. But it's a challenging environment from a supply chain standpoint. And we're, of course, got a lot of experience managing supply chain in our memory business. And I think that, you know, that's an advantage for us in an environment like this.

Matt Calatree | Analyst, Needham & Company:

Awesome. Thank you guys so much.

Nate Olmstead | Chief Financial Officer:

Thank you.

Operator | Conference Operator:

Your next question comes from the line of Sameek Chatterjee from JP Morgan. Please go ahead.

Sameek Chatterjee | Analyst, J.P. Morgan:

Hi. Thank you for taking my question. This is MP on behalf of Sameek Chatterjee. So my first question is I just want to double click on your advanced computing guidance. You mentioned that a lag of three to six months for the revenue which you will book in your

Nate Olmstead | Chief Financial Officer:

second half but was there a change observed for the for the bookings which you did in first quarter or any change in relative relative to what were you expecting to do in 2q and i have a follow-up as well yeah hey mp um you know i think bookings were strong in q2 really good growth sequentially and year over year i do think that the deployment cycle has lengthened a little bit with some of the supply constraints in particular on memory things have gotten a little bit longer But we're really pleased with the five new logos, and I think demand is good. We're seeing good strength in the pipeline, and it's also diversifying nicely across the non-hyperscale segments such as enterprise and neocloud and sovereign. So I think we feel really good about the demand. I think this is just an issue of a little bit of timing as we can convert bookings into revenue.

Sameek Chatterjee | Analyst, J.P. Morgan:

Okay. Okay. And my second question would be also on advanced computing and your AI factory-related business. Does NVIDIA coming up with their own reference designs for factory-level solutions, how does that play relative to you? Is that a tailwind for you or is that a headwind for you? Can you please help us understand that?

Kash Shaikh | Chief Executive Officer:

Yeah, we believe this is an advantage for us. So we work very, very closely with NVIDIA. And some of the wins that I mentioned, for example, the tier one financial institution recently, along with our memory AI product in this transaction, NVIDIA worked very closely with us. And we are working with NVIDIA, leveraging their reference design, combining that with our AI factory platform and complementing NVIDIA's NVI as an example to provide full stack to our customers. So their blueprints are more complementary to our AI factory platform and the components that make up for it. So we are actually quite excited about those blueprints and working very closely with NVIDIA to capture the opportunities, especially as NVIDIA is increasingly focused on enterprise. It aligns with our strategy and go-to-market. Thank you. Thank you very much. Thanks, MP.

Operator | Conference Operator:

Your next question comes from the line of Anata Baro from Loop Capital. Please go ahead.

Anata Baro | Analyst, Loop Capital:

Thank you for taking the question. Good afternoon. A couple, if I could. And maybe Nate as well. So earlier remarks were that you're seeing increased momentum across neoclass sovereign and enterprise. And you mentioned one or two of the new wins. And I think, Cash, you had mentioned you've made some specific or at least general influencing remarks, including around the Gentic. Do you have any specific... context you can give us around what your customers are telling you their thrust in inferencing is right now, and maybe the degree to which Agentsic is showing up there. Like, if we just wanted to get a sense of what the customer's activity tone is, like, behaviorally, say, over the last 90 to 180 days. Do you have anything there you can

Kash Shaikh | Chief Executive Officer:

share with us to make it a little bit more experiential for us and then i have a quick follow-up two things sure we believe uh we are early in the adoption of inference with these customers but it is increasingly um deployed as in customers as they move towards agentic inference provides the opportunity for powering the agentic and in when when you think about uh when you think about I'll give you an example of why the architecture is changing and why memory is becoming increasingly critical in inference as compared to the model training. So, for example, let's say if you are writing a book and if you have to write a new sentence without having the memory as a supporting component for you, you will have to reread the entire book before writing the next sentence. So in the inference, you know, you're doing an inference on a lot of data you already have. And if you have a component where the book you have written so far is stored, so before writing the new sentence, you don't have to reread the book. That's kind of how it is changing for the enterprises and other segments. And we see customers already deploying it and architecture is changing, which is why not only we have the opportunity and advantage to provide them our AI infrastructure as well as the services, increasingly we are seeing the demand for our memory AI portfolio where they are deploying AI infrastructure and increasingly inference. They need products like that to be able to provide that memory component for the inference so that the responses of LLMs can be much more faster than they would be otherwise.

Anata Baro | Analyst, Loop Capital:

I got it. That's helpful. And just one quick follow-up on mine for the time here, in case there's anybody behind me. The CXL product, it sounds like you, to the earlier question, it sounds like you guys are a little bit more enthusiastic about the CXL sleeve today than you were maybe 90 days ago. You have the new product out of GTC. Is that accurate statement that you're expecting? Maybe it's because of these new products. a little bit more, and certainly some of the NVIDIA announcements to CES as well. But are you expecting a little bit more revenue a little bit sooner than maybe you were CXL-wise 90 days ago? And then a quick second part to that, do you need photonics to work before you really get CXL amplification? Like, do you need CPO or photonics to work before you can really amplify CXL in CXL?

Kash Shaikh | Chief Executive Officer:

in scale out for scale up thanks that's it for me yep so let me address your cxl question first i think cxl adoption is timely given the transition to inference because as i mentioned with inference you need increased memory for faster llm responses and what cxl provide compute express link is you can share the memory between for GPUs and CPUs. So what it allows is new memory pooling, which is an advantage in inference workloads. So while CXL was obviously available for the last, I'd say, few quarters, it is driving that inference adoption is driving the adoption for CXL and this transaction that I mentioned where we received an order It's actually an enterprise generative AI company working on inference workloads. So you can imagine CXL cards make sense for them because those workloads need increased memory and the memory pooling capabilities provided by CXL between GPUs and CPUs are an advantage for those kind of customers. And then in terms of photonic memory appliance that we are working on, in our partnership with Celestial AI, which is now obviously Marvell, that provides increased capability because obviously when you have photonic connectivity, then you have increased capacity to share the memory. So it takes it to the next level. However, CXL in itself is an advantage. We can take it to the next level with the photonic appliance. There is another element, which is KV cache that I mentioned. memory AI KV cache server, which is essentially providing much more responsiveness for larger context workloads, again, used in inference. So various requirements, you can think of it as, inference has various requirements related to memory and the type of workloads it has, and some of it is latency. So these components between CXL or the CXL-based KV cache which provides increased responses and larger memory size, largest context sizes, and then taking it to the next level, photonic memory, make up various use cases for inference. So inference gets mainstream. We will have an advantage of this fair portfolio helping with various use cases of inference.

Anata Baro | Analyst, Loop Capital:

That's great context. Kash, thanks so much. Thank you.

Operator | Conference Operator:

Your next question comes from the line of Kevin Cassidy from Rosenblatt. Your line is open. Please go ahead.

Kevin Cassidy | Analyst, Rosenblatt:

Yes, thanks for squeezing me in. And just the gross margin for the memory, you know, your gross margin was up in the quarter and memory revenue was up strong. And I just want to understand what the dynamics were there.

Nate Olmstead | Chief Financial Officer:

Yeah, sure, Kevin. You know, we saw a little favorability in memory margins. Some of that is mixed. A little bit stronger demand than Flash, actually, which is a little bit higher margin product for us within the portfolio. And then also some of the pricing increases, we were able to capture a little bit of margin upside on that just based on the timing of our inventory purchases relative to the timing of, you know, shipments and sales to customers.

Kevin Cassidy | Analyst, Rosenblatt:

Okay. So you kind of, as you look out to the second half of the year, you see that catching up, you know, the price increases compared to?

Nate Olmstead | Chief Financial Officer:

Yeah, so as the price increases slow, right, if that's an assumption that you use, that price increases are going to slow, then we would see, we would expect to see less margin favorability from that because it would be less of a timing difference between, or less of a price variation between the timing between purchasing inventory and selling. But, you know, we have been using the balance sheet to try to secure inventory where we can. It's a tight market, so it's not unlimited supply, but where we can, we're using the balance sheet to try to gain a little bit of an advantage.

Kevin Cassidy | Analyst, Rosenblatt:

Okay. And maybe just as we're talking about memory, as you get to these CXL systems, would you expect that's going to be a higher margin than the module business?

Nate Olmstead | Chief Financial Officer:

Yeah, we do. It's really a solution. It's got software aspects to it, some good differentiation on the hardware as well. So I see that as a nice margin opportunity for us down the road.

Kevin Cassidy | Analyst, Rosenblatt:

Okay, great, thanks.

Nate Olmstead | Chief Financial Officer:

Thank you.

Operator | Conference Operator:

At this time, there are no further questions. I will now hand the call over to Kash Shaikh, CEO, for closing remarks.

Kash Shaikh | Chief Executive Officer:

Thank you, operator. We see AI shifting toward inference with demand expanding beyond hyperscaler to enterprise, neocloud, and sovereign AI customers. We are still in early shift in this transition, but the combination of our customer demand, product innovation, and booking momentum gives us the confidence in the path ahead. We believe we are well positioned at the intersection of AI compute infrastructure and memory, and we are making good progress diversifying our customer base. My focus is on strong execution across product innovation, customer engagement, and diversification, disciplined capital allocation, and investment in our AI HPC business to support the long-term growth. We look forward to updating you on our progress.

Operator | Conference Operator:

This concludes today's call. Thank you for attending. You may now disconnect.