

Advanced Micro Devices (AMD) / 4 Feb 25 / 2024 Q4 Earnings call transcript

Company Profile

⋮ Transcript menu

Matt Ramsay	executive
Lisa Su	executive
Jean Hu	executive
Aaron Rakers	analyst
Timothy Arcuri	analyst
Vivek Arya	analyst
Joshua Buchalter	analyst
Harlan Sur	analyst
Blayne Curtis	analyst
Stacy Rasgon	analyst
Toshiya Hari	analyst

Print

Operator

Greetings, and welcome to the AMD Fourth Quarter and Full Year 2024 Conference Call. [Operator Instructions]. And as a reminder, this conference is being recorded. It is now my pleasure to introduce to you Matt Ramsay, Vice President of Investor Relations. Thank you, Matt.

You may begin.

Matt Ramsay

Thank you, and welcome to AMD's Fourth Quarter and Full Year 2024 Financial Results Conference Call. By now, you should have had the opportunity to review a copy of our earnings press release and accompanying slides.

If you have not had the chance to review these materials, they can be found on the Investor Relations page of amd.com.

We will refer primarily to non-GAAP financial measures during today's call. The full non-GAAP to GAAP reconciliations are available in today's press release and slides posted on our website.

Participants in today's conference call are Dr. Lisa Su, our Chair and Chief Executive Officer; and Jean Hu, our Executive Vice President and Chief Financial Officer and Treasurer. This is a live call and will be replayed via webcast on our website.

Before we begin, I would like to note that Jean Hu will attend the Morgan Stanley Global TMT Conference on Monday, March 3.

Today's discussion contains forward-looking statements based on our current beliefs, assumptions and expectations speak only as of today and as such, involve risks and uncertainties that could cause actual results to differ materially from our current expectations. Please refer to the cautionary statement in our press release for more information on factors that could cause our actual results to differ materially.

With that, I will hand the call over to Lisa. Lisa?

Lisa Su

Thank you, Matt, and good afternoon to all those listening today. 2024 was a transformative year for AMD. We successfully established our multibillion-dollar Data Center AI franchise, launched a broad set of leadership products and gained significant server and PC market share.

As a result, we delivered record annual revenue, grew net income 26% for the year and more than doubled free cash flow from 2023.

Importantly, the Data Center segment contributed roughly 50% of annual revenue as Instinct and EPYC processor adoption expanded significantly with cloud, Enterprise and supercomputing customers.

Looking at our financial results. Fourth quarter revenue increased 24% year-over-year to a record \$7.7 billion, led by record quarterly Data Center and Client segment revenue both of which grew by a significant double-digit percentage. On a full year basis, annual revenue grew 14% to \$25.8 billion as Data Center revenue nearly doubled and Client segment revenue grew 52%, more than offsetting declines in our Gaming and Embedded segments.

Turning to the segments. Data Center segment revenue increased 69% year-over-year to a record \$3.9 billion. 2024 marks another major inflection point for our server business as share gains accelerated, driven by the ramp of fifth-gen EPYC Turin and strong double-digit percentage year-over-year growth in fourth gen EPYC sales. In cloud, we exited 2024 with well over 50% share at the majority of our largest hyperscale customers. Hyperscaler demand for EPYC CPUs was very strong, driven by expanded deployments powering both their internal compute infrastructure and online services.

Public cloud demand was also very strong with a number of EPYC instances increasing 27% in 2024 to more than 1,000. AWS, Alibaba, Google, Microsoft and Tencent launched more than 100 AMD general purpose and AI instances in the fourth quarter alone. This includes new Azure instances powered by a custom-built EPYC processor with HBM memory that delivers leadership HPC performance based on offering 8x higher memory bandwidth compared to

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

I Agree

competitive offerings.

We also built significant momentum with Forbes 2000 global businesses using EPYC in the cloud as enterprise customers activated more than doubled the number of EPYC cloud instances from the prior quarter. This capped off a strong year of growth as Enterprise consumption of EPYC in the cloud nearly tripled from 2023.

Turning to Enterprise on-prem adoption. EPYC CPU sales grew by a strong double-digit percentage year-over-year as sell-through increased, and we closed high-volume deployments with Akamai, Hitachi, LG, ServiceNow, Verizon, Visa and others.

We are seeing growing enterprise pull based on the expanding number of EPYC platforms available and our increased go-to-market investments. Exiting 2024, there are more than 450 EPYC platforms available from the leading server OEMs and ODMs, including more than 120 turn platforms that went into production in the fourth quarter from Cisco, Dell, HPE, Lenovo, Super Micro and others.

Looking forward, Turin is clearly the best server processor in the world with more than 540 performance records across a broad range of industry standard benchmarks. At the same time, we are seeing sustained demand for both fourth and third gen EPYC processors as our consistent road map execution has made AMD the dependable and safe choice.

As a result, we see clear growth opportunities in 2025 across both cloud and enterprise based on our full portfolio of EPYC processors optimized for leadership performance across the entire range of Data Center workloads and system price points.

Turning to our Data Center AI business. 2024 was an outstanding year as we accelerated our AI hardware road map to deliver an annual cadence of new Instinct accelerators, expanded our ROCm software suite with significant uplifts in inferencing and training performance, built strong customer relationships with key industry leaders and delivered greater than \$5 billion of Data Center AI revenue for the year.

Looking at the fourth quarter, MI300X production deployments expanded with our largest cloud partners. Meta exclusively used MI300X to serve their Llama 405B frontier model on meta.ai and added Instinct GPUs to its OCP-compliant Grand Teton platform designed for deep learning recommendation models and large-scale AI inferencing workloads.

Microsoft is using MI300X to power multiple GPT-4 based Copilot services and launched flagship instances that scale up to thousands of GPUs for AI training and inference and HPC workloads. IBM, DigitalOcean, Vultr and several other AI-focused CSPs have begun deploying AMD Instinct accelerators for new instances. IBM also announced plans to enable MI300X on their watsonx.ai and data platform for training and deploying enterprise-ready generative AI applications. Instinct platforms are currently being deployed across more than a dozen CSPs globally, and we expect this number to grow in 2025.

For enterprise customers, more than 25 MI300 series platforms are in production with the largest OEMs and ODMs. To simplify and accelerate enterprise adoption of AMD Instinct platforms, Dell began offering MI300X as a part of their AI factory solution suite and is providing multiple ready-to-deploy containers via the Dell Enterprise Hub on Hugging Face. HPC adoption also grew in the quarter.

AMD now powers 5 of the 10 fastest and 15 of the 25 most energy-efficient systems in the world on the latest top 500 supercomputer list. Notably, the El Capitan system at Lawrence Livermore National Labs debuted as the world's fastest supercomputer, using over 44,000 MI300A APU's to deliver more than 1.7 exaflops of compute performance.

Earlier this month, the high-performance computer center at the University of Stuttgart, launched the Hunter supercomputer that also uses MI300A. Like El Capitan, Hunter will be used for both foundational scientific research and advanced AI projects, including training LLMs in 24 different European languages.

On the AI software front, we made significant progress across all layers of the ROCm stack in 2024.

Our strategy is to establish AMD ROCm as the industry's leading open software stack for AI, providing developers with greater choice and accelerating the pace of industry innovation. More than 1 million models on Hugging Face now run out of the box on AMD. And our platforms are supported in the leading frameworks like PyTorch and JAX, serving solutions like vLLM and compilers like OpenAI Triton.

We have also successfully ramped large-scale production deployments with numerous customers using ROCm, including our lead hyperscale partners. We ended the year with the release of ROCm 6.3 that included multiple performance optimizations including support for the latest flash attention algorithm that runs up to 3x faster than prior versions and SGLang runtime that enabled day 0 support for state-of-the-art models like DeepSeek-V3.

As a result of these latest enhancements, MI300X inferencing performance has increased 2.7x since launch.

Looking forward, we're continuing to accelerate our software investments to improve the out-of-the-box experience for a growing number of customers adopting Instinct to power their diverse AI workloads.

For example, in January, we began delivering biweekly container releases that provide more frequent performance and feature updates in ready to deploy packages, and we continue adding resources dedicated to the open source community that enable us to build, test and launch new software enhancements at a faster pace.

On the product front, we began volume production of MI325X in the fourth quarter. The production ramp is progressing very well to support new customer wins. MI325 is well positioned in market, delivering significant performance and TCO advantages compared to competitive offerings.

We have also made significant progress with a number of customers adopting AMD Instinct.

For example, we recently closed several large wins with MI300 and MI325 at LighthouseAI customers that are deploying Instinct at scale across both their inferencing and training workloads. We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

Looking ahead, our next-generation MI350 series featuring our CDNA 4 architecture is looking very strong. CDNA 4 will deliver the biggest generational leap in AI performance in our history, with a 35x increase in AI compute performance compared to CDNA 3. The Silicon has come up really well. We were running large-scale LLMs within 24 hours of receiving first Silicon and validation work is progressing ahead of schedule. The customer feedback on MI350 series has been strong, driving deeper and broader customer engagements with both existing and net new hyperscale customers in preparation for at-scale MI350 deployments.

Based on early Silicon progress and the strong customer interest in the MI350 series, we now plan to sample lead customers this quarter and are on track to accelerate production shipments to mid-year.

As we look forward into our multiyear AMD Instinct road map, I'm excited to share that MI400 series development is also progressing very well. The CDNA Next architecture takes another major leap enabling powerful rack scale solutions that tightly integrate networking, CPU and GPU capabilities at the Silicon level to support Instinct solutions at Data Center scale.

We designed CDNA Next to deliver leadership AI and HPC flops while expanding our memory capacity and bandwidth advantages and supporting an open ecosystem of scale up and scale out networking products.

We are seeing strong customer interest in the MI400 series for large-scale training and inference deployments and remain on track to launch in 2026.

Turning to our acquisition of ZT Systems. We passed key milestones in the quarter and received unconditional regulatory approvals in multiple jurisdictions, including Japan, Singapore and Taiwan. Cloud and OEM customer response to the acquisition has been very positive as ZT Systems expertise can accelerate time to market for future Instinct accelerator platforms.

We have also received significant interest in ZT's manufacturing business.

We expect to successfully divest ZT's industry-leading U.S.-based Data Center infrastructure production capabilities shortly after we closed the acquisition, which remains on track for the first half of the year.

Turning to our client segment. Revenue increased 58% year-over-year to a record \$2.3 billion. We gained client revenue share for the fourth straight quarter, driven by significantly higher demand for both Ryzen desktop and mobile processors. We had record desktop channel sell-out in the fourth quarter in multiple regions as Ryzen dominated the best-selling CPU list at many retailers globally, exceeding 70% share at Amazon, New AG, MineFactory and numerous others over the holiday period.

In mobile, we believe we had a record OEM PC sell-through share in the fourth quarter as Ryzen AI 300 Series notebooks ramp.

In addition to growing share with our existing PC partners, we were very excited to announce a new strategic collaboration with Dell that marks the first time they will offer a full portfolio of commercial PCs powered by Ryzen Pro processors. The initial wave of Ryzen-powered Dell commercial notebooks is planned to launch this spring with the full portfolio ramping in the second half of the year as we focus on growing commercial PC share.

At CES, we expanded our Ryzen portfolio with the launch of 22 new mobile processors that deliver leadership compute, graphics and AI capabilities.

Our Ryzen processor portfolio has never been stronger with leadership compute performance across the stack.

For AI PCs, we are the only provider that offers a complete portfolio of CPUs, enabling Windows Copilot+ experiences on premium ultrathin, commercial, gaming and mainstream notebooks.

Looking into 2025, we're planning for the PC TAM to grow by a mid-single-digit percentage year-on-year. Based on the breadth of our leadership client CPU portfolio and strong design win momentum, we believe we can grow client segment revenue well ahead of the market.

Now turning to our Gaming segment. Revenue declined 59% year-over-year to \$563 million. Semi-custom sales declined as expected as Microsoft and Sony focused on reducing channel inventory. Overall, this console generation has been very strong. Highlighted by cumulative unit shipments surpassing \$100 million in the fourth quarter.

Looking forward, we believe channel inventories have now normalized and semi-custom sales will return to more historical patterns in 2025.

In Gaming graphics, revenue declined year-over-year as we accelerated channel sellout in preparation for the launch of our next-gen Radeon 9000 series GPUs.

Our focus with this generation is to address the highest volume portion of the enthusiast gaming market with our new RDNA 4 architecture. RDNA 4 delivers significantly better rate tracing performance and add support for AI-powered upscaling technology that will bring high-quality 4K gaming to mainstream players when the first Radeon 9070 series GPUs go on sale in early March.

Now turning to our Embedded segment. Fourth quarter revenue decreased 13% year-over-year to \$923 million. The demand environment remains mixed, with the overall market recovering slower than expected as strength in aerospace and defense and test and emulation is offset by softness in the industrial and communications markets.

We continued expanding our adaptive computing portfolio in the quarter with differentiated solutions for key markets.

We launched our Versal RF series with industry-leading compute performance for aerospace and defense markets, introduced our Versal premium series Gen 2 as the industry's first adaptive compute devices supporting CXL 3.1 and PCI Gen 6 and began shipping our next-gen Alveo card with leadership performance for ultra-low latency trading. We believe we gained adaptive computing share in 2024 and are well positioned for ongoing share gains based on our design win momentum.

We closed a record \$14 billion of design wins in 2024, up more than 25% year-over-year as customer adoption of our industry-leading adaptive computing platforms expands and we won large new embedded processor designs. In summary, we ended 2024 with significant momentum, delivering record quarterly and full year revenue. EPYC and Ryzen processor share gains grew throughout the year, and we are well positioned to continue outgrowing the market based on having the strongest CPU portfolio in our history. We established our multibillion-dollar Data Center AI business and accelerated both our Instinct hardware and ROCm software road maps.

For 2025, we expect the demand environment to strengthen across all of our businesses, driving strong growth in our Data Center and Client businesses and modest increases in our Gaming and Embedded businesses. Against this backdrop, we believe we can deliver strong double-digit percentage revenue and EPS growth year-over-year.

Looking further ahead, the recent announcements of significant AI infrastructure investments like Stargate, and latest model breakthroughs from DeepSeek and the Allen Institute highlight the incredibly rapid pace of AI innovation across every layer of the stack, from Silicon to algorithms to models, systems and applications.

These are exactly the types of advances we want to see as the industry invests in increased compute capacity while pushing the envelope on software innovation to make AI more accessible and enable breakthrough generative and agentic AI experiences that can run on virtually every digital device.

All of these initiatives require massive amounts of new compute and create unprecedented growth opportunities for AMD across our businesses. AMD is the only provider with the breadth of products and software expertise needed to power AI from end to end across data center, edge and client devices.

We have made outstanding progress building the foundational product, technology and customer relationships needed to capture a meaningful portion of this market. And we believe this places AMD on a steep long-term growth trajectory led by the rapid scaling of our Data Center AI franchise for more than \$5 billion of revenue in 2024 to tens of billions of dollars of annual revenue over the coming years.

Now I'd like to turn the call over to Jean to provide some additional color on our fourth quarter and full year results. Jean?

Jean Hu

Thank you, Lisa, and good afternoon, everyone. I'll start with a review of our financial results and then provide our current outlook for the first quarter of fiscal 2025. AMD executed very well in 2024, delivering record revenue of \$25.8 billion, up 14%, driven by 94% growth in our Data Center segment and a 52% growth in our Client segment, which more than offset headwinds in our Gaming and Embedded segments. We expanded gross margin by 300 basis points and achieved earnings per share growth of 25% while investing aggressively in AI to fuel our future growth.

For the fourth quarter of 2024, Revenue was a record \$7.7 billion, growing 24% year-over-year as strong revenue growth in the Data Center and Client segment was partially offset by lower revenue in our Gaming and Embedded segments. Revenue was up 12% sequentially, primarily driven by the growth of Client, Data Center and the Gaming segments. Gross margin was 54%, up 330 basis points year-over-year due to a favorable shift in revenue mix with higher Data Center and Client revenues lower Gaming revenue, partially offset by the impact of lower embedded revenue.

Operating expenses were \$ 2.1 billion, an increase of 23% year-over-year as we invest in R&D and marketing activities to address our significant growth opportunities. Operating income was a record of \$2 billion, representing 26% operating margin. Taxes, interest and other was \$249 million net expense.

For the first quarter of 2024 diluted earnings per share was \$1.09, an increase of 42% year-over-year, reflecting the significant operating leverage of our business model.

Now turning to our reportable segment. starting with the Data Center segment. Revenue was a record of \$3.9 billion, up 69% year-over-year driven by strong growth in both AMD Instinct GPU and the fourth and the fifth generation AMD EPYC CPU sales. Data Center segment operating income was \$1.2 billion or 30% of revenue compared to \$666 million or 29% a year ago.

Client segment revenue was a record of \$2.3 billion, up 58% year-over-year, driven by strong demand for AMD Ryzen processors. Client segment operating income was \$446 million or 19% of revenue compared to operating income of \$55 million or 4% of revenue a year ago, driven primarily by operating leverage due from higher revenue. Gaming segment revenue was \$563 million, down 59% year-over-year, primarily due to a decrease in semi customer revenue. Gaming segment operating income was \$50 million or 9% of revenue compared to \$224 million or 16% a year ago. Embedded segment revenue was \$923 million, down 13% year-over-year as end market demand continues to be mixed. Embedded segment operating income was \$362 million or 39% of revenue compared to \$461 million or 44% a year ago.

Turning to the balance sheet and cash flow.

During the quarter, we generated \$1.3 billion in cash from operations and a record \$1.1 billion of free cash flow. Inventory increased sequentially by \$360 million to \$ 5.7 billion. In the end of the quarter, cash, cash equivalent and short-term investments were \$5.1 billion. In the fourth quarter, we repurchased 1.8 million shares and returned \$256 million to shareholders.

For the year, we repurchased 5.9 million shares and returned \$862 million to shareholders.

We have a \$4.7 billion remaining in our share repurchase authorization.

Before I turn to our financial outlook, let me cover our financial segment reporting. Beginning with our first quarter fiscal year 2025 financial statement disclosures. We plan to combine the client and the gaming segment into one single reportable segment to align with how we manage the business. Therefore, reporting three segments: Data Center, Client with Gaming and the Embedded. We'll continue to provide distinct revenue disclosures for our Data Center, Client, Gaming and Embedded businesses consistent with our current reporting.

Now turning to our first quarter of 2025 outlook.

We expect revenue to be approximately \$7.1 billion plus or minus \$300 million, up 30% year-over-year, driven by strong growth in our Data Center and the Client segments, modest growth in our Gaming business and a modest decline in our Embedded business.

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

We expect revenue to be down sequentially approximately 7%, driven primarily by seasonality across our businesses.

In addition, we expect first quarter non-GAAP gross margin to be approximately 54%. Non-GAAP operating expenses to be approximately \$2.1 billion. Non-GAAP other net income to be \$24 million, non-GAAP effective tax rate to be 13% and the diluted share count is expected to be approximately 1.64 billion shares.

In closing, 2024 was a strong year for AMD, demonstrating our disciplined execution to deliver revenue growth and expand earnings at a faster rate than revenue. All while investing in AI and the innovation to fuel long-term growth.

Looking ahead, we will build on the momentum to drive double-digit percentage revenue growth and further accelerate earnings in 2025 and beyond.

With that, I'll turn it back to Matt for the Q&A session.

Matt Ramsay

Thank you very much, Jean. We're now ready to start the Q&A session.

As the operator pulls for questions, we remind each participant to please ask one question and a brief follow-up. Operator, please poll for questions. Thanks.

Operator

[Operator Instructions]. And the first question comes from the line of Aaron Rakers with Wells Fargo.

Aaron Rakers

I guess I'll just ask it right out of the gate as we think about the GPU business, and I appreciate you talked about delivering north of \$5 billion of revenue, which is extremely impressive in 2024. I'm curious if you -- how we should think about framing the GPU, the Instinct business as we think about 2025? And any kind of color you can provide us as far as kind of the progression of revenue, the pace of revenue first half versus second half as we think about some of the product cycle dynamics.

Lisa Su

Sure, Aaron. Thanks for the question.

So first of all, look, we were very pleased with how we finished 2024 in terms of the Data Center GPU business.

I think the ramp was steep as we went throughout the year, and the team executed well.

Going into 2025, as I mentioned in the prepared remarks, we're actually very happy with the progress that we're making on both the hardware road maps and the software road maps.

So on the hardware side, we launched 325 at the end of the fourth quarter, started shipments then.

We have new designs that have come on both 300 and 325 that we'll deploy in the first half of the year. And then the big news is on the 350 series.

So we had previously stated that we thought we would launch that in the second half of the year. And frankly, that bring-up has come up better than we expected, and there's very strong customer demand for that.

So we are actually going to pull that production ramp into the middle of the year, which improves our relative competitiveness.

So as it relates to how Data Center -- so the overall Data Center business will grow strong double digits certainly, both the server product line as well as the Data Center GPU product line will grow strong double digits. And from the shape of the revenue you would expect that the second half would be stronger than the first half, just given 350 will be a catalyst for the Data Center GPU business. But overall, I think we're very pleased with the trajectory of the Data Center business in both 2024 and then going into full year 2025.

Aaron Rakers

Yes. And as a quick follow-up, just thinking about the guidance overall relative to that down 7% sequential, I know you mentioned seasonality across the business segments. Are you assuming that you're down sequentially in Data Center in total in 1Q? And how do I frame that relative to seasonality? Thank you.

Lisa Su

Yes, sure, Aaron.

So let me give you some more color on the Q1 guide.

So Q1 guide was down 7% sequentially, as Jean mentioned. And the way that breaks out in each of the segments assume that Data Center would be down just about that average, so the corporate average. We would expect the Client business and the embedded business to be down more than that.

Just given where seasonality is for those businesses. And then we would expect Gaming business will be down a little less than that. And that's a little atypical from a seasonality standpoint, but we're coming off of a year when there was a lot of let's call it, inventory normalization. And now that inventory has normalized, we would expect that would be down a little bit less than the corporate average.

Operator

The next question comes from the line of Timothy Arcuri with UBS.

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

I wanted to ask about the server CPU business. Jean, I think you have said in the past that core count is going to grow mid- to high teens. And as long as your competitor is not super aggressive on pricing that your business should grow roughly that much as well. Are you expecting or are you already seeing them become a little more aggressive on pricing as they attempt to shore up their share. It sounds like they're getting a bit more aggressive on pricing.

So wondering if you still think that the server CPU business can grow in line with that core kind of mid- to high teens.

Jean Hu

Yes.

First, we always assume server CPU is a very competitive market but we currently have the best lineup of portfolio from not only Turin generation, but Genoa and then even Milan, we provide the best TCO for our customers based on the product portfolio.

So overall, we are actually quite confident about continuing to drive the server CPU businesses not only growing from a unit perspective, ASP perspective and continue to gain share.

Timothy Arcuri

And then, Jean, can you just give us a sense of where Data Center GPU came in for December? I'm thinking it's probably in the \$2 billion range. And then is it assumed to be down, flat or up, would you be willing to give a number for March?

Jean Hu

Yes.

I think the way to look at our Q4 performance is our Data Center business overall did really well. Actually is consistent with our expectations.

Of course, when we look at the server and the Data Center GPU, server did better than Data Center GPU. But overall, it's very consistent with our performance.

Lisa Su

Yes. Maybe I'll just add, Tim, on your question as to what you would expect as we go into 2025.

I think you should assume that the first half of 2025 Data Center segment will be consistent with the second half of '24. And that's true for both businesses on the server side as well as the Data Center GPU side.

Operator

The next question comes from the line of Vivek Arya with Bank of America Securities.

Vivek Arya

Lisa, a few questions on the Data Center GPU business.

I think last year, AMD was very explicit about setting and beating or meeting expectations. This year, you have not set a specific forecast, and I'm curious what has changed. And then if I go back to your Analyst Day in December, I think at that time, you had set a long-term 60% CAGR. Is it fair to assume that you can grow at that for '25, right, versus the \$5 billion plus that you did last year.

So just contrast the 2 years and then whether AMD can grow at that 60% trend line.

Lisa Su

Sure.

So Vivek, thanks for the question.

I think what we look at is certainly for the first year of the Data Center GPU business, we wanted to give some clear progression as it was going. The business is now at scale, actually now at over \$5 billion. And as we go into 2025, I think our guidance will be more at the segment level with some color as to -- some qualitative color as to what's going on between the two businesses. And relative to your question about long-term growth rates, you're absolutely right. I mean I believe that the demand for AI compute is strong. And we've talked about a Data Center accelerator TAM upwards of \$500 billion by the time we get out to 2028.

I think all of the recent data points would suggest that there is a strong demand out there.

Without guiding for a specific number in 2025, one of the comments that we made is we see this business growing to tens of billions as we go through the next couple of years. And that gives you a view of the confidence that we have in the business and particularly our road map is getting stronger with each generation, right? So MI300 was a great start. 350 series is stronger and addresses a broader set of workloads including both inference as well as training. And then as we get into MI400 series, we see significant traction and excitement around what we can do there with rack-scale designs, and just the innovation that's going on there.

So yes, we're bullish on the long term, and we'll certainly give you progress as we go through each quarter in 2025.

Vivek Arya

Thank you, Lisa. And for my follow-up, I would love your perspective on the news from DeepSeek recently, right? And there are kind of two parts to that. One is once you heard the news, do you think that should make us more confident or more conservative about the semiconductor opportunity going forward? Like is there something so disruptive in what they have done, that reduces the overall market opportunity. And then within that, have your views about GPU versus ASICs how that shifts developers over the next few years? Have they been involved in any way at all?

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

Lisa Su

Yes. Great. Thanks for the question, Vivek. Yes, it's been a pretty exciting first few weeks of the year.

I think the DeepSeek announcements, Allen Institute as well as some of the Stargate announcements, talk about just how much the rate and pace of innovation that's happening in the AI world.

So specifically relative to DeepSeek. Look, we think that innovation on the models and the algorithms is good for AI adoption. The fact that there are new ways to bring about training and inference capabilities with less infrastructure actually is a good thing because it allows us to continue to deploy AI compute in broader application space and more adoption.

I think from our standpoint, we also like very much the fact that we're big believers in open source. And from that standpoint, having open source models, looking at the rate and pace of adoption there, I think, is pretty amazing. And that is how we expect things to go.

So to the overall question, of how should we feel about it. I mean, we feel bullish about the overall cycle. And similarly, on some of the infrastructure investments that were announced with OpenAI and Stargate and building out, let's call it, massive infrastructure for next-generation AI.

I think all of those say that AI is certainly on the very steep part the curve. And as a result, we should expect a lot more innovation.

And then on the ASIC point, let me address that because I think that is also a place where there's a good amount of discussion. I've always been a believer in you need the right compute for the right workload.

And so with AI, given the diversity of workloads, large models, medium models, small models, training, inference when you're talking about broad foundational models or very specific models, you're going to need all types of compute. And that includes CPUs, GPUs, ASICs and FPGAs. Relative to our \$500 billion plus TAM going out in time, we've always had ASIC as a piece of that. But my belief is, given how much change there is still going on in AI algorithms that ASICs will still be the smaller part of that TAM because it is a more, call it, specific workload optimized, whereas GPUs will enable significant programmability and adjustments to all of these algorithm changes. But when I look at the AMD portfolio, it really is across all of those pieces.

So CPUs, GPUs. And we're also involved in a number of ASIC conversations as well as customers want to really have an overall compute partner. Operator

And the next question comes from the line of Joshua Buchalter with TD Cowen.

Joshua Buchalter

Obviously, it was good to see MI355X pulled into midyear. But I wanted to clarify, you said first half '25 Data Center GPU likely consistent with second half '24, and I was wondering if you could speak to whether or not the shape of the first half changed over the last few months and is potentially related to this pulled in time line, it could be a potential air pocket ahead of that launch? Or this was sort of consistent with how you saw things playing out as MI350 and MI325X ramps more fully?

Lisa Su

Yes. Thanks for the question, Joshua. No, I would say from our standpoint, we've gotten incrementally more positive on the 2025 Data Center GPU ramp.

I think 350 series was in second half always, but pulling it into midyear is an incremental positive. And from a first half, second half statement, as I mentioned, we have some new important AI design wins that are going to be deployed with 300 and 325 in the first half of the year. But with 350 series, we end up with more content. I mean it's a more powerful GPU, ASPs go up, and you would expect larger deployments that include training and inference in that time frame.

So the shape is similar to what we would have expected before.

Joshua Buchalter

And believe it or not, going to ask a question on Client. Obviously, the growth number in the fourth quarter, I mean, it was certainly higher than our model. Could you clarify the drivers of the strength across desktop, notebook and enterprise? And how we should think about 1Q and in particular, to put it bluntly, I mean are you worried at all about inventory buildup, given how much your client revenue has outperformed the broader PC market in the second half of the year?

Lisa Su

Yes. Thanks for the question.

Our Client business performed really well throughout 2024 and Q4 was a very strong quarter. There are a couple of reasons for that.

So we should go through that. We don't believe there's some substantial inventory buildup. We actually think that what we're seeing is very strong adoption of our new products.

So on the desktop side, we saw our highest sell-out in many years.

As we went through the holiday season, launching our new gaming CPUs, frankly, they have been constrained in the market, and we've continued shipping very strongly through the month of January as we're catching up with some demand there.

So desktop business was very strong. And on the notebook side, we also saw a number of our OEM partners launching new AI PCs with the slew of new mobile part numbers that we announced at CES.

We have our strongest PC portfolio on the mobile side with top to bottom Copilot+ PC compatible products, and those are playing very well into the market.

I know that there was some commentary about whether there were pull-ins relative to tariffs. I don't -- we didn't see that in the fourth quarter.

I think, as I said, we saw strong sellout.

Going into the first quarter, we do expect seasonality in there. But the part of our business that is performing better than seasonality is the desktop portion of the business. and the mobile portion of the business is, let's call it, more typical seasonality. But overall, I think we're very bullish on our prospects to grow clients in 2025.

Just given all of the drivers from product portfolio to some of the market dynamics as well as our new commercial PCs portfolio.
Operator

The next question comes from the line of Harlan Sur with JPMorgan.

Harlan Sur

For the fourth quarter, did your overall server CPU business grow double digits sequentially. And as a follow-on to that, I think, Q4 was the sixth consecutive quarter of double-digit year-over-year increases for your on-prem service solutions. On a sequential basis, I know you guys did start to see recovery in Enterprise in the second quarter of last year, I think, it was strongly up sequentially in the third quarter, a pretty broad based. Did Enterprise servers grow sequentially in Q4? And Lisa, how do you see the share prospects in this segment as you step into 2025?

Lisa Su

Yes, Harlan, thanks for the question.

So I think as Jean mentioned earlier, so in the fourth quarter, we did see a sequential double-digit growth in our server business. We saw that in both cloud and Enterprise.

I think the server business has been performing extremely well. We're continuing to grow our cloud footprint with more workloads as we have the strength of the Turin portfolio in addition to Genoa and Milan.

And then to your question on Enterprise, I do believe we're seeing some strong traction in the Enterprise.

I think what's helping us there is, frankly, we've invested a lot more in go-to-market and the go-to-market investments are paying off. The Enterprise sales cycle is often a 6- to 9-month sales cycle. But as we've invested more resources in it throughout 2024, we've seen that convert into a significant number of new POCs that are now converting into volume deployments. And as we go through into 2025, from a competitiveness standpoint, we have a very strong portfolio across every price point, every core count, every workload.

So I think we see a strong 2025 for server GPUs.

Harlan Sur

I appreciate that. Networking is a very critical part of the AI infrastructure becoming even more important. There seems to be this misconception that AMD is behind the curve here, yet you you're keeping pace kind of leveraging the incumbent Ethernet technology, strong installed ecosystem.

You guys are spearheading the Autonet Ethernet Consortium.

You've got your Infinity Fabric technology for scale-up connectivity.

So as you continue to drive customer adoption of your overall AI platforms. What's the feedback been like on your AI networking architectures? And any networking-related innovations the team is going to be bringing to the market this year?

Lisa Su

Yes. Thanks, Harlan, for that question. No question. Networking is an extremely important part of the AI solution, and it's an area that we have been investing and spending quite a bit of effort with our customers and our partners jointly. The way to think about it is that our networking proof points are actually increasing as we're going from MI300 to MI325 to MI350 to MI400.

So in each of those points, we're increasing with a number of proof points.

I think people want to see more clusters of ours. Certainly, on inference, we've shown great performance and total cost of ownership.

We are now -- also have a number of training systems that we are putting on board. And the important part there is the networking.

We have worked very closely with our partners on Ethernet. We believe that this is the right technology for the future.

In addition to third-party networking solutions, we're also with our Pensando team developing our own in-house AI NIC that Forrest mentioned at our Q4 Advancing AI event. And as we look forward working with our customers, we are actually standing up full rack solutions at both the 350 level as well as in the MI400 series.

So I think the net of it is we believe that, yes, it is absolutely very important. And in addition to all of the hardware and software work, the system level scaling is super important and we are on track to deliver that with our road map.

Operator

And the next question comes from the line of Blayne Curtis with Jefferies.

Blayne Curtis

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner I just want to follow up the data synergy business, obviously, very strongly year-over-year, but it seems for your commentary kind of the sequential growth indicates agreement. See our [Cookie Policy](#) for more information.
kind of slows for the next kind of 3 quarters.

So I just want to understand the why. Obviously, you have some new products coming.

So maybe it's just the shift to the new products. I also want to just tinker brain on in terms of when you look at the ASIC story lines, there tends to be kind of a shift to focus on training versus inference.

So just your perspective, I know a lot of your workloads initially were inference. Are you seeing any shift in terms of the demand from our customers between training inference as well?

Lisa Su

Yes. Sure, Blayne. Look, the way I would say it is we saw a tremendous growth as we build up the Data Center GPU business throughout 2024.

So I think what we're seeing is we're continuing to do new deployments. We're continuing to bring on new customers. Clearly, we are going through a little bit of a product transition time frame in the first half of the year. But the key is bringing in the MI350 series was very, very important for us and for the customer set.

So the fact that, that hardware has come on clean. And we've learned a lot from the initial deployments of MI300, I think, is very positive. And this is as we might expect, given the overall landscape of deployments.

And then to the second part of your question, as it relates to ASICs, I really haven't seen a big shift at all in the conversation. I will say that the conversation as it relates to AMD is kind of the following. People like the work that we've done in inference. But certainly, our customers want to see us as a strong training solution. And that's consistent with what we've said, right? We've said that we have like a step-wise road map to really show each one of those solutions.

On the software side, we've invested significantly more in some of those sort of the training libraries. We talked -- Harlan's question earlier about networking. And then this is about just getting into data centers and ramping up tens of thousands of GPUs.

So from my standpoint, I think we're making very good progress there. And I just want to reiterate on the ASIC side, look, I think ASICs are a part of the solution, but they are -- I want to remind everyone, they are also a very strong part of the AMD sort of toolbox.

So we've done semi-custom solutions for a long time.

We are very involved in a number of ASIC discussions with our customers as well. And what they like to do is they'd like to take our baseline IP and really innovate on top of that. And that's what I think differentiates our capability is that we do have all of the building blocks of CPUs, GPUs as well as all of the networking technologies that you would need to put the solutions together.

Matt Ramsay

Operator, I think we have time for two more callers, please.

Operator

Okay. The next question comes from the line of Stacy Rasgon with Bernstein Research.

Stacy Rasgon

I want to ask this a little more explicitly.

So you said your server business was up strong double digits sequentially in Q4. My math suggests that could have even meant that the GPU business was down sequentially. And given your guidance for, I guess, flattish GPUs in the first half of '25 versus second half of '24. Again, does the math not suggest that you'd be down sequentially both in Q1 and in Q2 to feel like -- am I doing something wrong with my math? Or like what am I missing here?

Lisa Su

Yes. perhaps, Stacy, maybe let me help give you a little bit of color there. I don't think we said strong double digits.

I think we said double digits.

So that perhaps is the -- so Data Center segment was up 9% sequentially. Server was a bit more than that. Data Center GPU was a little less than that.

I think for some of the models that are out there, you might be a little bit light in the Q3 Data Center GPU number.

So there might be some adjustments that need to be done there. But I think your suggestion would be incorrect. We -- if you just take the halves, second half '24 to first half '25, let's call it, roughly flattish, plus or minus. I mean we'll have to see exactly how it goes. But it is going to be a little bit dependent on just when deployments happen. But that's kind of currently what we see.

Stacy Rasgon

Got it. And I guess for my follow-up, maybe to follow on there, do you think your exit rate on GPUs in '25 is higher than your exit rate in '24. Are you willing to commit to that?

Lisa Su

Absolutely. Yes, of course. It would be hard to grow strong double digits otherwise, right?

Operator

And the final question comes from the line of Toshiya Hari with Goldman Sachs.

Lisa, I had a question on the server CPU business. I'm curious how you're thinking about the market this year? And if you can delineate between cloud and Enterprise, that would be really helpful. And then kind of part B to that question. In your prepared remarks, you talked about you all having more than 50% share across the major hyperscalers. How would you characterize the competitive intensity at those customers vis-a-vis some of the internal custom Silicon that's expected to ramp over the coming quarters and years?

Lisa Su

Sure, Toshiya.

So let me say, as we look into 2025, I think we see a good server market between cloud and Enterprise. The -- I think as we went into sort of the early part of '24, there was a little bit of, let's call it, less investments on the CPU side as people were optimizing investments for AI. We saw that sort of pick up in the second half of the year in '24, and we would expect that to go into '25.

So the Enterprise refresh cycles are coming in again. And certainly, there are a number of cloud vendors that are now, let's call it, reupdating some of their data centers.

And then your second question was, as it relates to...

Toshiya Hari

Competitive custom Silicon.

Lisa Su

Yes. Look, I think it's about the same. What I would say, Toshiya, is it's less about custom Silicon versus x86. It's much more about do you have the right product for the right workload. And look, the server market is always a competitive market. What we've done, and you've seen it in our Zen 4 product line as well as in our Zen 5 product line. We've expanded the design points for each of the core generation.

So that we have cloud native and then we have enterprise optimized low core count, high core count, highest performance, best perf per dollar.

And I think as we do those things, I think, we are continuing to grow share across both cloud and Enterprise. And look, it's always very competitive. We take every design win with -- very, very seriously, but we're winning our fair share. And I think that's the strength of the product portfolio. And also, I think there's a good amount of trust for our delivery capability as we've built up our -- sort of our franchise over the last number of years.

Toshiya Hari

That's great. And then as a quick follow-up, maybe one for Jean.

So you're guiding gross margin to 54% in the first quarter. I'm curious what some of the major puts and takes are and the things that we should be cognizant of going into Q2 and more importantly, the second half.

Given your Data Center commentary skewed more to the second half, I would expect margins to improve in the second half. But yes, if you can kind of run through the pluses and minuses, that would be really helpful.

Jean Hu

Yes. Thanks for the question.

You're right.

Our gross margin is primarily driven by our revenue mix.

I think when you look at -- looking to 2025, Q1 guide, not only Data Center continued to grow significantly year-over-year. At the same time, client business is also growing year-over-year.

So overall, the revenue mix is quite consistent with the Q4.

So the gross margin guide is 54%.

I think for the first half, if the revenue mix is at this level, we do feel the gross margin will be consistent with 54%. But going into second half, we do believe the Data Center is our fastest growth driver for the company and that will drive the gross margin to step up in second half.

Matt Ramsay

All right. With that, I think we are ready to close the call now, operator. I just wanted to say thank you to everybody that listened in and participated today and for your interest in AMD. Thank you very much.

Operator

Thank you. And ladies and gentlemen, that does conclude today's teleconference. We thank you for your participation.

You may disconnect your lines at this time.