ATIF MALIK: Welcome to day one of Citi Global TMT Conference. My name is Atif Malik. I cover US semiconductors and semiconductor equipment stocks. It's my pleasure to welcome Jean Hu, EVP, CFO, and Treasurer, Matt Ramsay, CVP Financial Strategy, friendly neighborhood IR from AMD.

I'll kick it off with my questions first. If you have a question, save it. Towards the end, we'll have the mic come to you and you can ask your question. Welcome guys. Jean, if you can take us through the lay of the land. It's been an exciting year. What has changed from January till now? You guys had your big AI day as well. You raised your TAM forecast. Just walk us through how the demand picture has evolved from January onwards.

JEAN HU: Yeah, first, thank you for having us, and it's great to be here. Thank you, everyone for joining us. Yeah, it has been a really exciting year. They have so many different changes in our industry and it's really about AI, and we do believe this is the most consequential technology transformation.

When you think about the pace, the scale and the rise of the AI, it's unprecedented. It's when you look at this year, not only the model capabilities continue to advancing very quickly. We also see inferencing outpace training from AI compute demand perspective. And agentic AI, that's another thing which really have another step change in demand, not only just for GPUs, but for CPUs.

So during our Advancing AI day, we did update our market opportunities from in the past, a small number, to now $2 trillion in 2030. We do see the demand for our product, not only just GPUs, but CPUs, and the AI PC and in the future, physical AI for our adaptive compute across all different areas.

And the most exciting thing is we just reported Q2 results, and we talk about our data center business are expected to double next year, our GPU business continue to ramp. At MI 450, we're going to launch this quarter at the very beginning of the ramp, then we're going to see very significant ramp in Q4 and into next year.

And on the CPU side, it was a supply constraint, but we have been increasing supply, so we do expect the second half, the CPU business is going to expand more than 80% year-over-year and into next year continue to expand the more than 70%, which should continue to be supply constrained, so it's a very exciting time for AMD.

When you think about it, the company has been very aggressively investing in high performance compute for more than a decade. As you know, we have been building methodically the compute platform from CPU, GPU to adaptive compute across all different end markets.

We think the most different thing this is AI super investment cycle is at the very beginning. And over time, we're going to continue to see strong demand for AMD's product, and the portfolio we have built will benefit from this AI super investment cycle.

ATIF MALIK: Great. Matt, Jean talked about inference being a big driver of demand, and we've been hearing about this disaggregated compute shift in the market, the Hot Chips Conference. And now you guys struck a partnership with Cerebras. More recently, you've acquired Taalas.

Can you talk us through what your strategy is on the inference side, and how you thinking about these different pieces of the different type of inference units?

MATT RAMSAY: Sure, and thank you all, and Atif, thank you and the folks at Citi for hosting us and for everyone to come and see us. And then maybe this ties a little bit into some of the stuff that Jean talked about in your prior question about, what's changed in the market in the last 12 months? It seems like a lot.

We were waiting for inference to become the majority driver of AI computing, and that's happened. And at the same time that that's happened, we've seen this radical change from what I kind of call chatbot inference to agentic inference, and that's really created big opportunities for both inferencing silicon and for the CPUs to run the agents.

We've talked about in a number of forums, obviously, the breadth of the inference market will be the majority of it driven by GPU-led computing, and our MI455 product is going to be ramping now with Helios, and we have a roadmap to continue to innovate there. There's obviously different ASICs and XPUs in the market that are going to do some of the inferencing work.

And there's this new market for disaggregated inference and ultra fast response time tokens. That's a relatively small piece of the market today, but depending on economics could grow into a larger piece over time, so we've taken a multi-pronged approach to the problem.

One is a partnership that we have announced with Cerebras, where our Helios systems will be in their cloud and compute alongside their wafer scale engine racks to expand the utility not just of ultra fast inference, but take it to a broader range more general purpose inference in their cloud for their customers.

And then longer term, we haven't given a ton of details yet, but we do have some of our own internal silicon ambitions for ultra low latency inference that would fit into our architecture via chiplets. And the Taalas team brings a lot of really good talent. We know the folks well. Some of them are former AMD ATI folks back in the day and are going to integrate really quickly into the team and add a lot of technology and horsepower behind the internal silicon work we're doing in that area.

ATIF MALIK: Awesome. Jean, let's talk about Helios. I believe you guys already had very high expectations on Helios exiting last year, but just in terms of your shipments and where you stand on Helios, how should we think about the ramp of Helios in Q4 and into Q1 next year? How is that ramp going? I do hear from clients around questions around execution. If you can just help us understand.

JEAN HU: I would say the Helios ramp, MI450 ramp is going very well. Q3 will be at the very beginning of our production shipment. We expect revenue in Q3, but Q4 will see a very significant step up, and then another step up in Q1 2027 and ramp through 2027.

It is a scale level, like rack-scale level. It's very different, very complex, so we have been very methodical and deliberate how we design the ramp process. Start ramp in the end of Q3 and step up in Q4, Q1 into 2027. We are working with all our ODM partners, and our overall supply chain to ensure we have all the components, so it's not about the GPU or CPU, HBM memory. There are also a lot of other small components that we need to ensure we have.

And with the ODM partners, it has been a long really, really durable work with them to make sure the manufacturing process can execute. That's why it's a methodical process. It's our team have been working with ODM partners. We also have been working with all the customers. We're going to have a production shipment to ensure not only all the component works, all the mechanical, all the software stack, everything works.

So that has been ongoing. It's weekly execution process. We actually feel pretty good about the ramp, both on the supply side. The demand and the volume for 2027, I certainly go above our original initial expectation. We need to continue to expand the supply.

We feel good about the financial plan. We have talked about it, but it's absolutely the case. We can use more supplies to make sure we meet customers demand.

ATIF MALIK: Just the customer profile, you guys had Meta and Microsoft as two customers last year. You're growing that demand further this year, and then you added Anthropic as a customer as well. Can you just talk about that customer funnel, particularly touch on the neo cloud opportunity, if that's an emerging area for you, and also touch on the dollar per gigawatt across that range of customers.

JEAN HU: Yeah, so we're really pleased with our strategic long-term partnership with Anthropic we announced. When you have Anthropic, basically, we have three major anchor customers, the meta and the OpenAI and Anthropic, and all three of them, the deployment are going to be multiple gigawatt scale deployment and multi-generational engagement with all of them.

When you think about our customer pipeline, not only all three of them give us the forecast more than what we expected in the beginning of the strategic partnership. Of course, we need to make sure we have supplies to support them, but more importantly, MI350, we have seen a tremendous demand from all the other third party customers, the model builders, the new AI companies, even the enterprise customers.

So we do think the neo cloud continue to be the area we're going to work with to make sure we meet all the other customers need. The pipeline is quite significant. For us, we really need to prioritize. We want to make sure we support the three major anchor customers for large scale deployment, but at the same time, you should expect us to work with neo cloud to make sure we support all the other customers in the market.

MATT RAMSAY: Atif, one thing I would add to what Jean mentioned is, when we think about these programs, we don't just think about it as what's going to happen with Helios and MI455 over the next 12, 15, 18 months? It's the engineering level and technical engagement that's influencing what the MI500 program will look like, the MI600 program will look like, what the design of the future racks will look like.

The amount of technical engagement across these the three leading model companies that we've announced as customers so far that it's really heartening to see that influence the roadmap over a multi-generational period, and think that that gives us confidence into not just where our roadmap is going, but what that engagement level is going to look like through the end of the decade, over multiple generations.

And we saw the same thing happen. It's a very different time and a different market, but at the same thing happened on the CPU side six or seven years ago, where people were announcing partnerships with AMD on the Rome generation, but it was really influenced by what they saw and their level of influence in the roadmap of multiple generations going forward from there, and I think that's where we are now with the AI business.

ATIF MALIK: Great. Jean, the MI series is below corporate average on gross margin side, and how should we think about the ramp and the move to MI355 to MI500 in the future and the impact to your gross margin expectations next year?

JEAN HU: Thank you for the question. I think, data center AI MI450 and the future generation, it's one of the most significant growth drivers in our data center business. What it's going to help us is to drive very significant incremental revenue and gross profit, both, in next year and beyond.

And even though the gross margin percentage right now is still a bit below corporate average, but the way to think about it is we actually have a broad portfolio at the company level, so we always talk about the gross margin is being driven by the mix of our different product portfolio.

And one of the things is going into Q4 and 2027, even though we're ramping MI450 significantly, but we do have some tailwinds on the gross margin side to help us. I think the first one is the server CPU business. When you think about server CPU business right now, within our data center, it's still a larger portion of our business, and it's actually going to continue to grow.

We expect second half to be more than 80% year-over-year increase, and next year more than 70%. So from that perspective, that business is gross margin accretive to corporate average, so we do see that is going to help to offset some of the dilution from data center AI business.

Secondly, our embedded business, after three years of inventory digestion, we have seen a significant double digit year-over-year increase, not only in Q2. We also guided in Q3 and going forward, so the business recovery is very broad based in embedded business. And we are also winning a lot of design wins with our embedded x86 business in data center, in networking, and all those businesses are margin accretive to us.

Third thing is probably small, but our gaming business, which tends to be the lower gross margin, which is at a later stage of product cycle, and the memory cost is pretty high right now, which also impacts the demand side, so the mix from a gross margin perspective, it actually we do have all those tailwinds.

In general, the way to think about it is when we ramp MI450 in Q4 and 2027, the gross margin will be slightly lower than what we guided the Q3. Q3, we actually guided our gross margin at 56%. You have seen we are expanding gross margin since last year, when we actually are ramping MI350, so going forward, gross margin in 2027, quarter-over-quarter will be different.

But the most important thing is the way to think about our business model is we are driving very significant data center revenue expansion and gross margin dollars expansion, which our investment and OpEx increase is slower than revenue and gross profit increase, which is going to driving very significant operating leverage and earnings per share expansion.

ATIF MALIK: Very clear. Let's talk about the server CPUs. When I visited you guys in, I believe, it was January or something for a bus tour, some bulb turned on in terms of the CPU demand this year, and you guys were talking about not finding enough CPUs internally to do agentic AI yourself.

So walk us through, and now the TAM is like 220 billion by 2030. And can you just talk about your aspirations around market share in this server CPU market? What have you seen so far, X86 versus ARM, and just your aspirations around market share?

JEAN HU: I'll start and Matt can add. It's actually, astonishing when you think about the server CPU market expansion. When we had our last financial analyst day in November 2025, we talk about server CPU market TAM to be at 60 billion in 2030. At that time, agentic AI was at the very early beginning and the really agentic AI adoption.

The diffusion into the business enterprise, the curve has been tremendous. So we see the agentic AI adoption very significantly starting January and just continue to be almost like a vertical in enterprise market, which that's when the demand for CPU continue to increase because as Matt mentioned earlier, when you think about agentic AI, it's about workflow execution.

In enterprise, that really requires retrieving data, execute, orchestration. All those are being done on the CPUs, so not only you need to go back to your foundational CPUs to run all the tasks. You actually have to increase the layer. Agentic AI sandbox, we think it's a new segment, which is to executing all the agents tasks and make sure it coordinate with every other compute.

That market has been just continued to increase. We can see the demand continue to go up. That's why we have just updated our TAM opportunities to more than 220 billion from 25 billion in 2025. That is how significant the expansion has been, is we see more than 50% CAGR in the next several years, largely driven by agentic AI.

They are different segments. The way to think about it is you do have the foundational CPU segment which handles all your enterprise applications, your SAP, your database. That continue to grow, but the growth is probably not as big as the other segment. Then there's the head node part coordinated with GPUs. That continue to grow.

But the largest is agentic AI sandbox, which is very small today, but it's going to be more than 50% of that 220 billion market. And we do think it's not only about ASP increase. The unit will increase very significantly too, so the market opportunity is tremendous, and we are very well positioned as the company from an investment perspective.

MATT RAMSAY: Just to add a couple of points, Jean. We feel like the roadmap is in about the best place that it's ever been in the server business. If you asked the question around instruction set, so x86 versus ARM, just to be clear, I mean there are areas particularly in the enterprise server applications, where x86 legacy is very important.

There are broad applicability, not really of instruction set, but of experience in security features, reliability, serviceability, really enterprise and cloud great features that we have in the sixth or seventh generation of the roadmap that are quite important. But we also feel like our own differentiation and our competitive lead in the server business will expand with Venice pretty significantly.

It's a very, very compelling program, and the amount of demand and visibility that we have in the server business is very significant from what it's been historically. And we get asked about competition a lot, and our approach to this is not an x86 thing or an ARM thing. It's a build the best server parts, period.

I mean, that's the priority of the business, and if we do that, we can give differentiation in the agentic world of threads per rack, or threads per megawatt. We can give over 5 gigahertz products that go into head nodes. We can have broad applicability across the enterprise stack, whether that's on prem or in the cloud.

You mentioned some of the growth rates, but one of the fascinating stats to me, having been in the server industry since 2000 is, I mean, our enterprise server business grew more than 70% in the second quarter. I mean, it used to be heroic if you had double digit growth in enterprise server, and now we're talking about 70% growth, so it's a really broad based portfolio.

And the going from the Venice generation into Florence and Ravenna and beyond that, to the much larger TAM that Jean described, we do intend to grow to 50% of that dollar TAM, inclusive of all, instruction sets of competition. And when you do the math and you talk about building $100 billion server business, and that's what we're intending to do.

ATIF MALIK: Great. Jean, let's talk about supply. This topic never dies. You guys have talked about 70 billion in data center sales next year and maybe low 40 billion GPU, and the remaining CPU, we saw purchase commitments, 29, 30 billion. And I mean, we can all track and hear about TSMC's allocation. You guys are seeing the biggest jump next year, but just help us understand what are the limitations around supply and your ability to upside next year?

JEAN HU: Overall, supply is very tight. There are multiple areas, not only wafers, advanced process, node, HBM tight packaging, substrate. Some of the components are very tight. We do have an excellent supply chain team. Operationally, we have been working with the whole supply chain to ensure we can support our top line revenue growth.

On the data center AI side, MI450, we have been preparing for the ramp for a long time, so the way to think about it is we have been working with the whole ecosystem to ensure not only we have wafers, HBM, and also advanced packaging capacity, as well as all the different components to support the Helios rack level solutions. That

Has been ongoing, and now the demand continues to go up, so we absolutely needed to get more supplies. On the server CPU side, as Matt and I talked about, it's actually, the demand start to accelerate this year, so we did need to catch up. During the process, over the first half, we have continued to increase the supply from wafers perspective, from advanced packaging capacity perspective, and that we continue to invest aggressively.

We talk about our CapEx increase. The primary increase of our CapEx is to build the capacity to support the server CPU ramp, especially Venice. As Matt talked about, we do see very significant ramp next year and that actually, our capacity is not enough. We need to build ourselves to buy the equipment to do the consignment to support the ramp.

So right now, we actually feel really good about all the supply chain work we have done to ensure we can support the data center revenue to double next year, and also to ensure the embedded businesses get the supplies to stil see very significant growth in 2027. PC and gaming, we prioritize, but we absolutely want to make sure we continue to gain share in the PC market too.

So we feel pretty good, but I would say, we can have more supplies. Demand continue to increase.

ATIF MALIK: Awesome. Let me pause here and see if there are any questions in the audience. If you have a question, please raise your hand. All right, let's move on. Yep.

UNIDENTIFIED AUDIENCE MEMBER: The CPU side. Can you talk about maybe price? Intel's been taking more price because it had older products, and just wanted to get your thoughts on price and how you guys see price evolving over the next couple of years.

JEAN HU: First is when we talk about our significant CPU growth. It increases both because of unit increase and the ASP increase. Unit increase probably is more than ASP increase. When you think about ASP increase, for us, the core count have been going up for each generation, so in general, that will drive the ASP increase.

And your question into the future, when you really think about it, is the way AMD we think about how we work with our customers is we absolutely need to make sure we get the gross margin to continue to invest in the future, but in general, if there's a component cost to increase, we do need to make sure we share that cost increase with the customers.

So for instance, if there's a wafer price increase, we absolutely want to make sure we share that with the customers. And if wafer price increase you should expect us to increase the CPU price too, but in general, that is how we are operating. We really want to make sure we provide customers the best TCO and have long-term strategic relationship with them.

And our gross margin really is important to us, but we're not going to increase the price just to expand gross margin.

UNIDENTIFIED AUDIENCE MEMBER: Thank you very much for your time. As a scale up domain gets bigger, at what point does copper stop being good enough, and when does optics have to move inside the package?

MATT RAMSAY: Thank you for the question. We have not disclosed a ton about the roadmap in this area, but we have disclosed some. And at our advancing AI conference a couple of months ago, we did talk about in the MI500 series program, which would be second half of 2027 introduction and be the primary product for the company in 2028, that we will have scale up domains that are larger than we're offering now.

So we haven't said a number, but greater than 72, and that we would have both copper-based and near packaged optics-based options for scale up connectivity. We're still going to run our Infinity Fabric traffic, which is the coherency protocol that we use to communicate between CPUs and GPUs.

And that technology was donated into the industry consortium, the UAL, and so we're still going to be doing traffic tunneled over both ethernet and other means on copper and over optics. I don't think you'll see it be a point in time thing, where everybody just decides as an industry, OK, no more copper. Now we're going to go optics. It's not going to be a student body left type of decision. You're going to see the technologies run in parallel.

And different versions of optics will be introduced with different risk tolerances over time, so I guess, that's a long way of saying we're going to start on that path in the products in 2027, and we'll give you guys more updates as we go forward, but you should think about copper and optics living side-by-side for a number of generations. It's not going to be a binary shift. We have partners.

UNIDENTIFIED AUDIENCE MEMBER: --and especially, what we're seeing in the private markets.

JEAN HU: On the MA environment, we have been doing bolt on acquisitions, which really focus on the software capabilities we want to add. Of course, we acquired ZT system a while ago, which was to prepare the system level expertise to get the talent capabilities there.

You should expect us to continue to focus on those areas to make sure we not only increase the capabilities of just building the system solutions, but also software side, the stack, the model to make sure we continue to invest for the AI.

UNIDENTIFIED AUDIENCE MEMBER: Is there any scenario under which you would consider using Intel as a foundry partner?

JEAN HU: Thank you for the question.

[LAUGHTER]

And so I think as most of you know, we have a long strategic partnership with TSMC. We have been working with TSMC for a very long-term, not only just as a supplier, but on the R&D side. If you look at a lot of our technology, we actually co-developed with TSMC.

And for us, when you look at our scale and the volume, the most important thing is we want to make sure the quality of the product, the advanced process technology, the 3D packaging. Everything we're doing, it can be supported by our suppliers. Definitely, they have a fab in Arizona, where diversifying from geolocation perspective with them, but you should expect us to continue to view TSMC as our primary supplier on the wafer side.

MATT RAMSAY: You should view us as anything that's advanced packaging or anything that's advanced wafers, we're going to evaluate and due diligence on technology from every vendor in the market. But as Jean said, or for the foreseeable future, I mean, the partnership we have with TSMC is going to be a significant one, and will continue to be for a long time.

UNIDENTIFIED AUDIENCE MEMBER: Think about your competitive positioning versus NVIDIA and proprietary ASICs. Where do you feel AMD has the clearest performance costs or availability advantage at this time? Thank you.

MATT RAMSAY: What you're seeing right now in market, and Lisa and Vamsi and others at the company have talked about this very publicly, is we feel like we have for in today's generation of product tokens per dollar advantages for large scale inference that we're going to be bringing to market and deploying in large volume with OpenAI, with Meta and with Anthropic.

And over time, and then through multiple generations, we're expanding our training capability with the MI450 series. We'll expand it further with generations beyond that, but as you think about things in market today and what we're going to be ramping significantly over the next number of quarters, we feel like our customers are great, great partners with us, but they do expect us to generate differentiated economic returns for them in terms of tokens per dollar. And that's what the market is demanding of us, and that's what we think the product delivers.

But when we talk about the entire breadth of the AI market, whether that be some level of customization, whether it be merchant computing, whether it be the server CPUs that run the agents, we intend to participate in all of that. And it might not look the same at every single customer, but we can bring at significant scale technologies across DPU, CPU, custom, scale up optics, and networking and system level design across the board. But where we differentiate the most today is in large scale inference.

ATIF MALIK: Great. We're almost out of time. Jean and Matt, thank you for coming to the Citi Conference.

JEAN HU: Thank you.

MATT RAMSAY: Thank you guys.