VIVEK ARYA: OK. I guess it's still good morning. Good morning. Welcome back to this BofA Global Tech Conference. I'm Vivek Arya from BofA's Semiconductor Semicap Equipment Research Team. And I'm really delighted to have the team from Advanced Micro Devices join us this morning. Jean Hu, chief financial officer, and Matt Ramsay, head of investor relations. And I'll go through my questions, but please feel free to raise your hand if you would like me to bring up something. But really warm welcome to you, Jean and Matt. Really appreciate you joining us at the conference.

JEAN HU: Yeah. Thank you.

VIVEK ARYA: And I guess, exciting times in AI and the semiconductor industry. So maybe Jean, as a start, just walk us through your state of the union. The growth has been very strong since the start of the year. What has surprised you? Which are the areas that you think you are seeing the best growth right now, and which are the areas which have been kind of pushed back because of this?

JEAN HU: Yeah. First, thank you for having us. Thank you all for joining us. I think the biggest change during the last few months is really the rise and inflection of agentic AI. You can see continued momentum from training to inference, from AI adoption, experimentation, to more adoption at scale. So when you look at that, this agentic AI, it's not about answering questions anymore. It's about orchestration. It's about database access and a lot of tool execution.

And all of those require significant CPU performance. What we're seeing is very significant and incremental demand for our CPU platforms. That has been really exciting. At the same time, we also are seeing the economics of AI keep changing. With the token generation going up so quickly, all the customers are really focusing on performance TCO, and they're really trying to figure out how to use different compute to address different application and workload. So we see demand for GPU going up, CPU, and also a lot of other ASIC LPUs.

I think from AMD's perspective, you know our Q1 performance. We had a record CPU performance. CPU business grow more than 50%. We guided the Q2. CPU performance is going to go up year over year, 70%. That has been really exciting. So all those things really benefit us because we have been investing in GPU, CPU, adaptive compute. Also address literally end to end applications from data center, to client, to gaming. That feels like right now all the engines are really driving the business revenue growth very significantly.

VIVEK ARYA: Got it. So I think the growth in CPU has been the biggest positive drivers this year. And we are kind of tracking this 2030 or late end of the decade type of addressable opportunity, which I think you were the first to point out, would be 60 billion. That doesn't sound that high enough. So they took it up to 100. Then you took it up to 120. And Jensen said, well, could be even more than that, up to 200.

What do you think is the difference between all these forecasts? This is one of the most frequently asked questions right from the investment community. What is the right number? How does one get to it? Is there like a simple unit, and are we just all using different ASPs to get to a different number and everyone has the same unit growth?

JEAN HU: You are asking a lot of questions here. But I think maybe let's take a step back. You are absolutely right. AMD was the first one. Literally last year, we started to talk about how AI drive the demand of CPUs. So last November, when we had our financial analyst day, we actually outlined how we think about this opportunity under the TAM.

We basically said, OK, we're thinking three segments. One is the traditional general compute, which we are all very familiar with, has been driving the traditional enterprise application across the board. The second category is the head node, which is really the communication with GPUs. You want the speed to make sure you have the communication.

We also outlined agentic AI last November, which is still very early, but our team already see the early signs of agentic AI. That has been the fastest growing market in our view. That's when we said, OK, we believe the TAM is going to grow 18% of CAGR to 60 billion in 2030.

Remember, the general purpose CPU used to be just a single digit growth. So that market expansion, we saw it earlier but we never estimate is going to be so fast. The pace and the speed of agentic AI adoption has been tremendous. So when you look at the January, February, March, all the enterprise adoption started. You see Anthropic OpenAI. Their revenue has been going up very significantly. The demand for CPU continue to go up across all our customers.

That's when we had our earnings call. We actually updated our market forecast to more than 120 billion. Frankly, it's still very early, when you think about agentic AI adoption. What we are seeing is with agentic AI, it's quite complex when you have millions of operations you have to execute, when you have a lot of agents concurrently working on different tasks. You actually need very high core count CPUs.

We don't know what other companies, how they think about the TAM and the SAM. For us, we go through the bottom up and top down approach. We know at this particular point, we feel pretty good about more than 120 billion. But it's evolving so fast and more and more complex. You should expect for the agentic AI portion of the market because you have so many diverse workload, so complex. The ASP will continue to increase because increasingly, you need a very high core count CPUs, high performance CPUs. Matt, I don't know if you have other you can add here.

MATTHEW RAMSAY: So first off, thank you, Vivek, for hosting us and for you guys all coming. I think there's a couple things, Vivek, that I would maybe add to Jean's comments. One is I think for this whole audience and us included, we were trying to figure out when was the primary driver of AI CapEx going to go from almost predominantly large model training and start the shift towards inference being a primary driver of CapEx?

And we're seeing that happen in real time. And I think the more powerful thing for our business and maybe the more transformative thing architecturally is during that transition, you're seeing chatbot inference become agentic inference. And what agentic inference does as a workload, you daisy chain or concatenate all these automated agentic flows together.

And between each inference task, there's a lot of CPU diverse work. So post-processing of data, figuring out what to tell the AI to do next, often based on the result of the prior inference. Where do I get the data from for the next step? Is it in the cloud? Is it in an ERP system, a payment system, a CRM system, wherever? Get all of that data, come back, post-process the data from that step and feed it into the next agentic step for the AI.

And that workload is quite diverse. And that's what we're seeing driving the demand for Turin now in real time and as the order book fills in for the 256-core, 2-nanometer Venice parts that are going to launch in a couple of months and be the primary workhorse for next year. That's where we're seeing the order book really expand.

VIVEK ARYA: Got it. Is there a way to segment that market, whatever it happens to be, right? $120 $20 billion plus in those three segments, Jean, that you described the traditional applications. Let's call it part of the AI cluster and then a generic kind of standalone CPUs?

JEAN HU: Yeah. I think a traditional segment, it's very clear. I think 2025, different third party would say that's a $25 billion to $30 billion market opportunity. It will continue to grow, as Matt is mentioned, is you actually have an agent doing more work on the traditional database. So that traditional general purpose CPU, it's probably steady growth but not as high.

The head node, which is working with GPUs, you can see that's when people talk about ratios. Traditionally, it's one head node CPU manager, eight GPUs, and then become two to-- yeah. It's getting more and more. So over time, that ratio is changing. So we do think the hidden node segment CPU will also grow very fast, much faster than the traditional gen AI CPU.

But the most exciting portion of the market is actually agentic AI. So agentic AI, actually, you are seeing agentic AI server rack sit in between the traditional servers and the GPUs. And those racks are handling all those different workloads to really make sure all agentic works. That market, we think, by whatever, the $120 or $200 billion market opportunity is majority of that large market.

VIVEK ARYA: Over 50% of that market you think is that--

JEAN HU: Is agentic, yes. It's still very early right now. But if you just think about what Matt said, how complex, how dynamic those workloads will be, and we do see significant productivity improvement, especially software engineering side. Those are very complex workload. You really need different call counts. You need really high performance, high call counts. You also need simultaneous multithread, all those high performance, you can manage millions of agents potentially.

VIVEK ARYA: Got it. And the final question there is, actually two last questions, one is x86 versus ARM. So if let's say that agentic opportunity does capture the dominant share of whatever that ARM number happens to be. What advantage, and if I dare ask, what disadvantage does x86 have versus ARM in capturing that?

JEAN HU: Yeah. I'll start. Matt can talk about the more technical side. It's the way AMD has been thinking about it has never been oh, x86 architecture or ARM architecture. It's more about how we provide the best TCO for customers. So for us, the performance is number one.

And secondly, what we have been trying to build is the breadth and the depth of platform portfolio. So if you look at our CPU platform, we have gone through five generations at the Turin, and Venice is next generation. We have the breadth from the core count to eight to 16-core count, addressing enterprise applications to-- the Venice will have 256 core count. So for us, we have all different configuration design points to meet all the different compute to give them best TCOs. I don't know. Technically, you can provide them.

MATTHEW RAMSAY: Yeah, Vivek, I think there are, as Jean talked about the three different buckets, the traditional server workloads for enterprise. I think there is an affinity towards x86 because of the code base. And many of those servers are deployed on prem. Some of them are deployed in third party cloud. But it's basically the same workload depending on where you want to deploy. And I think as these agents generate more traffic to all those traditional systems, then the x86 ecosystem is well positioned there.

Then you have the head nodes. I think that what you really need in head nodes is really high single thread performance, really high bandwidth to, it's not a technical term, but feed the beast of these really expensive accelerators that are in the box. We will do our own head nodes for Helios. I think our large competitor in their NVL racks will basically do their own head nodes.

And then there's a bunch of different XPU programs that are going to be launched by the industry. And I think when we have those conversations with customers, it's much more about what is the performance of your CPU such that I get the best utilization out of this really expensive XPU that I'm buying? It's not an Intel or AMD or x86 versus ARM conversation. It's what's the best server part?

And then when you talk about these agentic racks, it's really about how many concurrent agents can you run per rack or per megawatt of compute? And as Jean mentioned, really, really high thread and core count products, I think our roadmap is differentiated there. And another thing that doesn't get mentioned but I think is important is if you think about the dollars of CapEx that are going to come into the server market, the mission critical nature of these systems, both head node and agentic racks, some of the RAAS features and the security features of the x86 ecosystem.

It's not an x86 architecture thing. It's just we and our x86 competitor have been put through the paces by every hyperscaler and most enterprises on all of these security features for a decade. And you might imagine, you get into these automated flows, where those agents have access to mission critical data, and you want to really have servers that have robust security features there. So we feel we get brought into essentially every RFQ that's in the industry for these servers. And I think we're positioned to win a very high percentage of them. Got it.

VIVEK ARYA: Now your x86 competitor right has their own fab, right? They are or they are planning to add a lot more capacity on the server CPU side. I guess this year, every CPU is being sold. But do you think that there is a scope for share shifts given that your competitor has incremental capacity to devote to this, or do you think you will be adequately served by your foundry ecosystem?

JEAN HU: Yeah. We have been very pleased with our share gain trajectory. I think the last Q1 we just announced, we actually, from a value share perspective, we got to 46%. When you think about this, our success is really tied together with TSMC. From day one, we have been using TSMC. We work together with our chiplet design, with the packaging technology to come up with the best performance server platforms. That continue to benefit us going forward.

We have been planning for this ramp since last year. So if you look at the Q1 performance and the Q2 guidance, when we guide year over year, 70% to increase, And a lot of you actually know is right. The waiver started six or nine months ago. So it is the early planning, how we work with TSMC to make sure we get the support to continue to drive the ramp of the supply.

The demand is tremendous. Supply is still tight. So I think for us for this year, next year, every supply we can get, we can provide to our customers. We're going to continue to do that. It's the planning beyond the 2027 right now. For the longer term, I think we'll continue to work with the TSMC and prioritize data center as how we drive the business growth going forward.

MATTHEW RAMSAY: Vivek, one little thing that I would add to what Jean mentioned there is when we have conversations with the investor community on this particular topic. I think one thing that I've noticed is supply is tight, three nanometers tight. There's a lot of other areas that are tight. And I think we're very well positioned with the relationships Lisa has personally with the executives in that space to get maybe more than our fair share of the incremental. But it is tight.

I think what I've noticed is maybe the investor scoping of what Lisa and the team initially asked for in '26 and '27 maybe didn't imagine how much growth we had already planned for in supply. So things are tight, but we already asked for and already been, quote unquote, granted from TSMC. I think we're having the right conversation there but maybe not at the right starting point from what we were allocated initially.

VIVEK ARYA: And then final one on the CPU side on price. So you mentioned 50% plus growth in Q1 and then 70% plus. I think there was a sense that one could sustain this for the rest of the year, how much of this, Jean, is pricing versus unit growth? And then let me just add quickly part B to that. As that agentic, that third segment becomes a bigger portion, do you think that can command much higher ASPs? When I talk with the ARM camp, they're talking about ASPs in the 3, 4, $5,000 range, which is well above the ASPs that AMD is able to get. So just unit versus ASP and just the trend of ASP.

JEAN HU: Yeah. I think as you mentioned, both the more than 50% growth in Q1 and that we guided Q2 70% year over year, more than 70% year over year increase, 2/3 of that actually are unit growth. So majority of our revenue growth actually come from unit expansion versus ASP increase.

But your point on the ASP going forward, especially with the agentic AI is a great point, is what we're seeing is we're at the early ending of agentic AI adoption. You can already see with the software engineering part of agentic work, autonomous agentic workflow, you really need a very high performance, high core count CPUs that tend to have higher ASP.

So I think a generation over generation, when we go to 2 nanometer with Venice and the future generation to handle complex workload to provide the performance, you tend to see the CPU price continue to increase.

The other thing I would just say is because we have a really large platform, the CPU ASP is really determined by the mix and by the different configurations for different workloads. So it's not necessarily-- for enterprise, the per call ASP is very high because the call counts are relatively low. So ASP is relatively low. But the gross margin is great. So it's very nuanced from our business because of how big our platform is and how complex our different design point we're helping our customers.

VIVEK ARYA: Got it. You spent 2/3 of the time on CPU. So let's talk about--

JEAN HU: It's great.

VIVEK ARYA: --GPU now. So you were at the start of launch of a major rack scale a system to your portfolio. Maybe give us a status update of how you're feeling. When your competitor launched their first large scale up product, it took a while for the industry to get ready, liquid cooling and all those things to get ready. So how do you think about your progress so far? How excited are you about the second half?

JEAN HU: Yeah. We are on track. We sampled my 450. We expected to be launching in second half for Q3 starting point and Q4. As you know, we have been preparing for this launch for a long time. We acquired ZT System in 2024. The whole reason we're acquiring ZT System is to help us to add capabilities in system level design. And we also have worked on the networking side.

So the whole team has been working to prepare this launch. We know it's very hard. It's very complex, very hard. But from an execution perspective, our team has been on track meticulously prepare for all different aspects launch, not only technology software side, but more importantly, the supply chain side, how you make sure the design redundancy with all the components but also supplies even very small components. So you have to make sure you have supplies. That has been the whole team's execution track record. We feel pretty good right now.

MATTHEW RAMSAY: Yeah, Vivek. I think we're anticipating, as Jean said, we already have sampled and we have a number of customers that have full HELIOS racks in their own data center running their full production workload now and testing and doing all those things that you would expect us to be doing at this point in time.

I mean, we learned the customers learned a lot. We learned a lot from some folks that blazed the trail, as you mentioned. And we're going to really focus on a concentrated set of ODMs as we launch for the first number of months until we get up to scale. And not like a few systems but billions of dollars of scale. And then we'll be expanding the ODM ecosystem pretty significantly as we move through next year and get towards the MI500 series and whatnot.

So we feel good about where we are. The initial work that the customer base is doing in their own labs with sampled systems has gone very, very well. And so now, it's about we'll see a pretty significant jump in the fourth quarter of revenue and other fairly significant jump in Q1, and then we'll kind of go from there.

So it is wood-- so touch wood everything's going exactly as we'd hoped it goes. I mean, when you're doing a thing like this, day to day, there's always something. But our job is to keep the duck calm on the top as we kick on the bottom and get ready to ramp this thing up starting in Q3.

VIVEK ARYA: Got it. When you've been asked, Jean or Lisa has been asked, you've always mentioned the opportunity for multi gigawatt scale deployments. I think we have heard about two of them, OpenAI and Meta. You think there is still the scope for us to be pleasantly surprised with more such announcements, even for '27, or do you think that just given how tight the industry is, that more incremental announcements could be further spaced out than that?

JEAN HU: Yeah. We talk about we're really pleased with our partnership, very long term partnership with both OpenAI and Meta. Last year, we established those relationship. And we actually see the forecast from those customers are actually above our original plan for 2027.

I think one of the key milestones we're really focusing on this year is to continue to expand and deepening the customer adoption of MI450. We definitely expect there will be other mega gigawatt customers. We have both across the board hyperscale customers, model companies, and even AI native companies who have been engaging, working with. I think it's important for us to continue to expand this customer base and with the large scale deployment.

VIVEK ARYA: It makes sense. And then how do you expect the two customers you have announced to allocate market share because they have a number of supplier options. I know you have a somewhat unique engagement with them through the granting of warrants and whatnot. But if we set that aside, how is your visibility into getting I think the two gigawatt each right from those customers for next year?

JEAN HU: Yeah. Vivek, as you know, the lead time for MI450 is quite long. So we really have to start to plan and have the orders from customers for next year for the second half launch and the next year. I think the supply chain overall is tight next year. But since we have been preparing to support our customers, we do want to ensure we get supplies for what they need.

I think both are coming in higher than our original plan, but we are continuing to increase the supply. Those are two of our most important anchor customers. Of course, we also have a warrant, which we're going to be able to share the upside if we get more revenue. Our stock price is higher. They will share upside to that is absolutely the common interest for both companies want to drive the adoption of MI450. We feel pretty good with both of them. We think we can get the right supplies to support them.

VIVEK ARYA: Got it. What do you think. Having someone like a TSMC is almost adding a level of discipline right in the ecosystem because it's just harder to double count if everyone has to go to the same source. I mean, I know it sounds obvious, but do you see that also?

MATTHEW RAMSAY: They quite disciplined. I think they'll continue to be. I mean, these are sensitive topics for us to talk about in these forums. But like we really, really, really appreciate the support that TSMC giving us as a partner and the way that they're working with us to bring more capacity online to support not just the HELIOS and 450 growth but the significant server TAM expansion that we're seeing and planning for, not just for this year but for all of next year and conversations going into 2028 on wafer capacity support demand.

So we really appreciate the work that they're doing on our behalf, and I think that there being, as they've always been, disciplined in the way that they bring capacity online and do their diligence two or three levels deep in the customer base to understand where the organic demand is coming from to support the capacity that they're bringing on.

VIVEK ARYA: Makes sense. And then the final one, Jean, is on just memory cost inflation. I guess, when you wake up and see DRAM. I'm sure DRAM price is not what you see when you wake up. But when you see that cost inflation every day, is that a good thing because it gives you more pricing power, or is that a bad thing? Because now, you have to absorb that cost and hunt for more supply. What are the pros and cons of that memory cost inflation for you?

JEAN HU: It's a really a great question. First, for us, we really want to make sure if there's a longer term agreement with our suppliers, we can secure the capacity. That is the number one thing. When you have a larger business at our scale, you have to plan ahead.

And secondly, we think about strategically, if there's memory cost increase, we do need to figure out a way to share with our customers to absorb that increase together. I think the most important thing for us is we actually don't want to use the shortage of control of memory to just increase the price. We want to make sure customers help us to absorb the cost increase, but we are more strategic with the long term relationship.

I think from an industry perspective, the memory cost increase at this kind of level, we have never seen. The price is really high. I think it does increase the CapEx spending for other customers where they spend the CapEx. I think in the longer term, just like all the industry always have a new capacity add up. You also have a new innovation on how you utilize memory more efficiently. It will figure out itself in the longer term. That's what we believe.

VIVEK ARYA: Got it. But do you have long term agreements in place for the majority of your memory needs, or do you think that this is a cyclical high, and maybe there will be more normalization of that input cost?

JEAN HU: Yeah. From a planning perspective, because our manufacturing cycle is so long, if we plan for the MI450 ramp this year, we had to start to talk to our suppliers last year or year before. Similar situation is right now, not only with the very tight supply environment. Not only we're planning 2027. We're planning 2028. So for us to continue to make sure we can support our ambitious plan to grow the company, we have to plan ahead yeah.

MATTHEW RAMSAY: In the server market and in the PC market, the vast majority of the DRAM is actually acquired by our OEM and ODM partners and the hyperscalers themselves. That DRAM doesn't run through our P&L. And we work really, really hard with all of the customer base to try to make sure that we're matching like the memory supply that they're able to get and the CPU supply or the GPU supply that we're able to provide together to where there's not mixed matching of supply in the industry or hoarding of anything. And so we've been working really hard on that.

But there's certainly increased input costs to end devices in certain parts of the consumer and gaming market where there have been some anticipated market impacts. And I think we've given you guys our current view of that and the guidance that we've talked about. But we're working really hard on behalf of the whole customer base to try to make sure that the memory pricing and the matching of components is relatively seamless in a time like this.

VIVEK ARYA: And just final question before we close. When you look across your customer base, Jean, how far is the visibility like that they are willing to provide to you now? And when you get that, what are you doing differently in your business operationally to deliver to those?

JEAN HU: Yeah. I think because the supply chain environment is really tight, customer really plan ahead. So we have a very good visibility into 2027 and beyond. I think our customers and ourselves, especially when you think about the data center investment, very large CapEx planning, very long planning cycle, they actually plan way ahead. 2027, we feel really good about both the visibility of demand side and the visibility of supply side. And of course, 2028 and beyond, that's everybody's working on.

VIVEK ARYA: On that positive note, thank you so much, Jean.

JEAN HU: Thank you.

VIVEK ARYA: Appreciate you.