MODERATOR: All right, everyone. Welcome back to the Barclays Global Tech Conference. I'm pleased to have Jean and Matt here from AMD. Thank you for joining.

JEAN HU: Thank you. Thank you for having us.

MODERATOR: No problem at all. So why don't we start with the question that's on everybody's mind as we exit kind of 2025 and going to '26 year. There's been a ton of AI spend announced. We aggregate over $3 trillion. The compute and networking portion of that, we can argue about all day. But I think the conversation has centered around the feasibility of actually deploying all of this spend in the timeline that's been laid out. Maybe talk to me about what you're seeing in terms of the ability to deploy this and then how it's benefiting AMD, in general.

JEAN HU: Yeah, thanks for the question. First is the way we look at AI is we're really in the early stages of a multi-decade investment cycle. If you think about, it's very transformational technology which will change the global economy fundamentally. So it's absolutely the case that if you have a more data center, more compute, you can actually generate more intelligence and more capabilities.

The CapEx spending is super high and it's quite significant. The way we think about it is when we talk to our customers, you can see they are the ones, the hyperscale companies, they are increasing CapEx spending. And frankly, they are all very well-capitalized companies. They are funding it through free cash flow. And so the whole ecosystem is really funding the investment.

More importantly, what we hear from our customers is they are increasingly more confident about the business model for AI. Not only they are seeing real workload, the cases, they can see the productivity improvement. Also the unit economics is also improving. Inference costs rates are coming down. So I think now what they're telling us is actually they're constrained by the compute, by the infrastructure. If they have more compute, they actually can support more applications. They can tie their investment with revenue return on investment from that perspective.

So we do think everybody is working very hard to bring up more capacity, which, of course, we provide the significant compute not only on the GPU side but also on the CPU side. We see the tremendous demand for our compute in both accelerator side and the CPU side. I think it will benefit us from longer term.

MODERATOR: And increasingly of very, very late, you've seen the debates shift more from general-purpose silicon to can custom silicon scale across multiple customers, and how does that impact general-purpose silicon providers. Maybe for the both of you, just what do you think about the ability for a chip that was designed for a specific customer to be used more broadly? And when you see someone like a Google having success externally, do you feel like that cuts into your TAM, or maybe lay out why that would be a different swimlane than what you're in today.

JEAN HU: Yeah, I'll start with a very high level then Matt can provide more colors on. If you really think about it, the AMD's view has always been-- we see $1 trillion data center market opportunity. Of course, majority of them are accelerated opportunity. We always said include both general-purpose compute and you call the ASIC or customer silicon. And we always have said, the ASIC or customer silicon is going to be 20% to 25% of that market opportunity. So it's huge, that we always believe.

And we always said it's really about different compute for different workload. But consistently, the programmable architecture we have, we can support the more variation of models, workload training, inference, pre-training, post-training, that continue to be the flexibility customers are requesting. Of course, there's the most recent debate about Google TPU and general-purpose GPU.

We always have the same consistent view on TPU, Google. What they have done with Broadcom, very good, but they are still very specific from workload support perspective. Customer wanted flexibility, overall. So majority of the market we continue to believe it will be general-purpose GPU. I think Matt--

MATT RAMSAY: Thank you, Jean and Tom. Thank you guys for everyone at Barclays for having us here. I think it's interesting. First, one perspective, too, is from the model company's perspective, whether these are AI native model companies like OpenAI and Anthropic and others, or whether they're hyperscale companies with their own models, that's a super competitive space in and of itself.

There will be-- recently Gemini 3 published, and it's an incredibly good model, and that got a ton of attention. Next month or the month after, another model that's trained on-- whether it's trained on ASICs or whether it's more likely trained on GPUs, that'll be a better model than that model. And they'll continue to be this leapfrogging. And what we've observed is as a big swing in investor conversation around this. But you should anticipate as investors this being a continuation of these model companies getting better and better.

And as Jean said, getting the right silicon to do the right type of work is super important. We've tried to architect our Instinct family as we go forward at the rack scale and MI450 to be general purpose in nature to serve all of the customers. The flagship product of that portfolio would be the MI455 that will ship to OpenAI and a bunch of other folks.

There's also an MI430 version where we've taken the main compute chiplet out and put in a separate compute chiplet that has more floating point that's akin to what's been done in the HPC market. The market doesn't have to go completely GPU or completely custom. There's a lot of semi-custom opportunities in between to get the right type of silicon to do the right type of work. And I would just encourage this audience not to maybe overreact to the news of the day. This is going to be a super competitive market on the hardware side. It's going to be a super competitive market on the model side, and you're going to get new data points that come out all the time.

As Jean said, we've been consistent in our own modeling inside of AMD that 75% to 80% of this market is going to be programmable load store architecture computing at the GPU level, and that's where our customers are asking us to provide consistent annual cadence, system-level competition. And that's what we're going to go and do. But there's certainly a model-- there's certainly a market for ASICs, 20% to 25% of a trillion plus TAM is a big market. And there'll be folks that are very successful in doing that. So that's kind of our perspective right now.

MODERATOR: Perfect. Yeah. So take that 25% out of the pie, the 75% left. If you look at your long-term kind of TAM talked about a trillion, NVIDIA talks about something 3 to 4 trillion. Could you maybe walk through why their TAM is so much larger? Is it a function of gross margin? Is it a function of networking? What are they adding in that you guys aren't, because you would assume you guys are probably closer apples to apples than those numbers would.

JEAN HU: Yeah so let me clarify our TAM. What we are focusing on is really silicon addressable market opportunity for AMD. So our TAM, when we talk about the over trillion dollar data center TAM, we include the accelerator, which general-purpose GPU, ASIC or customer ASIC, or how you call it. We also include our expanded TAM on the CPU side, also networking, scale-up networking which we also have an offering.

So those are what we focus on. We actually don't include the rack. We don't sell rack. We don't include cable, all the other solutions component to that build up to the rack or clusters level. Of course, we also don't include the data center infrastructure build out. Those are not what AMD is focusing on. So of course, what other competitors talk about, their TAM, it's very different. So that is what AMD is focusing on.

MATT RAMSAY: Yeah, Tom, I think the growth rates of the TAMs, regardless of how you define them, all those curves look very similar. We have a data center business segment that is our server CPU business, our data center AI business, our scale-up business. What we tried to forecast at the analyst day a few weeks ago was AMD's TAM.

We're not in the business of forecasting data center CapEx or NVIDIA's TAM or Broadcom's TAM or anyone else's TAM. We're thinking about our Silicon TAM that we can directly address with products that AMD will and could offer us. That's all we've included. So there's certainly, if you want to forecast data center CapEx, that would include power and buildings and water and cement and all kinds of other things that AMD's never going to sell. So we just tried to forecast our own TAM.

MODERATOR: I want to move to something a bit more customer-specific in OpenAI. I thought that it was really an unlocking of investors' minds when they saw the deal with OpenAI and was like, wow, this really brings AMD to the center fold of the conversation with NVIDIA and Broadcom in terms of, one, ability to provide compute that is very, very real in the next 12 months. And then, two, had a structure of the deal, which was a bit unique, but also very interesting in that your economics kind of scaled with the deployments as well.

Maybe, one, talk about why you structured the deal you did and the way you did with OpenAI. And then two, just judging by general math and what you've said, it's about a gigawatt of deployment in the back half of next year, how ready is the ecosystem to get that out there with all the other compute announcements? And do you feel secure in your ability to get the product that you need and have those deployments go to market?

JEAN HU: Yeah, yeah. Thank you for the question. We are very pleased with the partnership with OpenAI. It is a definitive agreement, not LOI. We signed with OpenAI for six gigawatt over several years. I think, as you mentioned, it is a win-win situation. The framework is really based on-- they scale up the deployment of AMD's MI450 and the next-generation product. And at the same time, there is a performance-based warrant, is when we ramp up our revenue, which creates value for shareholders, then they also get a warrant from the partnership that we have.

So that is how it's designed. But to be clear, we have been working with OpenAI for a long time, multi-generations starting with MI300 and then MI355 and then now trying to deploy MI450. So the first gigawatt is a commitment. We'll start to deploy in the second half of 2026, but it will ramp into 2027.

And the whole ecosystem we are working with really focus on the planning from the data center, CSP selection, the power to supply chain, our ecosystem partners to help us to ramp the MI455. Those are the overall system we have been working with the partners. So we feel pretty confident about the execution part, where the starting ramp of second half and then going to 2027. Of course, the relationship is multi-year multi-generational. I think we are both very motivated to continue to drive the future partnership too.

MODERATOR: Yeah, we saw earlier this year at both the analyst day and then previously Sam was on stage with you guys for a current period of time talking about how they were very involved in the design of this product. And then you actually got to see Helios in person. Can you talk about where the differentiation is versus other rack architectures?

And then maybe customer engagement, since you've had that out there, I would assume customers get a little bit of an earlier peak than us, but something that customers are coming to you and saying, wow, this is really unique. We would prefer this solution versus what we've seen so far.

MATT RAMSAY: Now it's a good question. So one of the things that we focus on really heavily with the work between AMD and OpenAI is with them being arguably the leading model company in the world, I mean, there were weekly level executive engineering engagements back 18, 24 months. It wasn't like we just popped out with a product and we had an announcement.

They, among other customers, have had influence and given us feedback on the design of the GPU itself and some work that we've done in our rack software stack itself. And then you think about what we're doing in the roadmap with the Helios rack and how we worked with Meta on that around OCP to have an industry-standard, compliant rack that you might imagine. We could make more dense as we move forward as because of the double wide rack footprint.

The engagement level across the board with customers has been a very deep one. I think Lisa has talked at the analyst day and in other forums about having multiple multi-gigawatt engagements over the MI450 frame, and OpenAI is a critical partner both there will be others as well. One of the really exciting things for us about the close, close partnership with OpenAI is that they do deploy their infrastructure in many places with a number of hyperscalers, with a number of neoclouds.

And the work that we were doing at AMD anyway, on MI355, MI450, and MI500 series after that was to partner with a very wide range of customers and push our infrastructure into all of the CSPs and all of the neoclouds on our own. And we were having great progress in doing that. And you saw the customers we had in our event back in June.

Now we have an additional really large customer pulling us to scale at all of those different platforms as well. And that gives a breadth of other customers confidence that look through the partnership with OpenAI at various places in the industry, AMD will have scaled infrastructure that we can then build our work on top of.

And so the engagements with customers that were happening anyway both deepened and accelerated in time, since people have gotten a view as to what the OpenAI deal looks like and the fact that the Helios rack has been unveiled to the world. So it's been an exciting six months, and we're really pleased to move forward with the breadth of the customer base.

MODERATOR: And then one for Jean, on that same topic, you talked about overtime with volume, the data center, GPU business, getting up to corporate gross margins and then potentially in the future maybe being better. But rack-scale architecture obviously brings into account a variety of other subsystems, components, et cetera, that generally are a margin headwind. Can you talk about, as you see Helios ramp, what that does to gross margins on the corporate level.

JEAN HU: Yeah. To be clear, we actually don't sell the Helios record level systems. Our focus as we talk about our TAM is really silicon. It's more focused on the high value added piece, which include GPUs and CPUs and sometimes scale up networking. So when you think about our business model, it's really not changing from what we do today. We really want to focus on the high value added piece. And at the same time, we do provide a reference design for our partners.

And we are committed to open standards so everybody can also make money. And from TCO perspective, it's better TCO for customers too. On the gross margin, we always have been focused on right now, the priority is market share expansion and the gross margin dollar pool. As you can see, the market expanding very quickly. That is what we focus on right now. So right now, the GPU gross margin is slightly below corporate average. But going forward, when we scale our business, when we really optimize the solutions for our customers, that we do think a gross margin will go up.

One thing to be clear is right, we talk about it at our financial and state, if you look at our strategy at company level is we're building a compute platform which including GPU, CPU. It all includes adaptive compute and other solutions for different end markets. From a company level, we always leverage our investment across all the platforms.

Same thing on the gross margin side-- we do have a multiple drivers. We can continue to improve the company's gross margin. On the CPU side, we're getting into commercial market, which has higher gross margin. Same thing on the client side. We see tremendous opportunities to continue to improve gross margin. And then our FPGA business is very gross margin accretive.

So when we add it together, take a step back at the company level, we are driving the gross margin to be at a 55% to 58% as our long-term model, and we feel very comfortable about that trajectory.

MATT RAMSAY: Yeah, Tom, just to reiterate what Jean said at the beginning, because we continue to get some questions about this, is we are not selling racks. We are not selling servers. Our OEM and ODM partners will sell the racks and sell the servers. We will work extremely closely hand in hand with them through our DP system services team to license the reference design. Often, games will license testing and testing programs to make sure they can test the racks and deploy the racks.

We'll help provision the supply chain for all of the other components, whether that's cables or connectors or power supplies, or a whole laundry list of things. And we will be at AMD responsible for delivering the servers to the model company, to the hyperscalers, making sure that they run workload and that they run efficiently.

But all the other pass-through components that are not part of the silicon TAM will not run through our P&L. So just to be clear about that because we've gotten this as you go to rack scale, what happens to margins. But we're going to be a fabulous semiconductor selling semiconductors the same way we've always been. So hopefully that's pretty clear.

MODERATOR: That's why we ask it on stage. All right. So next thing is NVIDIA has brought to market a CPX, which is an interesting, at least from my perspective, a new type of compute where you would imagine it be doing something like a pre-fill functionality. You're seeing this ecosystem evolve very rapidly.

That to me looks more like a custom piece of silicon or a CPU in general. But does that design choice mean that you will necessarily fall in that direction? Is there a reason why they would go in that direction? And a better question, because you obviously don't want to talk about your competitors, is what could you guys do in next generations that CPX chip does that would improve your performance?

MATT RAMSAY: No, that's a good question. I think we do a ton of work with the customers on workload characterization of AI workloads. So there's obviously this growth at different rates of prefill and decode. And they've made a certain design decision around in certain instances doing dedicated piece of hardware for that.

We've evaluated it extensively. In the MI450 time frame, we're doing PD and software. We're not yet convinced that the relative ratios between free field and decode and other part of the inference workload pipeline are yet fixed enough to make dedicated hardware decisions. But we have some flexibility as well.

I mentioned earlier, the ability to maybe take our overall platform and substitute in different compute chiplets into the roadmap over time. So you don't need to do, in our architecture, at least a brand new piece of total silicon to subsegment parts of the workload. There are certain places where pre-training and training and things are getting closer to inference in the way the workload is characterized.

There's certain parts of the algorithm stack that might slow down and be more amenable to a fixed piece of silicon versus other pieces that continue to evolve very, very quickly where you want flexibility. And so I think for us, we've not yet made that choice.

And we're in the current gen doing PD and software, but we're evaluating all parts of both the training, pre-training, and inference software stacks as to which part might require some more general silicon and which part might require some more dedicated silicon over time. And our customers all have a view of that as well. But we've evaluated it super closely. And right now we're doing PD and software. But that may change going forward as the algorithms mature.

MODERATOR: Another one on the technology side, scale-up architectures is a huge debate today. You guys have been committed to UAL longer term. First generation is UAL tunneled over Ethernet. More recently, you've seen with Amazon T4 using NVLink Fusion, at least in some SKUs.

With you guys offering a system architecture or a footprint for others to engage with, how do you see the world evolving? Do you think that ultimately, everyone interacts with the large, general-purpose silicon providers in terms of back-end ecosystem? If you get UAL up and running, will people use yours as well? How do you see the world evolving, and where do you see scale-up architectures moving in the next three to five years?

MATT RAMSAY: Tom, I think what we care about is driving TCO at the rack in the data center level and adding-- one of the areas that we want to support open standards, just like we do in our ROCm software stack where we provide a lot of openness to the ecosystem, is on the networking architectures we choose. For example, for scale-out networking over Ethernet, we have built into Helios the flexibility to have different switch vendors that do scale-out ethernet.

On the scale-up domain, as you mentioned, we've been doing a technology inside of AMD for five or six generations in our server business called Infinity Fabric that's done coherency across chiplets, across sockets, across racks in our server business. Out to supercomputing scale, we've licensed that to the UAL consortium that they've ratified to be 1.0 standard of UAL standard in the initial implementations of Helios that are going to launch in the second half of next year.

We're using that traffic that we're really-- that's critical-- the UAL traffic, that's Infinity Fabric, coherency traffic. That's what we're really, really focused on. The transport layer, we're a bit more agnostic to what the customer wants to do. And there may be some in the 2027, '28 products, there May be some opportunities for us to support native UAL silicon that can have some power and latency advantages. And I think we would expect many of our customers to adopt that because there are some technical advantages to doing that.

But if there are customers that want to continue to tunnel that traffic over Ethernet or scale-up ethernet or other protocols, we're totally fine with that. What we want to do is make sure that the coherency works on the functional level, and it's performant, and the underlying silicon transport protocol is going to be driven by the needs of the customer. And so we have some of our own technical opinions about which one might be better than others. But that's not our business. I mean, the customers are going to decide what their scalable architecture is going to look like, and we're going to make sure that our coherency protocol is validated over whatever transport they decide to use.

MODERATOR: So we went very deep in the tech, pulling back out to the macro. News on China, again, over the last week, we see several iterations of this. I would say the most recent was, there was some ability to sell, but it seemed like customers in China were not taking that product. Maybe just whatever. I know it's a sensitive issue. How do you feel about the current arrangements? What's changed for you, and do you think it really changes the dynamic of Chinese customers taking your product?

JEAN HU: Yeah, the situation with China probably is probably the most dynamic. Every day there's some news. I think we do expect-- based on the most recent news on H200, we do expect we would expect to be treated the same for our MI325 product, which is similar to H200. Of course, we support administration's effort to help the whole industry, but at the same time, they're still working through the details.

Just like all the different complications with the situation in China, so on the MI325, we will apply for licenses once they work through the details. But then as you mentioned, there's still China customer demand question. We still need to figure it out.

On MI308, as we guided the Q4, we did not include any revenue from MI308 because of the uncertainties we have. We did obtain a few licenses. We are working with our customers on the demand side. They're just always very uncertain about what's going to come or not. So we're going to monitor the situation, make sure we comply not only with the US government's export control rules, but also on the China side.

MODERATOR: Great. I want to hit a couple of rapid fire as we wind down time here. And client continue to see really good share gains. ASPs have been a huge positive story as the year has gone along. I actually think that ASPs have held them a bit better than even you guys have described in the back half of the year. What's driving that, and can that continue-- should we be seeing some normalization there into Q4, Q1?

JEAN HU: Yeah. First, we are very pleased with our client business performance. If you just look at the last three quarters, we literally increased revenue by 60%. And the majority of them actually is driven by ASP expansion. The major reason is not only we have been going up the stack to really go to the premium PC, not only desktop side and also on the mobile side. And secondly, we're getting to enterprise commercial market, which is also a higher margin product.

So overall, that has been our strategy. We do believe we have the best technology and product portfolio right now in the PC market. So we'll continue to drive. We should expect the consistent ASP trend just like what we have seen in the last three quarters. The team is very excited, not only about Q4 and the next year, how we can continue to execute to expand our market share.

MODERATOR: And then your competitor has talked about supply tightness, incline as well as server. Are you guys seeing this as well? And is this an opportunity for you guys to gain more share, or how do you view this dynamic?

MATT RAMSAY: Yeah. It's a good question, I think, for two things. One, in the client side, as Jean said, we're going to continue to push to gain share in enterprise in particular, hold a very, very, very strong position we have in premium desktop, where the ASPs and margins are quite strong. And we'll certainly work as best we can to support our customers. If there's any shortages in the industry, we'll have to be really strategic about that from a margin perspective, but make sure that we can step in and help the customers where needed.

And then on the server side of the business, which is something that we didn't get to quite in this conversation, we continue to see a pretty rapid expansion in our enterprise footprint. One of the statistics that got maybe overlooked with all the things that we threw at the investment community at the analyst day, was we've expanded almost doubled our enterprise customer count during 2025, and we'll see how the land and expand goes there.

In addition, pretty much at all of our top hyperscale customers where our market share in server is fairly high, we've seen an expansion of the TAM. As those folks have deployed inference, you've seen significant amount of additional CPU demand to support the inference traffic, whether it's agentic inference, whether it's storage servers. I mean, there's head nodes, there's some places where people are running inference on server just across the board in the server portfolio.

We've seen there was a thesis in market for some period of time that AI was going to be cannibalistic to the CPU server market, and I think we're seeing the exact opposite happen and in an accelerated way and broadening out of that trend. So yeah, the CPU portfolio-- the shining light of AI has got a gravitational pull to it with investors, but the underlying CPU businesses and AMD are in a great spot.

MODERATOR: Well, we've run out of time here. I very much appreciate you both being here. Thank you so much. And it sounds like things are going quite well.

JEAN HU: Yeah, thank you so much.

MATT RAMSAY: Thank you, everybody.