9-9-2024 Goldman Sachs Technology Communacopia and Technology Conference

TOSHIYA HARI: We'd like to get started. Good afternoon, everyone. My name is Toshiya Hari. I cover the semiconductor space for Goldman Sachs. I am very honored, very happy, very excited to have Dr. Lisa Su from AMD, chair and CEO. I'm pretty sure everyone knows Lisa. So we will go straight into questions, skip the intro.

First of all, Lisa, thank you so much for coming.

LISA SU: Yeah, thank you for having me. It's great to be here.

TOSHIYA HARI: So I think this time last year, we were on this stage. We kind of kicked off the conversation by me asking you, what are your key priorities. And you said something along the lines of AI number 1, AI number 2, AI number 3.

LISA SU: I might have said that.

TOSHIYA HARI: I think you've executed really well since last year. You've grown your data center GPU business from essentially 0 last year to per year guidance 4.5 billion this year. Reflecting back, in what ways have you and your team outperformed your expectations, again, specifically in the field of AI? And going forward, again, what are some of your focal points?

LISA SU: Yeah, absolutely. Well, again, thank you for having me. It's been a remarkable year, I would say. So much has happened. I think, all in technology, we're moving faster than ever. And in the last year, I mean, if you look at what we've been able to do, we've launched MI300X in December. It's had just tremendous customer traction. And customers have been really excited about it.

We have several large hyperscalers, including Microsoft, Meta, Oracle that have adopted MI300 as well as all of our OEM and ODM partners. When I think about what do I believe we've done the best over the last, let's call it, nine months or so, it's really been the progress on software. That was always a big question around how hard is it to get people into the AMD ecosystem. And we've just made tremendous progress with our overall ROCm software stack.

We've now worked with some of the most challenging and largest models, and we've seen them get performance. And in some cases, with certain workloads, even better than the competition, which is exciting. And then we're continuing to build out the entire infrastructure of what we need. So we just recently announced several software acquisitions, including the acquisition of Silo AI, which is a leading AI software company. And we just recently announced the acquisition of ZT Systems, which also builds out the rack scale infrastructure necessary.

So sitting here and talking about priorities going forward, certainly AI is a huge priority for us. But when I think about AI, it's actually end to end AI. Of course, the data center component is very important. But I'm a big believer in there's no one size fits all in terms of computing. And so our goal is to be the high performance computing leader that goes across GPUs and CPUs and FPGAs and also custom silicon, as you put all of that together.

So I think lots of opportunity, lots of focus on the roadmap going forward. But it's been a pretty exciting year.

TOSHIYA HARI: That's great. You shared a 2027 AI accelerator TAM forecast of $400 billion earlier this year. A lot has happened since then. How have your long term expectations evolved since that time? To the extent you are more bullish on the opportunity set, which applications, which end markets have you seen the most upside, if you will?

LISA SU: Yeah. When we originally talked about a $400 billion TAM in the 2027 time frame, I believe many thought that that was high. And actually, I think, as time has passed over this last year, I think we feel very good about that overall TAM. And I think the main reasons for that is we are still so early in this AI computing cycle.

And whether you're talking about training of large language models, or you're talking about inferencing, or you're talking about fine tuning, or you're talking about all of these things, the workloads will demand more compute. And for that reason, we feel very good about the overall market. Now, within that market, when we talk about the accelerator TAM, it's not only GPUs. We believe GPUs will be the largest piece of that $400 billion TAM. But there will also be some custom silicon associated with that.

And when we look at our opportunity there, it really is an end to end play across all of the different compute elements. So from that standpoint, we feel good about it. We're also seeing-- many people have said, inference will continue to increase over time. We're certainly seeing that. Training is very, very important. But inference is increasing over time.

And then the fact that you actually see some mixture of the workloads where people are doing inference and continuous training, as you think about how to really tailor these models, those are all important trends that we're seeing that are leading to the belief that the TAM growth will be there.

TOSHIYA HARI: Got it. I have one hardware question, and then a software question. On the hardware side, you announced at Computex, I believe, that you'll be transitioning to a one year product cadence in data center GPUs. I'm curious, what catalyzed this change? Was it based on customer feedback? Are they asking for higher frequency, if you will, or was it a competitive response?

LISA SU: Yeah, definitely, when we look at the roadmap today for AI-- and we have announced a one year cadence, we've accelerated our investments in both hardware and software as well as systems-- it is all customer driven. We spend a lot of time with our largest hyperscalers and our overall partners. And what we see in the ecosystem is the idea that people have different data center needs.

Of course, you have the largest hyperscalers who are building out these huge training clusters, but you also have a lot of need for inference, and you have a lot of need-- some are more memory intensive workloads that would really focus there. Some are more power, data center infrastructure constrained. And so they want to reuse some of their data center infrastructure.

And so what we've been able to do with our MI325-- that's planned to launch here in the fourth quarter-- and then the MI350 series and the MI400 series is really just broaden the different products such that we are able to capture a majority of the TAM with our product roadmap. So lots of conversations with customers on what they need and where they're going, and ensuring that we're aligning our roadmap and our investments with that going forward.

TOSHIYA HARI: Got it. Software used to be one of the sticking points for AMD, and what I would have conversations with investors that that was the commonly asked question. You touched on this a little bit at the very top of the session. But where do you see yourselves today, from a software perspective, given the recent iteration of ROCm?

You've also made M&A moves, if you will, from a software perspective. Where are you today and what's still to do going forward?

LISA SU: Yeah, absolutely. Look, software has been a huge priority for us. And if you think of all of the steps, ROCm has been around for a while. Actually, ROCm is our version of the ecosystem. And we use open source ecosystem. But what has been necessary is for us to really practice ROCm in the most difficult environments.

So over the last 9 or 10 months, we've spent a tremendous amount of time on leading workloads. And what we've found is with each iteration of ROCm, we're getting better and better. So in terms of the tools, in terms of all the libraries, in terms of knowing where the bottlenecks are in terms of performance. So if I just give you an example, customers that we worked with, let's call it, early on, we've been able to demonstrate on some of the most challenging workloads that we've consistently improved performance.

And in some cases, we've reached parity. In many cases, we've actually exceeded our competition, especially with some of the inference workloads. Because of our architecture, we have more memory bandwidth and memory capacity. And what that is it's really good for large models when you can fit them on a single GPU versus having to go into multiple GPUs.

But the key with the software is how long does it take to get to performance because time is money in this world. And whereas with earlier versions of ROCm, it might have taken a couple of months for workloads to get performance, we're seeing, in the latest iterations of ROCm-- there was one company that we were recently working with, which was very much using PyTorch as their framework foundation. And we saw, in this case, it was out of the box performance on PyTorch and, within a week, exceeding our competition.

So it just shows you that there's been a ton of heavy-lifting on ensuring that the entire software ecosystem is there. And we're not done. I mean, that's part of the reason that we announced the acquisition of Silo AI, which is a very, very talented team that is really there to help our customers migrate to the AMD ecosystem as fast as possible.

TOSHIYA HARI: Great. You mentioned time is money. You also announced the acquisition of ZT systems recently. I know the deal hasn't closed. But what specific capabilities and competitive advantages do you attain once ZT is integrated in AMD vis a vis you going at it as you are today?

LISA SU: Yeah. So maybe if I take a step back and talk about what we think success factors are in the AI world. I think with our size and scale, we believe that we can be one of the most strategic computing partners to the largest hyperscalers as well as the largest enterprises. And as we spent time with our customers and really looked at what would be necessary three to five years down the road, it was clear that the hardware roadmap, super important. We've made significant investments in there.

The software roadmap, we just talked about with ROCm. We've made significant investments there. But the rack scale infrastructure, because these AI systems are getting so complicated, really needs to be thought of and designed at the same time in parallel with the silicon infrastructure. So we're very excited about the acquisition of ZT.

As you said, it hasn't closed yet. So we expect to close in the first half of 2025. What we see is a couple of major factors in terms of really addressing the future. And these are the largest scale AI systems. The first is just designing the silicon and the systems in parallel. So the knowledge of what are we trying to do on the system level will help us design a stronger and more capable roadmap. So that's certainly a big advantage.

The second reason that we're quite excited about it is back to this comment of time is money. The amount of time it takes to really stand up these clusters is pretty significant. And we found in the case of MI300, we finished, let's call it, our validation. But our customers needed to do their own validation cycle. And much of that was done in series. Whereas now, with ZT as part of AMD, we'll be able to do much of that in parallel. And that time to market will allow us to go from, let's call it, design complete to large scale systems running production workloads in a shorter amount of time, which will be very beneficial to our customers.

And the largest thing is, look, we believe collaboration is key. And so this is an area where there is no one size fits all as it relates to a system environment either. Different hyperscalers want to optimize different things in their systems environment. And we want to be able to have the skill set to do that and do that really with what I would call best in class talent with the ZT team.

TOSHIYA HARI: And again, was this an example of a customer or customers of coming to you and saying, hey, why not make this move to time up, speed up your process, or how did it sort of come about, if you will?

LISA SU: Yeah, I would say it's actually the opposite. If you think about-- and I've said this before, Toshiya-- everything that we do is really making bets for what we think are important three to five years from now. And so the work that we're doing today on the MI300, 325, 350 series was actually decisions made a few years ago, our decision to focus on chiplet architectures and really do that.

This is also a bet for what we think the future is going to be like. And we spent a lot of time with our largest customers. And when I look at what our priority is, look, we can build great technology, which I think we are doing. But by really making it easier for customers to adopt, its time to market is ease of adoption and its adding more value into the equation, it became clear that we wanted more systems capability. And again, ZT is one of the leaders in AI systems. And they're also similarly-- their customers are very much our customers. And so it made it a very logical choice.

TOSHIYA HARI: Got it. I have a ton more AI questions, but I want to shift gears a little bit. The server CPU market, which continues to be a very important market for AMD, went through an extended correction. The market finally seems to have turned the corner from a demand perspective. What are your forward expectations for server CPU, and how would you differentiate what you're seeing in the cloud hyperscale space versus enterprise?

I think some of your customers are increasingly sort of worried about things like space and power consumption. Could, innovation like Genoa and Turin catalyze a replacement cycle in server CPUs?

LISA SU: Yeah, absolutely. I am pretty happy with some of the server CPU market trends. I think what we've seen is traditional compute is important. So as important as accelerated compute is, there are lots of workloads that run on traditional CPUs. And from a upgrade cycle standpoint, although there was a little bit of a degrade, a sort of a delay in the upgrade cycle, we are seeing customers upgrade today. And that is both cloud and enterprise.

I think from the cloud standpoint, it's very, very beneficial to upgrade some of the infrastructure that is four or five years old. You get a significant power savings, you get a significant space savings, and overall TCO benefit. Genoa or our Zen 4 family is extremely well-positioned. And so we've seen very strong adoption with the new capabilities their.

We're very excited about our Zen 5 cycle. Our turn cycle is coming up shortly. We'll be launching that here in the fourth quarter. And we see lots of excitement around that as well. And then going forward, as we think about just sort of decisions that people make, whether you're talking about cloud or enterprise environment, I think people are just becoming much smarter about what a difference it makes if you talk about the underlying silicon.

So whether you're making a choice of something that's optimized or let's call it performance optimized, we actually expanded our CPU portfolio because we believe that different variants would get you better TCO. And we're seeing that play out with our customers.

TOSHIYA HARI: Got it. In terms of the competitive landscape in server CPU, 5, 6, 7 years ago, you were low single digit market share, I believe. And today, from a revenue standpoint, I think you're in the low 30s. I do think you've had massive success on the hyperscale side. You're at or above 50%, I believe.

On the enterprise side, it's been a little bit slower. But at the same time, you've been much more vocal in terms of the penetration or the momentum you have. So what are your thoughts on the enterprise side? And what needs to happen for you to inflect higher and for your market share position to mirror what you have in hyperscale?

LISA SU: Yeah. I mean, it's been really exciting to see how the data center market has grown for us as a business. When you think about where we started data center business, as you said, we were low single digit share. It was similar percentage of our revenue. In our last quarter, I think data center was over 50% of our revenue. So we really are a data center first company.

And when you look underneath that, customers are really adopting when they need the best technology. So for the hyperscalers, I think their adoption rate was faster and earlier, especially on first party workloads, because it was so clear that the TCO advantage of adopting AMD was so clear. As you look at enterprise and some of, let's call it, the third party adoption, they've had many other things on their mind. And so they weren't necessarily focused on CPU versus CPU.

But at this point, it's all about TCO and it's all about efficiency. And one of the things that we've done is the more we have interacted directly with end enterprise clients, they want the best technology. And so we've put more field application engineers in place. We've done quite a bit more of these larger complex POCs for customers to try in their environment. We're helping customers with, again, software support.

There's not a lot of software support that's needed on the CPU side, but there's some for people to get comfortable. And we've seen the adoption increase on the enterprise side. So if you talk about our market share being in, let's call it, low seconds revenue percentage, on the hyperscaler side, we're well above that. And on the enterprise side, we're well below that. And I think we have a lot of opportunity to continue to grow in enterprise.

TOSHIYA HARI: And there's really no fundamental reason why your enterprise share should be so much lower than hyperscale.

LISA SU: Not at all.

TOSHIYA HARI: From a technology perspective.

LISA SU: Yes. From a technology standpoint, I think we feel extremely good about our competitive positioning. And it is really about being a trusted supplier. One of the things that we find in the data center is customers want to know that they can count on you, count on your roadmap, count on your reliability, all of those things. And I think we've demonstrated that over the last few years.

TOSHIYA HARI: Many of your cloud customers have custom CPU and accelerator programs that are running. Some are way ahead, some are fairly nascent. How do you see the mix of merchant versus custom evolving over the long run, again, both on the CPU side and the accelerator side? And as a supplier of, for the most part, merchant silicon, how do you plan or how do you strategize competing with essentially some of your customers?

LISA SU: Yeah, I find this to be an interesting question because people are always wondering, well, is it going to be x or y. And, I say, look, it's going to be both. I mean, you're absolutely-- when I think about the investments that we're making in a competitive CPU and GPU roadmap, they're huge. And we're getting economies of scale over all of that investment in architecture, in software, in yields and reliability and all of those things.

And our largest hyperscaler customers want to leverage that scale. That's a good thing. And so we expect that our job is to continue to move, let's call it, the merchant roadmap as fast as possible to get all those efficiencies of TCO and new technology, new architectures going forward.

As expected, there should be custom silicon. I think custom silicon will come into play. It will typically come into play for, let's call it, less performance sensitive applications. So that's where you see sometimes, let's call it, good enough performance can be done in custom silicon or in areas, especially on the accelerator side, where it's a more narrow application. So if you don't need a lot of programmability, if you're not upgrading your models every 12 months, in that case, you may trend towards that.

But that being the case, when we think about, for example, our $400 billion accelerator TAM, we think the vast majority of that will remain GPUs. And then I also look at it as an opportunity to partner closer with our largest customers. I don't view it as competition. I really view it as partnership because we also have a semi-custom capability which allows-- if you look at what we've done, for example, in our game console business with Sony and Microsoft, what we say is, hey, come use our IP and figure out how you want to differentiate yourselves. And I believe that that's a very effective model when you get into a time frame when the models and the software are a bit more mature. And in which case, that would be an opportunity for us.

TOSHIYA HARI: So something like that, we might be able to see on the data center side?

LISA SU: I do believe so. Yes. Look, I think at the end of the day, we're all about how do we drive more value in our overall technology equation. And again, we have very deep partnerships with all of our IP investments. There are definitely ways that we can do even more together with our largest customers.

TOSHIYA HARI: On AI PCs, from a financial markets perspective, CES was very much sort of an AI PC fest, and then Computex was also another one. More recently, I think Sentiment on our side, if you will, has come down a little bit. What are your thoughts on AI PCs, what are you focused on as it pertains to Killer apps, and how would you characterize your competitive position in AI PCs vis a vis traditional PCs?

LISA SU: Yeah, I believe that we are at the start of a multiyear AI PC cycle. So again, you guys are always trying to go a little bit too fast. So we never said AI PC was a big 2024 phenomenon. AI PCs is a start in 2024. But more importantly, it's the most significant innovation that's come to the PC market in definitely the last 10 plus years. And I view it as a very, very natural thing.

If you're thinking about PCs as a productivity tool, you can definitely use AI. And in this case, we call it AI PCs have these NPUs that are in the silicon. You can definitely use this AI technology to make your PCs more useful. So why wouldn't people want to adopt AI PCs?

It is one of those things where you have to do a lot of hardware software co-optimization. We've done a tremendous amount of work with Microsoft on their Copilot plus initiative. They just announced last week at IFA that they will have, let's call it, x86 support for hours and other technologies later this year. We think this is the beginning of the AI PC cycle. So next year, as we think about commercial PCs and commercial refresh cycle, we actually see AI PCs as a driver of that commercial refresh cycle.

TOSHIYA HARI: And then from a competitive standpoint, I think, historically, you've been better positioned on the consumer side and maybe a little bit less so on the commercial side. Going forward with AI PCs, could that be sort of a catalyst for you to improve your position on the commercial side?

LISA SU: Yeah. Again, on the PC side, we have traditionally been underrepresented overall, but particularly in the commercial PC side. One of the things that as we have really focused on sort of future go to market, our investments in the enterprise and commercial go to market have increased quite a bit. I think we lead with server CPUs. Server CPUs, the value proposition is very, very strong for AMD.

And then we find that many of these enterprise customers are pulling us into their AI conversations because, frankly, enterprise customers want help. They want to know, hey, how should I think about this investment, should I be thinking about cloud or should I be thinking about on prem, or how do I think about AI PCs. And so we found ourselves now in a place of more like trusted advisor with some of these enterprise accounts.

And so I do believe that when you look at the overall choices that enterprise CIOs have to make between their traditional compute, what should they do, cloud versus on prem to their AI compute, how much is being done on CPUs versus how much is being done on GPUs, how much of that you have to worry about privacy and security and all of that stuff to AI PCs, when to adopt, I think all of those are part of a broader commercial go to market that I believe is a great opportunity for us. And frankly, it's an important opportunity for the industry because CIOs have more choices today than they've ever had. What they need is some help to go through all of that and figure out, where are the priorities for investment.

TOSHIYA HARI: Shifting gears a little bit. Your embedded business or primarily FPGA business is about 40%, 45% off the recent peak. You did, I believe, guide that business up going forward. What are you seeing from a customer order pattern, perspective. You service industrial, automotive, consumer, etc. Are there any applications or end markets that stand out from a demand standpoint?

LISA SU: Yeah. So again, the embedded business is a business we don't quite talk about as often as it relates to AMD. But it's a very, very good business for us. When we look at the diversity of customers and the diversity of applications, we continue to believe it's a strong pillar of our overall strategy. We are coming off the bottom. So the first quarter was the bottom for the embedded business after there was just a lot of inventory that was gathered at end customers.

We do see some improving order patterns, certainly in the second quarter and going into the second half of the year. It's probably a little bit more gradual than everyone would like. We do see some markets better than others. Like aerospace and defense, very strong test and measurement, sort of emulation related needs, strong industrial, a little bit slower in the overall recovery. But what I'm most excited about with the embedded business is we're starting to see some real synergies in our overall portfolio.

So if you think about it, our embedded customer set that based on FPGAs was like over 6,000 plus customers. Many of them had not really even understood the technology that AMD had. And what we're finding now is, especially in this world where I said, CIOs and CTOs are finding this really complex environment that they're dealing with, they actually don't want more and more suppliers. They actually want more partners that can help them navigate the overall roadmap.

And so we've seen very significant design win synergy between our embedded FPGA business and our embedded CPU business with design wins in the first half of the year being up sort of 40% year on year to over $7 billion in new design wins. And we see multiple customers saying, you know what, I want to standardize on AMD. I trust you guys. I trust that you'll be a good partner in all other respects. Now let's talk about how we move more and more of our portfolio.

TOSHIYA HARI: Got it. Coming back to AI. Just on how you think about the portfolio and potentially M&A going forward. You've had Xilinx, Pensando, multiple software assets, and now-- again, hasn't closed-- but ZT Systems. At this point, do you believe you have the portfolio and the right assets to be very competitive or are there still holes that you feel like you need to fill?

LISA SU: Yeah. So we've always thought about our portfolio and our capital allocation very strategically. So these are long term bets. From the standpoint of each of these acquisitions and our organic investments have been towards really positioning us to be a leader in high performance computing and AI. So I think with Xilinx, Pensando, our software acquisitions, and now with ZT Systems, we're extremely well-positioned.

And I'd like to say, well-positioned in the bigger I conversation, not just data center AI, but really end to end AI infrastructure across cloud, edge, and client. And I feel really good about our portfolio. So yes, we're in good shape.

TOSHIYA HARI: OK, great. The other question that we often get is on the supply chain and what's going on there. Nothing specific to AMD, but I think generally speaking, things like advanced packaging and high bandwidth memory have been fairly tight from a supply perspective. In '24, and going into '25, how supply constrained are you and your data center business? And I know this is a tough question, but at what point do you feel like supply could potentially catch up to demand? I know it's a moving target.

LISA SU: It is. Toshiya, as you said, it's a moving target. Look, I think as an industry, we've put a lot more supply capacity on board. So we've certainly ramped up our ability to service AI revenue in 2024. We will take another big step up in 2025. The constraints are, like you talked about, advanced packaging and some of the high bandwidth memory.

I think it continues to be tight, frankly, because although we're bringing overall capacity up in the industry, demand is also very strong. And then we find that with the new generations, die sizes are larger, the memory capacities are larger. And so all that says, we're still going to be in a relatively tight supply environment going into 2025.

TOSHIYA HARI: Got it. On supply and how you think about your manufacturing strategy, the other question we often get is how do you think about your foundry strategy going forward. You have a lot of concentration at TSMC and specifically in Taiwan. And this certainly isn't specific to AMD, but how do you think about plan B, if you will, if there is one, when you're thinking about 3, 4, or 5 years down the line?

LISA SU: Yeah, it's clear that we all have to think about resiliency in our supply chain. So COVID certainly taught us that. We continue to look at diversification of the supply chain. TSMC is a fantastic partner. I mean, they have been an excellent partner to us across all of the various aspects of technology and manufacturing.

We are big supporters of the CHIPS Act. We're happy that people are building in the US. We're happy that TSMC is building in Arizona. We're taping out products and ramping that. And we'll continue to look at how to de-risk the supply chain with the notion of this is an industry wide problem. And all of us are looking at how do we create just more geographic diversity.

TOSHIYA HARI: OK, great. In the last two minutes, just one last question. And how we should be thinking about OpEx leverage. Your investments in the near term versus generating profits and free cash flow, if you will, for investors. Obviously, you have a rich set of opportunities, as we've discussed.

You do have a lot of competition with very strong companies. You're a strong company as well. How do you think about that balance, investments versus showing returns, if you will, for the investor base?

LISA SU: Yeah. Look, capital allocation is incredibly important for us. And we do have many more opportunities than-- every year, we seem to get more. I think the key principle is we are investing in the business. I mean, this is an opportunity for us. I think this AI sort of technology arc is really a once in 50 years type thing. So we have to invest.

That being the case, we will be very disciplined in that investment. And so we expect to grow OpEx slower than we grow revenue. But we do see a huge opportunity in front of us.

TOSHIYA HARI: In the last minute or so then-- we have a little bit of time-- anything that perhaps we didn't touch in the session? Or as you have discussions with investors and analysts, as a collective unit, any aspects of AMD or your markets that we either overlook or underappreciate?

LISA SU: Yeah. I think the main thing is, look, this is a computing supercycle. So we should all recognize that. And there's no one player or one architecture that's going to take over. I think this is a case where having the right compute for the right workload and the right application is super important. And that's what we have been working on building over the last 5 plus years is to have the best CPU, GPU, FPGA, semi-custom capability such that we can be the best computing partner to the ecosystem.

TOSHIYA HARI: Great. Thank you so much for the time and hope to have you back next year.

LISA SU: Fantastic. Thank you.

TOSHIYA HARI: Thanks so much.

[APPLAUSE]