Please welcome Arm's Executive Vice President of Edge AI, Chris Bergey.

Chris Bergey - *Executive Vice President, Edge AI Business Unit, Arm Holdings plc*

All right. Thank you all for being here. Obviously, an exciting day for Arm.

I was asked just to give a little bit of my background. I see some old faces here in the audience. For those who don't know me, I'm actually going to celebrate my 30th year in the semiconductor industry this summer, after starting out of engineering school at AMD, then spending almost a decade at Broadcom, and actually joining Arm a little over six years ago. That was the first time I actually worked for an IP company versus a silicon company, and I did some startups in between there.

So for me, it's exciting to see Arm get into silicon, but I'm super excited about what I represent and what we're doing inside of the Edge Business Unit.

As you're probably aware, the Edge Business Unit actually represents a lot of the legacy Arm business. As large as it is, and as strong a footprint as we have, many of you may be surprised to see the amount of growth that we believe we're going to see over the next five years from a TAM perspective, and that's about a 40% increase in TAM.

Now, why is this happening? It's not a big surprise. AI. Essentially, what is happening is AI workloads are coming to all kinds of devices. Obviously, it's been pretty exciting to see the OpenClaw moment that Rene mentioned earlier, and this has been really fun for us because we've been talking about agentic AI for over a year now. It was really hard for people to necessarily connect and have that "aha" moment around why I would want a personal agent and why I would do these things.

Now you're starting to see these technologies come to market. But what you're seeing is that you start seeing the concepts, and that's great. But I think the other thing you see with many of these early agentic use cases, or even some of the prototyped agentic OS work being done on the development side, is that the concepts are great, but the performance is not there.

I can tell you that this has come through in spades for me. I actually just spent the last several weeks traveling between Mobile World Congress in Barcelona and Embedded World in Germany, and the number of customers that came to me and said, "Chris, you know what? We need to up our compute. We need to think about going to more advanced nodes than we were thinking about. I need to think about a more powerful CPU complex. I need to think about more memory bandwidth."

So, this is really what's going to drive much of this growth that we're expecting.

When you look at where the edge AI opportunity is for Arm and inside of my business unit, obviously mobile continues to be an incredible driver for us, and these devices continue to be essential to us. We believe that many of the agentic services that are going to start running on top of those devices are going to drive an incredible amount of additional silicon content.

One of the questions I get there is, "Wait a minute, Chris, what about the dynamics around pricing?" We're obviously seeing some of the tensions that are happening already today around memory and storage. But when you actually look at it economically, and you look at the cost of many of those tokens in the cloud versus doing them locally, there are quite a few economic drivers, quite a few performance drivers around latency and real-time response, and then there are also a lot of requirements around privacy and where that data stays, and how do you protect some of those things, and where the agents have access to that data.

So, we see a huge amount of growth opportunity in mobile, and I'm going to talk more about that as we get into our CSS strategy.

The second thing is intelligent edge. Obviously, this is a huge category of devices, but what you see across these devices - whether they're consumer devices, enterprise devices, or industrial devices - is AI coming into those devices, and people looking for new use cases and new capabilities.

One of the analogies I like to use is touchscreens. Remember, the iPhone is almost 20 years old, the iPad is about 15 years old, and that was when we really started using touch. Now if you give a child a screen and it doesn't have touch, they basically give it back to you and say, "I don't like this. It's not the experience I know."

What you're really going to see in all these devices is that's how essential AI is going to be. The AI experience, the device intelligence, being able to sense things you want to do, being able to have conversations with it - all those kinds of activities. So we see this across the board, both around existing categories of devices and then new devices that you can see around XR platforms and the like.

Lastly, we see a lot of exciting things going on around personal AI computing, and how AI changes the way we have these compute paradigms in our homes and in our lives.

The first thing to touch upon - and Rene obviously did a great job this morning talking about this - has been the journey from IP to CSS. This has been quite transformative for us.

The conversation that I would have with so many of our customers were: "Chris, we're not having the best performance in the market. If you could just help me get the best Geekbench score, if you could just help me get the best SPEC score, I could get more for my product, I could get more market share, I could do better, and I'd be willing to share some of that with you, Arm, because you're helping me be successful."

So, this was really the birth of CSS.

What we did for CSS inside of the Edge AI Business Unit is we looked at what it really takes to make a world-class product. Really, what you need to do is optimize all the way from the transistor, through the libraries, up through the RTL, up through the physical implementation, and then as well work very closely on the software stack.

This is what we have done with CSS. I think you heard Y.H. mention it in his video - the president of Samsung Semiconductor - around the amount of work that we do with leading companies like Samsung, like TSMC, and the like, to make sure we are optimizing that full stack.

So what we do is we give customers not just the RTL and the IP - the GPU IP, leading CPU IP, system IP - but we actually are able to stitch that together, and then we also give them the recipe. We say, here you go: this product on the latest node, at 4 GHz, gets you this score.

For some customers, that's amazing because they say, "You know what? I'm new at this. I'm just an OEM. I want to get into silicon. Give me that recipe, and I'm going to cut and paste that right into my chip."

Now we have more legacy customers that say, "Hey, we think we're really good at that too," and we say, "Great." But what we give them is a benchmark, because many times their engineers say, "We don't think 4 GHz is possible." We say, "Here it is." Then they get a chance to try to beat it and try to do 4.1, 4.2 GHz.

This has become a very virtuous cycle for us in really making sure that our partners are able to do very well in the markets that they participate in.

You can see this in our results. At this point in time, as of the base of the quarter that's ending this month, we have now achieved the point where 25% of our mobile royalties will be coming from CSS. But we're not stopping there. We actually have a tremendous amount of growth, as you can see from the slides, that we believe we're going to be able to pick up as we're really able to enable even more partners to do even better in their markets.

Obviously, this agentic drive for more performance across all price bands becomes very, very interesting to us.

But the opportunity goes beyond mobile.

A couple of segments that I like to talk about: one is kind of the emerging XR-type segment. I don't know about you guys, but obviously I like Meta on the data center side - it was great to see Santosh this morning - but I think their glasses are awesome, and I think I own three pairs at this point in time, so I definitely am a full-time user.

It gives you an idea of next-generation platforms, and all of those glasses are Arm-powered. In fact, the wristband shown there is actually a neural wristband that you can use - when they have the display glasses, you can start turning up and down the volume, doing those kinds of activities - and that also is Arm-powered, based on our Ethos NPU and MCU.

Again, we're really able to help enable these next-generation platforms that are starting to use AI-type features in all kinds of different power envelopes, because that is one of the most difficult computing platforms right now because of battery life, sensitivity to weight on your face, temperature concerns, and all those kinds of things.

But beyond that type of platform, which to me is also driving a whole rebirth in wearables, we're also seeing a tremendous opportunity around personal AI computing.

One of the products I'd like to point out is another CSS product: the GB10 coming from NVIDIA. What you're seeing here now is this new category of devices where people are saying, "Hey, I want to run agentic AI." You can see what we're talking about from an OpenClaw point of view. You see great YouTube videos and

examples as people are playing with these things. We very much believe that there's going to be a tremendous amount of computing platform opportunity here.

In fact, some of the products in this category are sold out - you can't get them.

What are the attributes of these products? Well, they have a tremendous amount of Armv9-powered Arm CPUs. Two, they've got great GPU subsystems, whether that comes from Arm or whether that comes from other partners' GPUs. And third, they have an amazing UMA memory system that's able to provide the bandwidth those AI platforms need.

So, we believe this is going to drive a huge growth platform.

I just mentioned I came back from China. It is OpenClaw crazy there right now around the number of platforms and what people are trying to play with, and the future types of products that they want to build.

What I think is also exciting is how this is translating into what you see happening in the PC industry. What I really see happening there is a really exciting transition happening. Obviously, Arm has made a big push around Windows on Arm. We've made some nice gains over the last several years. But what we're seeing now is something totally different.

At the premium level, you're seeing people thinking about AI workloads in the personal computing space. It could be running agents, but it's also about the next-generation creator. When I think about PC creating, it always was about publishing websites, those kinds of devices. But now, that's not a creator. A creator is creating videos - AI-generated videos. They're using Stable Diffusion to create new graphics capabilities. Those kinds of platforms, again, are a tremendous strength for Arm's partners, and we believe that Arm will be able to participate in a very meaningful way at that premium part of personal AI computing.

What's also happening, interestingly, is at the opposite end of the market, which is really focused on the more traditional client use cases, you're starting to see all of these efficient computing platforms come out, whether it's Google thinking about how they're going to merge Android and Chrome together for Aluminium, or potentially Microsoft and how they're doing 365 Link, and then lastly you have Apple coming in with their neo platforms.

With the tightness around memory and storage, I really believe you're also going to see this more efficient category really start to eat up a significant amount of PC market share. That's a great story for Arm and our partners because they participate in many of those markets today. We have amazing power efficiency, we can get amazing memory efficiency, and not to mention our GPU footprint.

I think many of you know, but maybe it's a secret to many, that Arm is the highest-volume GPU shipper in the world. In fact, we've shipped over 12 billion chips that utilize our GPUs. We're also making a significant amount of AI investment in this space around our GPU platform, so expect to see more from us and our partners in that area.

So, when you think about the CSS opportunity in this space, this is very different than mobile. This is not a space that we have a significant amount of market share in today. This is really a growth opportunity for us. You're seeing the projections that we believe, and this will be a significant driver for us as we bring those CSS platforms into many of these other computing areas. So this is a key strategic area for us, and we're getting tremendous leverage from what we're doing already on the mobile and GPU side.

Lastly, I want to touch upon just a little bit about AI software. Sharbani is going to talk more about the importance of software. Rene obviously touched upon it earlier today, but I wanted to just nail this point: Armv9 is absolutely the most secure and advanced AI CPU architecture out there.

Richard Grisenthwaite was up on the screen - he is our Chief Architect - and what he and the team have put together with our partners is the most advanced platform. SME2, our Scalable Matrix Extension 2, is now shipping in the leading handsets, both from iOS as well as Android. What you're going to see is those platforms proliferate.

At this point in time, we're approaching about 50% value share on Armv9 penetration. That's going to actually go up to 85% in just the next two years. So you're seeing Armv9 go everywhere, and AI is going to enable us to take that into other markets as well because of the AI footprint that we have here.

Again, you can see it's the leading OSes, the leading applications and tools, and then lastly the leading AI frameworks.

Many of you know that many of the really tough challenges around AI are around software frameworks. How do you make it easy for users to migrate their workloads, for them to reduce the quantization challenges, and for it just to be easy to run AI?

A year and a half ago, or almost two years ago, we introduced KleidiAI at Computex, and it has been transformational for us because KleidiAI is a library that allows developers to integrate into their platform such that when their AI workload lands on an Arm CPU, it is able to use the advanced instructions that are possible in that platform.

This is what has been integrated now in these frameworks that you see below. But I'll let Sharbani tell more about that to all of you later on today.

With that, I'm going to hand it over to Drew Henry.