Arm Everywhere event Media Q&A full transcript

March 24, 2026

Question: Hi I'm Peilin from China. So my question is about competition and cost. You used to only do IP, and now you are doing products, you are now competing with some of your customers. So how do you think about the competition? Also, can you provide any projection of the Chinese market? The CPU products sold into the Chinese market.

Rene: I could take that. Oh, I have a mic. Sorry. Technology is amazing. So if you guys can hear the question, his first question was, you guys were primarily an IP company, and now you're doing chips, are you competing with your customers and then a question about China. I think Paul (Saab of Meta) I think Paul said it best when he started talking to us, saying that the market was underserved. It is a massive, massive market. We think there's room for lots of different players, so we're not really too worried about that. I think also you saw the commentary from our leading partners, Amazon, Google, Microsoft, all with in-house efforts and Jensen as well. So we just view the market opportunity as very very large and quite underserved at the moment. To your question on China, no China customers to announced today, but we think China can be a very good market for us.

Question: Jim McGregor, TIRIAS Research, hi guys, good to see you. Quick question for you. First off, glad to see you, but since you're targeting AI applications and accelerator, can you talk a little bit about your decision to go with the PCIe Gen 6, and did you consider possibly licensing NVLink since we're seeing some uptick with NVLink Fusion?

Mohamed: So two things, so we chose PCI Gen 6 because that is the furthest along in terms of speed, et cetera, in terms of standard IP implementations when we began, certainly. In terms of NVLink and as we look forward, one of the things that we're absolutely looking at is entire system performance and ways to kind of optimize how it engages in the system. So NVLink is obviously one of the options, NVIDIA is a great partner for us, and so we're certainly looking at that for future revisions. In fact, we already have announced that we will be supporting NV link in a future revision of our CSS. And as I said earlier, the AGI CPUs are based on the CSS so there's a natural extension there if we decide we want to head in that direction.

Question: Great presentation, Ben Pouladian with BEP Research. So I guess most of the market, when they're focusing on CPUs, are focusing on AMD and Intel, but it seems that Masa working with you guys at SoftBank and Arm did a great job. You guys were able to make this big bet into agentic CPU and to me it looks like you have a lead in the market

versus AMD and Intel on any type of agentic processor that they have. How long as this been going in development and can you tell us a little bit more about the genesis of how you cam about this story, it's pretty interesting.

Rene: I alluded to it during the keynote, CSS has been a bit of magic for us, because what CSS has done is it's pulled forward the ability to do these complex chips a heck of a lot faster. So we were having lots of discussions with customers about CSSs, and that moved pretty rapidly into, "Hey, couldn't you take a CSS and do something for us?" So I think the primary thing to start with was we were asked to do. That's kind of number zero. But now what we're seeing and you can see based upon the huge amount of customer attraction, we think the opportunity is pretty large.

Question: I was curious about the name AGI, calling this an AGI CPU. How much should we be reading into that? How much have you been thinking about a chip design for AGI? Is that actually the way you thought about it? What does that mean in practice? What does it say about timing of AGI?

Rene: Do you like the name?

Question: I mean, yeah, it's a good name. I'm just wondering is it more than a name?

Rene: When we thought about it, you know obviously, branding and naming is a very passionate topic that one would talk about internally and then ultimately externally. And there's a lot of opinions on it. When we think about the trajectory of where technology is going, two things kind of came to mind. One is CPUs will be everywhere no matter whether it's a very small use case or very large use case in terms of power. And secondly, there's going to be an AI component to it. So we just felt that intersection of AI everywhere and CPUs being table stakes is just felt right as far as the naming goes. We thought we would make it very clear.

Question: How do you want the industry to interact with Arm's first chip? Is it another example of Arm innovating with a specific particular vertical or a disruption of the entire business model of computing infrastructure?

Rene: I think your question was about how should we think about the disruption of the business model?

Mohamed: Her question was is it a business model disruption or is it a continuation of innovation in a particular vertical. I'll take a stab at it. The way to think about it is an extension of what we've always done. We look to meet our customers where they are. That's always been Arm's business model, which is, how do we develop some technology and allow it to get amortized across lots of different customers? When we saw the

opportunity, we had multiple customers coming to us and asking us to go off and take things a little bit further. It was a natural sort of extension of the business model we've always employed.

Rene: I think maybe the way to think about it is to go back to CSS. When we started talking to folks about CSS, and particularly, I think I hearken back to investors two and a half years ago, the CSS traction we had there that time was with Neoverse, but we were planning on expanding it out to automotive and mobile. One of the questions we got was, aren't you competing with customers? And aren't people going to hate it? And when we fast forward now, we have customers for CSS in all those markets, which kind of comes down to, well why are they choosing it? It's a better choice. Are we going to ask everyone to buy Arm AGI chips? We may ask, but if they don't, that's okay. If they take CSS, that's fine. If they take IP, that's fine. To Mohamed's point, it's just an extension of the offering.

Question: Hey, this is Stephen Nellis with Reuters. Thanks for doing this and taking the time. I have two two-part questions, so evade as necessary. One is, when you showed the die shot, it kind of looks like the two dies are different. And I'm wondering if you can talk about that and if it have any effect on yields? The number two is just, if you could talk about the processing technology or packaging technology that will go either into the 26 chip or the 27 chip that you showed on the roadmap slide:

Mohamed: Yeah so they're not different, it's the same die. I don't know what you're specifically pointing at but it's the same die. And to your second question, we're not going to talk about the next chip if you will. But the current chip is a sort of standard package. There's nothing especially exotic about it. And so again, we were looking to TSMC, three nanometer standard substrate, off you go.

Question: So the current chip was not on 3nm?

Mohamed: It is on 3nm, the current AGI CPU is a 3nm product.

Question: Hello this is Nitin Dahad with EE Times. I'm curious about, you talked about a trillion dollar market opportunity. I'm not sure whether that was total for Arm or for the AGI chip series, but can you just sort of elaborate on how that market will be addressed by you? I mean currently a small volume of customers, the hyperscalers, where do you see the market opportunity? And that leads to a question of pricing and distribution channel. I'm not sure those are addressed. I'm not sure whether you have had those yet or whether you're saying it publicly.

Rene: So what we were trying to, and thanks for the question Nitin. What I was trying to illustrate on that slide is there's a lot of market opportunity beyond just the Arm AGI CPU. When you think about the size of the semiconductor market, in 2030, $1 trillion is not a

arm

stretch. Won't go any further in terms of how that's sliced and diced, but it's an opportunity that we think we have rights to.

Question: Hi guys. Ian king here from Bloomberg, and thanks for this opportunity. I wonder if you can just talk a little bit about how it's going to work with your customers. I mean, we saw endorsements from people in that video who are actually making their own CPUs and committed to CPU programs which would seem to have significant overlap with the product that you're also going to try and sell them. So can you just unpack a little bit about more about the interaction?

Mohamed: Yeah, I think the way to think about this, right? And it's actually sort of a unique strength of Arm, which is, we can walk into one of these customers and say, "Hey, here's a portfolio of products. We can give you IP, we can give you CSS, we can give you AGI CPU," sort of full range. And so one of those customers may decide for certain use cases they're going to use IP. In other cases, they want a CSS. In other cases, they want to take an AGI CPU off the shelf or a complete finished piece of silicon. And guess what? They get software leverage across all those products. And guess what? Now they can customize their entire data center regardless. And so, what it really gives them at the end of the day is optionality. So do we want all those customers, to Rene's point, buy AGI CPU? Absolutely, right? But the reality is, many of them are going to continue down the path of doing what they're doing and possibly using it for certain areas or not, and that's okay. In fact, we think that's a strength, and strengthens the overall Arm ecosystem, and, frankly, strengthens our presence with those customers in the long run.

Rene: It actually expands the market opportunity, Ian. If you think about a company like SAP, SAP is going to be building a system based on the Arm AGI CPU. They may not be able to serve all the workloads on their on prem. They may want to go to AWS. That opens up an AWS opportunity that may not have existed before. So it all comes down to the software ecosystem to Mohamed's point. The richer and broader and deeper the software ecosystem becomes for Arm, if you're building your own chips on Arm, it's actually beneficial. It broadens the market opportunity.

Mohamed: Sorry, just one more highlight. That point that Renee made, that is an incredible one. You heard me say there are more than 10,000 customers that are using arm in the cloud today. They don't have access to something that looks like that for on-prem. And this is really, Arm AGI CPU gives them access to that exact, you know, something that looks very similar to that based on the same IP base I think you use on-prem that creates upside for the hyperscalers. It creates upside for them for the customer.

arm

Question: We have a question coming in online from Tobias Mann of The Register. He says, "You now have a something unique for agentic compute. Does Arm have any plans to bring Arm-branded silicon to other market segments, and if so, where next?

Rene: So all we're talking about today is the Arm AGI CPU, which was plenty, we took two hours. So nothing more to talk about.

Question: George Cozma from Chips and Cheese. On the product brief, I notice that you call it CXL 3.0 but only type three devices. Are you also supporting type one and type two devices as well or just type three?

Mohamed: Type one, two, and three.

Question: And second question, I saw that the TDP goes up to 420 watts on that product brief. Why the disparity between the slide saying 300 watts?

Mohamed: That could be a typo, I'm not sure. It certainly doesn't go that high, so I'm not sure. But we'll go take a look at that and come back to you.

Question: Neil Shah from Counterpoint Research. Congratulations, First of all, on the AGI announcement. My question is about talking about software ecosystem and the developer strategy. What is the developer strategy? We didn't hear much about that. So if some developers were developing already for Arm Neoverse or AWS, can they easily develop for the same?

Mohamed: It just works. I mean, it's the same it's the same micro architecture, it's the same CPU Code, it's is the same interconnect. This is actually, by the way, I'm going to riff for a minute. This is actually a big misnomer in the industry. So for all those partners out there that are deploying Arm Neoverse in the cloud, whether you're Google or Microsoft or AWS, the software that's tuned for one of those just works on the other one. That's very different than the x86 ecosystem, where you have two completely different implementations, right? And so there's a big software lift. We don't actually have that problem. And in this case, AGI CPU just falls into that same sort of...In fact, a lot of the early porting work that that folks like Meta have done, and even us getting ready for the silicone to come back, we did in the cloud.

Question: And if someone buys a chip, I guess the IP is paid for? They don't have to pay IP again.

Mohamed: If they want to pay us twice, we'll take it.

Question: Hi, Vlad Galabrov from Omdia. I have a question. So today, the players use the only one that has SMT. And I'm curious about your thoughts through the design process of

should you have that you know, what the benefit was, the drawback? How did you decide not to do that? Because I'm guessing it's not something that AGI will have.

Mohamed: I'm so glad you asked that question, because I was just commenting, you all noticed that my slides got a little messed up around that point, and I was going to try to make that point during the discussion. So here's the deal with SMT, right? SMT makes a lot of sense if you got, if you need very little bandwidth, right? Because you're going to split the bandwidth, and you're effectively, you know, waiting around a lot, right? Because each process, as it waits around, can sort of yield the CPU resources to the other processors, so there are use cases where that's going to make sense. What we see in an agentic AGI type use case is where you've got, you know, a ton of threads you're trying to support. You've got expensive accelerators and expensive, you know, infrastructure sit around and waiting. The last thing you want to do is, you know, split your IO bandwidth, or split your memory bandwidth, right? You want to tailor that IO and memory bandwidth to the process right? And we think that that number is about four gigabytes per second. Four to six gigabytes per second is sort of the sweet spot. And so you know, if you go implement SMT in some of those use cases, it might make sense. For us, not so much. And so right now, we have no intentions to kind of head down that path.

Question: Hi guys, Katie, CNBC. Was excited to see Open AI on stage. I'm wondering if you have anything to share about that deal we might see coming or when that might be, and along those lines, any additional details about the Meta deal, financials, number of chips, number of years?

Rene: Yeah, I think we're going to be doing investor session, Katie, so we're going hold off when this session is over to talk anything about financials. Open AI is an amazing partner. Having Kevin on stage was great. Then you heard Santosh particularly talk about use of the Arm AGI CPU as a head node and Mohamed also alluded to that in his presentation. So Open AI has been an amazing partner.

Question: Stephen Nellis with Reuters. I just wanted to follow up on what the interplay is going to be like between when IP is available to Arm's internal silicon design teams and when it's available to external customers. Is everything going to be available to those design teams at the same time? Also, you know, what kind of sharing, if anyone, will be as that implementation happens because, you know, we've got IP, but we've got implementation, and there's a lot in there. So for example, if they are available at the same time to internal and external teams, if you find out something in the implementation internally, how is that going to work? Does it get pushed back up to the IP team and it sends over to the external customers. But just wondering if you could get into the timing and

arm

footing of how these design teams will be able to access IP like the CSS IP and some of those other really important IP blocks that you'll be selling.

Rene: Yeah, so we have long term agreements with many of our IP partners, and in those agreements are commitments and understanding to deliveries on schedule. We have a very detailed schedule for ip implementation in terms of alpha, beta drops, LAC, EAC, which are basically production drops. We can't deviate from those, and our internal team certainly don't want to take anything that's not ready, because we're not going to be able to do anything with it. So nobody is going to get anything any earlier internally than folks get externally. There's going to be no issue with that.

Question: Leonard Lee at Next Curve. Congratulations, really exciting announcement. My question is regarding your rack business. You know, Rene, you stopped at the chip, but Mohamed, you talked a lot about the rack. So I don't know if I, you know, this question has already been asked, but because I was out getting your delicious lunch here. But yeah, what's the deal? What's the go to market?

Mohamed: Well, I'm glad you're enjoying the lunch. Now, in all seriousness, so, so here's the deal, right? We're not selling racks persay. We've got a bunch of ecosystem partners. We talked about some of them Quanta, Lenovo, Supermicro, others that are building racks and what have you. The point that I was making with the racks is that, you know, you can't think about an AGI CPU, a CPU designed for an AI data center, in isolation. You need to think about it in the context of these massive systems. And that's really where it comes back to those three things I talked about, performance, scale and efficiency. And by the way, performance, scale and efficiency, interesting, nobody buys them based on the chip. They buy them on the data center is going to go in, and what sort of impact is at that level. And so what I was showing you was that, hey, we've just got off the shelf rack, an OCP off the shelf rack, and we have partners who have done that, we've been able to sort of breakthrough performance, efficiency and scale that you just can't do an existing things because of because of the nature of the design.

If those partners, and we're working with them, want to go out and optimize that rack even further, they could probably do even better. And the best example of that is the liquid cooled rack, right? I told you, it's a 200 kilowatt rack. I got 45,000 CPU cores in there. But I'm only consuming 100 kilowatts per pound. So theoretically, I could do a lot more in that rack, but there's physical constraints associated with it. So it was more about showing that, hey, we're just at the beginning of this thing. The sky's the limit and it's been optimized for it.

Rene: That's a great question. One of the things that is sometimes misunderstood about our business is that we develop a piece of IP, we license it, and then we're off and to the

next customer. We work very deeply with customers who build chips, who build systems, who build reference boards, so we have a fairly deep understanding of what it takes to get the product out to market. When we thought about this particular product, to Mohamed's point, there's no way you do this without engagement of an ODM community, because people don't buy chips. They buy buy racks, but we're not selling them, but we have a very deep understanding about how the value chain works.

Rene: The question was we could provide professional services, design services, et cetera et cetera. Yes we could, but we're not talking about that today.

Mohamed: But we certainly work closely with the ODMs.

Question: I'm Jerry from Commercial Times. I have a question about some of Arm's previous partners who have built Neoverse chips, or used Arm Total Design, like Mediatek and Novatek. Will these partners also become competitors in the future?

Mohamed: I don't think so. I mean, the way that we think about this again, this again, it's a sort of trillion dollar market. There's lots of opportunities for customization. There's lots of opportunities to go off and expand. Again, the strength of Arm is really the breadth of the ecosystem. If you look at those partners, many of them are building chiplets, and they're type of coupling over partner accelerators. There's different core count configurations that go off and target different markets. We're building 128 core device, super high performance designed for the core of the data center. There's maybe smaller core count devices that are needed for things like DPUs or for edge platforms, etc. So it's a big market, and I don't think there's any overlap right now.

Question: Hi, Anand with TechInsights. I'm curious to understand your vision, how this rack will work with the rack of GPUs. You know, you would want a CPU sitting on the same board as GPU, for good performance and agentic AI, but if you have a rack of CPU and a rack of GPU, how would that work?

Mohamed: So I think there's a couple of things there. So there's the case of, you know, you've got, there's multiple use cases here, and I just want, let me just break it down real quick, right? So the simplest use case is a use case you're talking about, where the CPU is acting as a sort of head node of sorts. And that's a great example what we're doing with SK Telecom and Rebellions. So, you know, Rebellions is using Arm AGI CPU as a head node, and then there are a bunch of accelerators sitting in the same server. There's an example of it on the floor, you should go take a look at it. So that's one, and Positron is another example of that. That's one use case. There's a second use case when you talk about a very dense rack, right? And this dense rack is handling all of the agentic use cases around orchestrating and managing all those agentic flows, taking, you know, deciding which

model to go call, or going off and probing applications and tools, databases, et cetera, et cetera. And in that case, you actually do want a large rack sitting next to your accelerator. They can very, you know, it's very low latency they can go very quickly go off and dispatch jobs and kind of organize through that.

And then the third use case that I would suggest to you, where we're seeing a lot of traction, is really about the sort of control plane of the data center, if you will. And so this is like, you know, as these agents, you know, to Rene's point, they don't sleep, right? They're constantly running. They're going to generate a ton of traffic, ton of traffic to applications, applications to networking devices, and all of that CPU usage is going up. And those are all opportunities as well.

Question: One more from Tobias at the Register. "Mohamed, your argument against SMT makes sense in this context, but it also strikes me as an argument for Intel Sierra Forest and Clearwater Forest parts which also lack SMT. What do you see as the AGI CPU's differentiation against something like an Intel e-core part?

Mohamed: Go look at the memory bandwidth on those parts. I think, I think what you'll find is that what the focus was on those parts was really about jamming as many cores into the device as possible, rather than thinking through the overall architecture of the system. It may have their place, but I don't think that this is necessarily the same sort of category of product. And the other thing that I would add is also the sort of performance of those of those cores and those CPUs.

Question: Stephen Sopko, from Hyperframe Research. You've been on a journey, for the last couple of years doing this, and you've had to reintroduce Arm to every single one of your suppliers, partners, customers. Other than, "can we please have it faster," what's been the biggest challenge you face in that reintroducing yourself, reintroducing the company, and adding this product?

Rene: I think what we've seen is an acceleration of just demand for compute at a baseline level. When you look at Mohammed's chart of 1.2 billion Neoverse cores, everything has been happening faster. There is a huge lift. You heard it from Kevin, you heard it from Santosh, you heard it from Paul. And when there's a huge demand for compute, and most of the compute the world is serviced by Arm, it's a huge tailwind for us. So that has really been probably the chief thing over the last few years.

arm