Arm Business Briefing Q&A

March 30, 2026

Presenter: Ian Thornton, VP Investor Relations, Arm Holdings plc

Q&A

Speaker 1

Q1:

I have two questions. The first is about how to think about SoftBank Group's semiconductor business as a whole. In addition to your company, there are Ampere and Graphcore within SoftBank Group. Given that your company has now entered the chip business, how should we view the overall semiconductor strategy of SoftBank Group? Should we understand this as creating greater synergies within the group, or rather as each company pursuing different directions and targeting different layers independently?

A1:

Within SoftBank Group's AI Computing Segment, there are Ampere, Graphcore, and Arm. However, the chip that we've just announced is independent of anything being worked on by SoftBank Group and its semiconductor companies, including Graphcore and Ampere. This is entirely an Arm initiative, and there is no link to what SoftBank Group is doing. Obviously, there is a lot of work being done. I think Graphcore had about 500 engineers, and I believe they are expanding their capacity. Ampere had about 1,000 engineers, and they may also be expanding their capacity as well. But those engineers are working on something completely different. As to when that will be announced, and what they're working on, that's probably a question for SoftBank Group. But for now, the only thing we can talk about is the CPU chip that Arm has just announced.

Q2:

My second question is regarding the CPU chip business expansion. It appears that your plan anticipates a significant increase in revenue around FYE2030–2031. Should we interpret this as the point when inference demand, which you mentioned earlier, begins to scale in a meaningful way? Or is this growth driven more by company-specific factors, such as securing customers or achieving mass production?

A2:

It's a combination of both of those. By the time we get to FYE31, we expect that our third chip will be ramping into high volume. So, the first bumps are the first chip, and then as it starts to ramp, the second chip starts to ramp, and then by the time you get to the final bar in FYE31,

you now have three chips from Arm all ramping. That is also combined with what we believe will be the expansion of inference within the large data centers, and particularly the opportunity from agentic AI. It is a combination of both multiple chips from Arm, and an expansion of the opportunity with more agentic AI.

Speaker 2

Q1:

Does launching the AGI CPU chip mean you are now competing with your big customers like AWS (Graviton) or Google (Axion)? The beauty of Arm's business model is neutrality, so I'm a little worried about that.

A1:

AWS is building chips to be deployed only in AWS's data center. Microsoft is building chips to go into Azure, and Google is building chips going into GCP. None of those companies are selling their chips to anybody else. It is only for their own internal use. Maybe NVIDIA is selling chips in this area, but even NVIDIA seems to be wanting to sell systems, not individual chips. So if you're Meta, Cloudflare or SAP, you have no choice. There is nobody to sell you a chip. And so, we've kind of found a market that is very underserved. We think one of the reasons why it is so underserved is that these chips contain largely Arm IP. So if you are a chip developer, there is very little opportunity for you to add value or to differentiate, so it has not been an attractive market for someone to go after. But for Arm, because it's already all based on Arm technology, it is a very easy and attractive market for us to address.

Q2:

My second question is about margins. It may be a bit early to discuss, but generally speaking, hardware sales tend to have lower margins than IP licensing. How should we think about the impact of this on your long-term profit targets and future margin mix?

A2:

You are correct. As we indicated on one of the earlier slides, the IP business in this timeframe will probably have about a 99% non-GAAP gross profit margin and a greater than 65% non-GAAP operating margin. The chip business, we think, will have in this timeframe a better-than-50% non-GAAP gross profit margin and a non-GAAP operating margin of 30%. So, you are absolutely correct. The gross profit margin is lower. But what we are expecting is that we will provide information so that you can do a sum-of-the-parts calculation on valuing the IP business as you do today, and also then the new chip business. The other thing I would point out is that today we have an IP revenue stream, we have the costs of an IP business, and we have the costs of a chip business already in the company today, and we are still delivering around a 40% non-GAAP operating margin. So any additional revenues from the chip business, even though it will have lower gross margins, most of those gross profit dollars should flow through to operating

profit. To a certain extent, we already have the benefit of having all the costs associated with the chip business in the business today. Looking ahead, as I mentioned, we expect cost growth in the mid-teens. Excluding inflation and some additional investment in the chip business, the majority of the remaining costs will be directed toward new technology areas that we are not discussing today. If these investments are successful, they should drive additional revenue; if not, margins could be even higher.

Speaker 3

Q1:

First, I have a fundamental question regarding the Arm AGI CPU chip. Compared to other Arm-based CPUs, such as AWS's Graviton, Google's Axion, or NVIDIA's Vera, how does the performance of this AGI CPU chip stack up? I understand that performance is more than twice compared to an x86 based chip, but how does it compare against CPUs based on the same Arm architecture? Additionally, can this AGI CPU chip truly work with any AI accelerator? For example, would it be compatible with Google's TPU, Amazon's Trainium, OpenAI's in-house accelerator, or Cerebras' accelerator? I would appreciate some technical clarification on these points.

A1:

Looking across the different chips that are available and how they fit with the accelerators from the cloud companies, most of them have been highly optimized to work in their systems. For example, Google's Axion chip has been targeted to work very well with the TPU. NVIDIA's Vera chip is being optimized to run with Rubin, which is the GPU. Although you could put an Arm AGI CPU in there instead of an Axion, it may not perform as well, because the Axion has been optimized to work with the TPU. But if you were to take an Axion and have it working maybe standalone, then we think we would be very competitive. So, if you are Meta, you will have a competitive product to Axion, or to Graviton, or to Cobalt. So, we think it's very competitive when not being paired with their own proprietary accelerator. Standalone, or with somebody else's accelerator, we think it would be very similar in terms of capability. But the key thing is that for most companies, they don't have access to these technologies, so just being able to provide them with the equivalent is very powerful for them.

Q2:

In that case, is it correct to understand that this would be best suited for players that do not have their own CPUs but do have their own AI accelerators, such as Meta, OpenAI, and Cerebras?

A2:

We worked very closely with Meta on the design of the AGI CPU, with Meta acting as a lead partner to help fine-tune the architecture to meet their specific requirements. At the same

time, we were careful not to develop a solution tailored to a single customer. Our objective was to create a product that could be broadly deployed across the market. In that sense, we were pursuing two goals simultaneously, and we believe we have delivered a solution that performs well both as a companion to Meta's accelerator and as a general-purpose CPU. It is designed to support a wide range of use cases, including general cloud workloads, AI workloads running natively in the cloud, and deployments alongside various accelerator chips.

Q3:

My final question is about guidance for future licensing revenue. Earlier, you mentioned three drivers of growth. The AI technology cycle, next-generation technologies such as CSS which increase pricing, and licensing to SoftBank Group. I believe you indicated that this would grow at a high single-digit rate. Should we expect this to continue, or is it more temporary in nature?

A3:

We have not yet seen any slowdown in the AI technology cycle. We are seeing many companies wanting accesses to the next generation of Arm technology because they need to run more advanced workloads in their next generation of chips. This trend is evident across multiple end markets, including data centers, automotive, robotics, and smartphones. Across virtually every segment, customers are embedding more advanced capabilities into their designs, with agentic AI expected to be a key driver. For example, we are already seeing local agent use cases—such as OpenClaw running on devices like a Mac mini or Raspberry Pi—demonstrating that AI workloads can increasingly run at the edge, not just in the cloud. We expect this to expand across a wide range of end products and consumer electronics, further supporting licensing growth. In addition, we have currently signed approximately 30 of our top 50 customers to subscription licenses, and we expect broader adoption over time, which should drive incremental license revenue. CSS has also only been adopted by a limited number of customers to date, with significant runway for further uptake. Finally, we have good visibility into SoftBank Group's roadmap and plans. As a result, we expect licensing and design services related to SoftBank Group to not only continue, but to grow over the next several years. Taken together, these three drivers are expected to support sustained growth in licensing revenue over multiple years. We have greater confidence in the growth outlook over the next few years. Visibility into the outer years is more limited, which is typical.

Q4:

Just to confirm, I had assumed that for SoftBank Group, once a project is completed, royalties would continue, but licensing revenue will not grow. Should we now understand that these licensing revenues will actually continue to increase?

A4:

Most semiconductor companies in general don't build one chip. They build a series of chips.

They have a product portfolio that will continue for many years. They plan for success, and so there will usually be a sequence. As we showed earlier, we are also building a product roadmap, starting with the Arm AGI CPU, followed by AGI CPU 2 and AGI CPU 3. Assuming SoftBank Group has a roadmap, then they will need more technologies in the future. So, we are anticipating that there will be multiple projects, not just one.

Speaker 4

Q1:

My first question relates to the diagram you showed earlier on AI architecture and computational algorithms. In that context, software becomes critically important. In the smartphone space, the ecosystem ultimately converged around two platforms—iOS and Android—which contributed to Arm's dominant position. In AI, do you see the central software layer being driven by platforms such as CUDA or Meta's PyTorch, or is there potential for a third major ecosystem to emerge? From your perspective, who are likely to be the key software ecosystem partners going forward?

A1:

I think you already touched on a few key examples. NVIDIA's CUDA, for instance, runs partly on the Arm CPU and partly on the GPU. Meta, with PyTorch and ExecuTorch, is also an important partner. We are seeing strong ecosystem momentum more broadly. Google recently announced, alongside the launch of Axion 2, that it has ported 30,000 workloads to run on Arm. Microsoft has also highlighted the migration of applications such as Teams to Arm. In the cloud, essentially all major software stacks now run on Arm. Amazon, for example, recently noted that over 99% of its top 1,000 customers are already running at least part of their workloads on Arm. As a result, there is now very little commercially relevant software that is not available on Arm-based chips.

Q2:

My understanding is that there is a large software ecosystem centered around NVIDIA—namely CUDA—and alongside that, there are cloud service orchestrators, largely OSS-based, running on containerized or Kubernetes-based environments. From your perspective, when it comes to running AI today, should we think of the key software stack as consisting of, for example, CUDA on one hand, and then OSS driven by cloud service providers on the other? Or do you expect a broader range of operating systems, or rather kernel-level platforms, to emerge going forward?

A2:

All of the software that runs in the cloud is now already running on Arm, so I think there is nothing that's missing. You mentioned containers, so yes, you can run Docker or Kubernetes on Arm. Python and Perl interpreters are optimized and running on Arm. So, it's all of it.

Q3:

My second question relates to the assumptions behind the 1GW figure and the $15 billion target for FYE31. I would appreciate a rough sense of how you are thinking about FYE31. For example, at around 1GW, is it reasonable to assume that each chip might have roughly 200–300 cores by that time? If we translate that into CPU units, say with 200 cores per chip, 120 million cores would imply roughly 0.5 million CPUs required for a 1GW data center. It would be helpful to understand whether this 1GW assumption and the $15 billion figure for FYE31 are aligned with your internal assumptions. The intent behind my question is to better understand what level of CPU shipments you are implicitly assuming when you refer to the $15 billion opportunity in FYE31.

A3:

I wondered if your question was leading to "how many chips". We're not disclosing the number of chips, mainly because per-chip pricing is very much related to volume. And we see multiple ways of getting to $15 billion. It could be just through a small number of large customers, or it could be through many smaller customers. So, we've decided not to disclose chip pricing and chip volumes, in the same way that most semiconductor companies also don't disclose chip volume and chip pricing. We'll keep that secret.

Q4:

Conversely, looking toward around FYE31, you are currently using two chiplets—should we expect that to increase to, say, four? Also, if you have any rough outlook, how much improvement do you anticipate in performance per CPU by that time? I would appreciate any guidance you can share.

A4:

Indeed, our first chip is a two-chiplet solution. One of the key initiatives we are working on with others in the semiconductor industry is the development of a chiplet marketplace. Today, companies such as AMD, Intel, and NVIDIA build chiplet-based designs using proprietary interfaces that they have developed themselves. What we are aiming to do is similar to what we did 30 years ago with IP—establish industry standards that allow multiple vendors' components to be integrated seamlessly. At that time, we enabled the use of standardized interfaces, buses, and protocols within a single SoC. We are now seeking to drive a similar standardization for chiplet-to-chiplet interfaces. This would enable customers to mix and match components—for example, combining a CPU chiplet from Arm, an accelerator chiplet from NVIDIA or Rebellions, and an I/O chiplet from another vendor**,** into a single system or package. Over time, this could provide significant flexibility in system design, including the ability to scale configurations—for example, from two CPU chiplets, as in our first product, to four, six, or eight. As I mentioned earlier, we are already seeing indications of significantly higher core counts.

Some companies are exploring designs with around 500 cores, typically using multiple chiplets, and we have also seen conceptual presentations suggesting configurations exceeding 700 cores. This points to a clear trend toward increasing core density. Our first chip, for example, has 136 cores, but we expect core counts in our future products to rise over time, driven by the growing demand for CPU performance in inference data centers. While I cannot provide specific guidance on chip or core counts for FYE31, we do expect these figures to increase meaningfully.

Speaker 5

Q1:

Just to clarify, your non-GAAP EPS this year is tracking at less than $2, and you're saying it's going to be $9, so it's going up by about 5x. That does not include anything from a potential accelerator market expansion? And could you say what you thought about the TAM of that accelerator market? My question is just to clarify the magnitude of profit increase, and then could you tell us a bit more about this potential accelerator market and any comments on what you would do with all that cash?

A1:

All right, so three questions there. We are anticipating and forecasting more than $9 worth of non-GAAP EPS in FYE31, so in five years' time. About $6 of those dollars come from our existing IP business, and $3 of those come from the new chip business. The TAM for the accelerator market is indeed we anticipate growing to more than $1 trillion, from about $250 billion this year so quadrupling in size. We do think that's quite an attractive number, and last week when we announced our CPU chip, we did sort of have a watch about this space for the future, but there's nothing to announce just yet. And in terms of what shall we do with the cash from that; we have nothing to update you on right now. We generate at the moment about $1 billion a year in free cash flow. We aren't doing a buyback at the moment. Most of the requests from investors are for more shares, not fewer, and so doing a big buyback doesn't seem to be a good idea. We remain open to considering a dividend, but that is something on which no decision has been taken just yet. We remain open to investor requests there. And we do occasional acquisitions. Most of our acquisitions, however, are really about accelerating hiring. We've been hiring between 1,000 and 1,500 engineers per year. As you hire them one by one, it takes 6 to 12 months to build an engineering team that can build new technologies for you. If you acquire a company as a whole, you end up with an actual team coming in that's already set up and running. It may be that you don't want them to develop their existing product, but you can then give them a project to go and do, and you've got the senior engineers, the junior engineers, project managers; they often come with a building and with facilities managers, so the whole thing comes in a nice package that you can then get ramping up very quickly. So that's something that we have done from time to time and will probably continue to do, but those typically would be a few hundred million dollars; they're not going to make a huge dent in the

billion.

Speaker 6

Q1:

My first question is regarding the revenue outlook shown on slide 37. Could you share how much of this is expected to be accounted for by the two major customers—OpenAI and Meta? Also, how should we think about the level of visibility there? For example, are there firm commitments in place?

A1:

Not going to give specific customer information; contracts with them obviously are confidential. But as I mentioned, we have visibility to more than $1 billion worth of revenues. And if you look on slide 37, there's a little bit in FYE27 and a bigger amount in FYE28. The sum of those two parts is more than $1 billion. And we have very strong interest in delivering that. So we are as confident as you can be in revenues that far out.

Q2:

My second question is about how we should think about the profit and loss trajectory of the CPU chip business before it begins generating revenue. In other words, while you mentioned a medium-term expectation of around $3 in EPS contribution, how should we view the investment phase leading up to that? Should we expect it to be accretive at any point, or will it act as a drag on EPS during the upfront investment period? I would appreciate your thoughts.

A2:

I guess there are two things that could push down EPS. Firstly, if the revenue doesn't come through in the way that we expected or if we increase investments. So staring into our crystal ball, if Rene (CEO, Arm Holdings plc) were here today, he would be telling you that he is very confident in the $15 billion and can see multiple ways that we can achieve that with just the customer set that we have today. And in reality, we'll probably be building a bigger group of customers, and so therefore it could be even greater. But it's five years out, so there are lots of uncertainties, and so it could be lower if we end up in another war which we seem to be brewing all the time or another credit crisis. There could be lots of uncertainty. But that's based on the customers and the customer demand we can see today.

Q3:

Maybe it was lost in translation. My question was: for the next couple of years, what should we expect from the CPU chip business? Is it going to be disruptive to EPS or accretive to EPS while you're running the business at an early stage?

A3:
On the basis that almost all the costs associated with that business are already in the business today, any gross profit is going to help grow EPS. So, it is going to be accretive in the near term, assuming we are executing that plan.

Speaker 7
Q1:
I have one question regarding CPUs. You mentioned the need to reduce power consumption for CPUs. However, it is often said these days that GPUs account for the majority of power consumption. Could you help us understand how critical CPU power consumption is in this context? Is reducing GPU power consumption the primary focus for hyperscalers, or will CPU power efficiency also become increasingly important going forward?

A1:
On a chip-to-chip basis, a CPU would be lower in terms of power consumption than a GPU. As we showed earlier, the maximum power used by one of our CPUs is 300 watts. That is a fraction of a GPU. However, as we also saw earlier, in an agentic data center, you may have many more CPUs than GPUs, whereas in a training data center you have more GPUs than CPUs. So I think as we move towards more agentic workloads, as we all start using AI more in our day-to-day lives, then the data centers are going to have many more CPUs deployed, and that's going to mean that the energy efficiency of the CPU is going to become increasingly important. We will need to deliver higher levels of compute within tight power constraints. For example, in a 36-kilowatt rack configuration, we are able to operate within that power envelope while delivering strong performance, and in many cases without fully utilizing the available power budget. Clearly, this is going to become a multi-year focus. Trying to get more and more compute out of the same amount of energy. And happily, that has always been Arm's focus. Ever since we started, we always were trying to create more compute out of fewer electrons, and that is what we'll continue to do going forward.

Speaker 8
Q1:
I have one question regarding the breakdown of the $15 billion AGI CPU target. I understand that you may not be able to disclose the specific contributions from OpenAI and Meta. However, could you provide a general sense of the breakdown—for example, how much is expected to come from existing hyperscalers? Also, on slide 46, you mentioned the opportunity to expand into customer segments that you have not previously captured. Could you elaborate on what types of customers these are, and whether they are included in the $15 billion target?

A1:
The $15 billion assumes that our existing customers in this area—Amazon, Microsoft, and

Google—do not buy chips from us. We assume that they continue to develop their own chips, and so the $15 billion is only made up of new customers, some of the logos that we showed before. And for the $15 billion, we see multiple ways of getting there. It does not necessarily need to be just Meta and OpenAI; it could be the other companies as well. We see opportunities much greater than $15 billion. We are just setting $15 billion as a target, because going from nothing to something is always hard to forecast. But we think it could be even higher than $15 billion.

Cautionary Note Regarding Forward-Looking Statements

This presentation contains forward-looking statements that involve a number of risks and uncertainties. Arm Holdings plc (the "Company" or "Arm") cautions readers that any forward-looking information is not a guarantee of future performance and actual results could differ materially from the information expressed or implied by these forward-looking statements. When used in this presentation, words such as "may," "might," "will," "could," "would," "should," "expect," "is/are likely to," "intend," "plan," "objective," "anticipate," "believe," "estimate," "predict," "potential," "target," "continue," "ongoing" and similar expressions and any other statements that are not historical facts are intended to identify forward-looking statements.

Such forward-looking statements include, but are not limited to, projections and estimates of the TAM for our products and our expectations regarding revenue, licensing and royalty mix and growth, in both the near and long-term; our expectations regarding the impact of the introduction of new products on our existing operations, customer base, and demand; our vision for the future of Arm and AI computing; our ability to implement new products and business initiatives, including the expansion of our business model into production silicon; Arm AGI CPU and its expected performance, scale, efficiency and projected energy savings; our annual product roadmap; data center and agentic AI growth generally, including anticipated data center capacity; the Company's partnerships and customer expectations; projections relating to our future financial results, growth, products and services; our financial position; our market opportunity, demand and growth drivers; and any other statements that are not historical facts.

Forward-looking statements involve a number of risks, uncertainties or other factors beyond our control that may cause actual results to differ materially. These factors include, but are not limited to, our ability to implement our strategic initiatives; our development of new products and technologies; our entry into new business areas, including production silicon, and the associated execution risks; our reliance on third parties to manufacture, assemble, package and test our products; market acceptance of our products; the accuracy of comparative performance benchmarks and claims; the impact of technological development and competition; the development and growth of the AI market generally; any potential design, manufacturing, hardware or software defects; changes in customer preferences and demands; changes in industry standards; global economic, political and market conditions and fluctuations; geopolitical instability, government and industry regulation; and global competition. For a complete discussion of factors that could materially affect our financial results and operations, please refer to the reports we file from time to time with the SEC, including our Annual Report on Form 20-F. Copies of reports we file with the SEC are posted on our website and are available without charge. The Company undertakes no obligation to publicly update any forward-looking statement, whether as a result of new information, future events or otherwise.

Non-GAAP Financial Measures

Arm utilizes, and this presentation includes, certain non-GAAP financial measures that differ from measures calculated in accordance with GAAP. Arm's non-GAAP financial measures include non-GAAP operating margin and non-GAAP earnings per share. Arm believes these non-GAAP financial measures provide useful information to investors and others in understanding and evaluating its results of operations, as well as provide a useful measure for period-to-period comparisons of its business performance. Moreover, Arm has included these non-GAAP financial measures because they are key measurements used by its management internally to make operating decisions, including those related to analyzing operating expenses, evaluating performance, and performing strategic planning and annual budgeting. Arm believes that the presentation of its non-GAAP financial measures, when viewed holistically, is helpful to investors in assessing the consistency and comparability of its performance in relation to prior periods and facilitates comparisons of its financial performance relative to its competitors, particularly with respect to competitors that present similar non-GAAP financial measures in addition to their GAAP results.

Non-GAAP financial measures are presented for supplemental financial purposes only, should not be considered a substitute for financial measures prepared in accordance with GAAP, and may not align with similar financial measures presented by Arm's competitors, which may limit the ability of investors to assess Arm's performance relative to certain peer companies.

Arm is unable to provide a reconciliation of certain non-GAAP guidance measures to the corresponding GAAP measures on a forward-looking basis because doing so would not be possible without unreasonable effort due to, among other things, the potential variability and limited visibility of the excluded items. For the same reasons, Arm is unable to address the probable significance of the unavailable information.