Nasdaq 44nd Annual Investor Conference (Virtual)

Company Participants

- Colette Kress, EVP and Chief Financial Officer

Other Participants

- Mark Lipacis, Analyst, Jefferies

Presentation

Mark Lipacis {BIO 2380059 <GO>}

Hi, and welcome to the afternoon sessions here for the London Virtual NASDAQ Conference. My name is Mark Lipacis, I am Senior Semiconductor Analyst at Jefferies. It is a great honor for me to host the NVIDIA fireside chat. I'm going to read a disclaimer and then we're going to jump into Q&A with NVIDIA, CFO.

So members of the media and press are not authorized to participate in this event. If you are from the media or the press please disconnect from the call now. The content presented on the conference call is proprietary to and/or subject to the copyrights of Jefferies or third parties, you may not publish or otherwise publicly disclose the name of or otherwise identify the speakers unless Jefferies permits it in writing. By attending this event, you agree to all of these restrictions.

Okay with the disclaimer out of the way, it's my great pleasure to introduce, Colette Kress, the CFO and Executive Vice President of NVIDIA. Colette is -- previously was the Senior Vice President and CFO at Cisco's Business Technology and Operations Finance organization, as well as the CFO at Microsoft's Server and Tools division for 13 years. And prior to that, she served at Texas Instruments in a variety of financial positions.

And with that, I just want to say to Colette, welcome and thanks for joining us today.

Colette Kress {BIO 18297352 <GO>}

Thank you so much Mark for hosting us and appreciate the time here with NASDAQ.

Questions And Answers

Q - Mark Lipacis {BIO 2380059 <GO>}

Great. So Colette if it's okay with you, I'd just like to jump right into the questions. So the first question that I have is on Arm. And so, I was hoping that we could kind of do an update here. So you guys are in the process of trying to get regulatory approval for Arm. At Computex earlier this month, Jensen expressed confidence that your proposed acquisition would be approved later this year or into 2022. And there has been other news reports that, that suggested that some players are against the deal. And because they are concerned about handling -- handing greater control of designs for Arm processes to a U.S. company's company. So I was hoping, for the first question, if you can reconcile your confidence with some of the reported push back from what we're hearing in the press from other players in the technology market.

A - Colette Kress {BIO 18297352 <GO>}

Our Arm acquisition is an important piece to our future. But we look at this as a true regulatory process and following through a process just like we had seen with our overall Mellanox deal. What we end up doing is a world tour discussing with each of the jurisdictions around the world, working, starting with the U.S, moving to the U.K, Europe and overall China. Right now that process is moving as expected, as each of the jurisdictions are working on their questions, understanding the technology and all is essentially intact. Now, when you think about the overall customers, it's interesting when you think about the customers for Arm and the customers for NVIDIA are often some of the same. And so, we do have the ability to discuss with many customers in terms of what we have as a opportunity for Arm post the closure of the deal. We look to invest in both Arm helping them really build their models passed what they had built for the overall mobile platform and very well in the mobile platform. But the additional investments that it will take them to build out data center automotive and/or IoT. Our investment is not something that Arm can do as a standalone company. There is so much work that needs to be done in terms of building out the ecosystem for many of the different types of players. And many of those types of industries, particularly data center and automotive are places that we have a long-standing relationship with many of the participants in that. So we are working with customers and honestly, many of our customers and Arms customers together have learned and understood that the great value that NVIDIA can provide to their ecosystem. We'll continue to invest not only in Arm, but keep in mind, the U.K. The U.K. is where Arm is currently headquartered, we plan to keep it there. And as you've seen us develop already super computers for the industry in Cambridge for them to focus on healthcare and other use cases around there, we will continue that investment in the U.K. So, keep in mind, it is a long process regulatory and probably will take us through the beginning part of 2022. But again we remain confident.

Q - Mark Lipacis {BIO 2380059 <GO>}

And Arm has a history of open licensing its processors. Is there any intent to change that approach should you get the approval?

A - Colette Kress {BIO 18297352 <GO>}

That's correct. Arm has a great business model of licensing -- open licensing to essentially any type of customer that would arrive. There is no desire for NVIDIA to

change that model. That model is essential to why Arm has been overall successful. What we can do is, we can add to it, we can add technology that NVIDIA has to their model and continue to license in that open framework that they have completed. But correct, there is no plan to change their open licensing model.

Q - Mark Lipacis {BIO 2380059 <GO>}

I've got you. And I believe that you've indicated that with Arm and NVIDIA with increased innovation and the computing ecosystem, can you give us an example of how you might be able to do that?

A - Colette Kress {BIO 18297352 <GO>}

Sure, let's think about the data center, the data center market opportunity, Arm has already started the work in terms of creating overall cores that would be great for data center. But keep in mind, there is more to that when you think about the future of acceleration in AI. Arm is known as a CPU, probably one of the most energy-efficient CPUs. Mirroring that with something as accelerated computing with GPUs can really build a great thing together. But they need assistance, they need assistance in terms of how you put that together in terms of the new instruction sets, how you bring that to market. So there is a great example of where we can assist them. We also have some of the same experiences as it relates to PCs. So we have the ability to help them think through acceleration and AI that may be available to the PC market, through the (inaudible) as well and add to that.

Q - Mark Lipacis {BIO 2380059 <GO>}

Got you, that's very helpful. So -- and on this topic Colette, I just about an hour ago, I got an email from your Investor Relations department, talking about a interview that Jensen is going to have with the Arm CEO at what it looks like as 1 PM Eastern Time tomorrow. It's for me, what's interesting is, it seems like the narrative has been largely one sided, we haven't heard a lot from you guys on this is, is this a new effort? Should we expect to hear more kind of conversations, interviews like this, talking about the benefits of this? And maybe you could just talk about what's the intent on this interview tomorrow?

A - Colette Kress {BIO 18297352 <GO>}

Yes. Mark, this is a interview that is being put together from Pat Moorhead, who has great industry experience. And really asked the two gentlemen to come together, the two CEOs to really talk about both the future of what they see in the industries. But most importantly to talk about the deal. Where do both of them see together with Arm plus NVIDIA, highly recommend this interview. I think it will be a great event.

Q - Mark Lipacis {BIO 2380059 <GO>}

It's great. Okay, looking forward to it. I'm sure there'll be a lot of interest in that. Okay, so that -- so I want to move to a different part of Arm. So, we just talked about the regulatory process, but you in April announced that you're going to develop your own Arm CPU that you're calling Grace. And that I believe that this will ship in 2023. What's the idea behind making your own Arm CPU? And to what extent are

you doing this to work with in conjunction with your GPUs for AI workloads or versus to solve this as a standalone CPU?

A - Colette Kress {BIO 18297352 <GO>}

Yes, we announced at our GTC, the upcoming Arm CPU Grace. Now keep in mind we have built other CPUs on Arm in many of our SoCs over the last several years. But this is our first time bringing an Arm CPU to the overall data center workloads. Grace is a unique overall CPU, it is intended to be a CPU for certain types of workloads. Those workloads that we are concentrating on are high-performance computing and AI workloads, or another way of saying this, that this is not intended to be a universal general purpose CPU, GPU together. This works with our overall GPU next generation, that will also come out and we'll work in terms of the interconnect and the links that function between CPU, GPU and all of the other different pieces within the computing infrastructure. This was something that we've been working on for several years. It takes quite a bit of effort to really fine-tune what we think we need for this industry. So it actually started well before the signature of the overall Arm deal. But yes, we plan to bring it to market and we plan that it will be available for shipping in 2023.

Q - Mark Lipacis {BIO 2380059 <GO>}

Got you. What will you be able to address with Grace that you can't do with x86 right now?

A - Colette Kress {BIO 18297352 <GO>}

The key thing with Grace is, it is engineered, it is designed in terms of working with accelerated computing and working with our GPU infrastructure. When you think about some of the work that we have started to do, our focus is not on just that time in acceleration and the time that the acceleration takes place with the GPU. You have to think about all of the different time and components around the GPU to also improve their performance for overall acceleration. So our direct connection with the overall CPU is we can now work on influencing the overall acceleration in AI process, together with that GPU. x86 will continue to be our number one CPU, form factor that we also connect with, but this is a one that can really focus on acceleration in AI.

Q - Mark Lipacis {BIO 2380059 <GO>}

Got you. And you've articulated a vision for data center scale architectures that include CPUs, GPUs, DPUs integrated into a single box. It is -- do you need to acquire Arm to achieve this vision or can -- can you get by with having an Arm architectural license and just designing and making your own CPU yourself without owning the instruction set?

A - Colette Kress {BIO 18297352 <GO>}

So that is correct. We've created a outlook about moving forward that really focuses on data center computing as a whole. The modernization of data centers as we see them today will continue. And that is a focus on the CPU, the GPU, and the new and upcoming overall DPUs that are referred to as the data processing unit. This dis-aggregation of data centers is so key to how we see data centers being built today

and in the future. Taking part on what the work needs to be accomplished and creating (inaudible) systems that those pieces are all separate. Now having an Arm architectural license has enabled us now to create a CPU to work in these overall modern data centers and our CPU and give a choice of an opportunity outside of x86. Is that enough to create our work that we have here for CPU, GPU and DPU going forward? Likely, absolutely that we have a great roadmap, we've got a great overall plan not only focusing on hardware, but also focusing on software, software systems as we put this together.

Arm will be a great addition when it closes, but again we always have an opportunity with our existing plans and our existing opportunities in front of us.

Q - Mark Lipacis {BIO 2380059 <GO>}

Great. And I want to pick up on the comment you just made about the software this is -- and talking about the data center ecosystem, it seems like one of the reasons you've been so successful in the data center and gaming for that matter is the ecosystem that you have created. Can you provide a framework for investors to think about your ecosystem and the data center. I know that you've been developing CUDA since I think it's 2005. But can you help us understand the vertical market, software stacks and the work you've put into those and how that impacts the customer stickiness?

A - Colette Kress {BIO 18297352 <GO>}

Absolutely. Let's try and work backwards in terms of how we have developed to where we are today. Thinking about our systems that we create for the data center, but also the ecosystem around that has enabled the adoption and where we've reached in terms of that adoption today. It started all the way back when we were focused on PCs and focus in terms of the (inaudible) in terms of gaming. Our desire was to move to other platforms, platforms outside of the desktop or overall notebook platform. And we did by moving to where CUDA was available on every single GPU that we've built. What is CUDA? CUDA is our development platform, CUDA is a development platform that allows developers out there to determine new use cases for the overall GPU and the accelerated capabilities. So starting all the way back more than 15 years ago, developers have come on board and have leveraged overall CUDA to work with the overall GPUs. This has burked our overall data center ecosystem. We have right now more than 2.5 million developers right now that are focused on CUDA and focused on that accelerated platform, that enabled the building of AI solutions, using GPUs. That early work have started with the hyperscales that built out a big part of their infrastructure to monetize for their business models and using overall AI.

But that was the first started out saying CUDA was that first underlying software opportunity that allowed us to build out this ecosystem. But so much more has happened. When you have seen the development of frameworks, frameworks that essentially sit on top of CUDA and these AI frameworks that enables hyperscales and/or enterprises to work in those frameworks to build out solutions that they may need for the different businesses. We work together in terms of stitching together those frameworks with CUDA, so that you can run on top of overall GPUs. But we

continue work even beyond just those frameworks in terms of application SDKs. Working together with the APIs that stem from the overall applications that are used every single day with the overall enterprises and many of their workloads. So you will see a proliferation of new SDKs for large markets that we think will benefit the future of using acceleration versus just a standard x86 server configuration. All of these enterprises need assistance and they need someone working to help them on that software, so the deployment and the ease of use is there for them. So when you think about NVIDIA and our work, sure, building and designing overall hardware, hardware silicon is top of mind. But we also have a very large focus on software today. Software that is included in almost each and every single one of our platforms that we.

Q - Mark Lipacis {BIO 2380059 <GO>}

Got you. And as part of that, also recently you introduced pre-trained neural networks for genomics and natural language processing. How do we think about that? Is that another layer on top of what you've been doing in the past? How does it fit into -- the pre-trained neural network fit into your ecosystem?

A - Colette Kress {BIO 18297352 <GO>}

The solutions that we bring to market, whether it is a enterprise that just launched by a GPU server or wants to buy a full system or wants to purchase a full system, rent a system, lease a system and come with pre-trained models, we are looking to provide as many different form factors and solutions to enable their adoption of acceleration. So that is another case where -- when you think about how infrastructure was built over the last 2 decades, it started from a bottoms-up approach that has, you build out the server, data center infrastructure and then start thinking about applying the application. In many cases now, enterprises can leverage the cloud to get a first look and well-designed ideas on how they can use that in their own on-premise if they desire to bring it on-premise. So these pre-trained models are an example to say a federated way to share some of the best work that is already in the market that they can add onto and also help sustain these overall models to even get better and better by returning them back to the models to add to. So these are all as a plan to expand the ecosystem, plan -- expand the use cases of using accelerated AI and that's just one of the many different offerings that we have.

Q - Mark Lipacis {BIO 2380059 <GO>}

And you know when we talk about software, you know, you look at your margins have grown remarkably over time and you start to look like have a operating model that approaches out of a software company. Can you talk about how you get paid for your software? To what extent is it embedded in the price of your GPUs and to what extent do you charge and license separately for your software?

A - Colette Kress {BIO 18297352 <GO>}

Yeah, great question. So with software being such a big component of many of our platforms today, the question is, well, where is that? How do I see that? So our software comes with each and every platform, whether that be gaming and gaming drivers. If you think about our work in terms of RTX, DLSS, all of that is incorporated

in what we provide today and meaning, we do not charge for it separately, we do not call it out overall separately, it is embedded within that. You can take that software all the way to our professional visualization as well. As you recall, this is a very important industry, we're working with the top 50 to 100 different enterprise applications as key. We are assured that the GPUs were stitched together with those applications both from a forward looking basis, but also a backwards looking basis that it always works successfully.

When you move to data center now with our overall CUDA, our overall CUDA libraries, our compilers, our frameworks and our SDKs, at this time most of that is free and available for download and overall use. So we have incorporated the used cases and when you think about our overall for example gross margins across all platforms, you would say that they represent the cost of the overall manufacturing. And the cost of the overall software is within our overall OpEx. So this has been a very successful model in terms of expanding out the ecosystems. But there is also now an opportunity for us to license software separately. An ability for us to continue our expansion with enterprises and the data center and putting together software solutions that are very similar to the system software licensing that they do with such partners such as VMwares and other. So when enterprises are thinking about building out their data centers, they want to assure that their accelerated computing is just like what they have in terms of their existing infrastructure. They want to schedule the jobs, they want to make sure that they can monitor the overall performance of that acceleration just on top of the VMware.

So, we have created a suite of products with our enterprise AI software that will be available very similar to VMware and connected to what VMware has in terms of vSphere as well. That's just one opportunity that we will have in terms of pricing separately. Additionally, we will look at a software, SDK such Omniverse. Omniverse is a ability to attract the overall designers and creators out there in a collaborative virtual world such that they can create their digital twin to be on the floor of the design and the manufacturing to fine-tune those overall designs. This has attracted quite a bit of attention from the 20 million or more designers that are out there and POCs are in place for them to think about the wide licensing throughout enterprise design teams to use. So these are some of the examples, so we'll look in terms of what we will have incorporated in every single GPU that you can buy. But also things that are specific for certain workloads and/or industries.

Q - Mark Lipacis {BIO 2380059 <GO>}

Great. So that's very helpful color, you know, as software becomes more important, would you consider breaking out software as a line item on your -- in your bottom-up reporting of your revenues? And why or why not?1

A - Colette Kress {BIO 18297352 <GO>}

Absolutely there is an opportunity for us to break that out when it becomes a meaningful number for to communicate. We'll be able to give you progress along the way in terms of the adoption of our software. What types of customers and folks are interested in there. But when it gets to a meaningful amount, and it would be something that we could separate, we would likely do so.

Q - Mark Lipacis {BIO 2380059 <GO>}

Got you. So I want to talk about your -- the competitive landscape and your customers a bit. In the olden days when I was a young analyst, semiconductor companies used to make semiconductors and system companies used to make systems. And now you guys have morphed into a company that's providing a whole ecosystem, system and then some of your customers like Amazon or Google are developing their own chip, so that there is definitely cross-currents in the value chain. And I think one of the concerns that I hear from investors is that most developers use these AI-programing languages that people call them deep-learning frameworks like TensorFlow or Keras and that the hardware underneath like NVIDIA GPUs could be abstracted out and effectively commoditized. And so, while it's not obvious that this is happening right now, this is a concern. So what do you say to investors who could voice that concern?

A - Colette Kress {BIO 18297352 <GO>}

So there is definitely a recognition of NVIDIA has changed from its early days of a semiconductor overall chip company. This is something that we have moved to in a full system and essentially a computing company rather than looking at us just as an overall chip company. We continued to work in terms of building out ecosystems, but building out essentially anything you may need in the data center to complete that and be a full computing company. What that enables is not just bringing to market astounding hardware that we do, great hardware of course is the table stakes of building out a computing company. But it is also important to think about how you have stitched together the ecosystem to use the overall hardware that is put together. We're in the early days of acceleration and AI workloads, meaning many more innings in front of us and we've even seen a wide adoption in the last couple of years. But we are still a long way from broad-use in terms of in the enterprise. Even with the great growth that we have seen, we have a huge opportunity in front of us.

From time to time, you'll find people that say, I would like to build a customized or a chip for a certain overall workload and give that to try. But one of the things that NVIDIA has that separates us is that development platform. The development platform is not always right in front and center and the first thing that you see when you start the use of an overall GPU. But it is essential to keeping the movement forward and expanding the types of workloads that are using AI. For example, you will see in the early days, we were focused on image detection, we focused on image categorization, it's one of the very first AI types of workloads. Well it's expanded greatly when we think about the introduction of natural language processing, speech to text or vice versa. These types of new expanding workloads have extended the use case for overall GPUs and the need for developing frameworks and stitching that together. So, although it may seem that there is not a need for that underlying CUDA development platform, it is essential. Everyone in terms of, as they think going forward is jumping into CUDA, to continue to expand the use of the GPU for new types of workloads that are developing.

That work still has to be created and the software it is needed to help them advance that work. So there will be from time-to-time an opportunity for custom ASIC, a

custom check for certain workload. But if we think about the future of overall AI, having a platform that is flexible, programmable to change with all of the advancements, we think we are well-positioned for that.

Q - Mark Lipacis {BIO 2380059 <GO>}

Well, Colette, we're coming to the top of the hour. We could talk -- I have a million questions, we could talk for hours and hours. But this is great color, it's really informative conversation. Thank you very much for joining us today at the fireside chat. And I look forward to seeing you in London in person next year, hopefully.

A - Colette Kress {BIO 18297352 <GO>}

Great. Thanks Mark. Enjoyed the time today. Appreciate it.

Q - Mark Lipacis {BIO 2380059 <GO>}

Thank you very much. And thanks everybody for joining. Bye, bye.