Citi Global Technology Conference

Company Participants

- Colette M. Kress, Executive VP & CFO

Other Participants

- Atif Malik, VP and Semiconductor Capital Equipment & Specialty Semiconductor Analyst, Citigroup Inc, Research Division
- Unidentified Participant, Analyst, Unknown

Presentation

Atif Malik {BIO 15866921 <GO>}

Good afternoon, everyone. Welcome to Day 2 of Citi Technology Conference. My name is Atif Malik. I cover semiconductors and semiconductor equipment stock here at Citi. It's my pleasure to welcome Colette Kress, CFO from NVIDIA. We also have Simona Jankowski from Investor Relations. I'm going to kick it off with the -- few of my questions. And then we'll open it up to the audience to ask their questions.

Questions And Answers

Q - Atif Malik {BIO 15866921 <GO>}

Colette, welcome. Colette, NVIDIA has gone through an amazing transformation from a PC hardware company five years ago to become a leader in AI platform computing. Some investors, mostly journalists, don't quite get your verticalization approach to various end markets. You recently launched a new platform called Turing. Maybe using that as an example, can you kind of explain to us how to use hardware, graphic processors to layer software on it for -- to address various end markets like gaming, data center, pro viz and auto?

A - Colette M. Kress {BIO 18297352 <GO>}

Sure. So good question to start with. When you think about our business of the past, it really has still stemmed from the piece of driving accelerated computing, using an overall GPU, which we began back in 1999 to really accelerate the overall computing. The greatest application for the use of that accelerated computing was overall graphics or essentially gaming. But a lot has transformed from there. We took that position at that time to say there are plenty other more use cases that we can think about, where accelerated computing will continue to take off. Many of the things that we have in place in the markets that we approach today still approaching overall gaming in a very significant part of our business. But many have also now followed us in terms of our work, in terms of the data center, our work in terms of

provisualization and graphics for the enterprise as well as our work in terms of automotive. Each and every single one of these is all based on the exact same underlying architecture, which allows us the overall leverage and the scale to approach each one of these markets. Now the markets that we chose were not something that we branched off and said, "let's give it a try." We significantly researched in terms of both the size of the markets, the possibility for the transformation and the hard aspect of being ingrained in that business from a platform approach of what is necessary. So when you think about the products that we bring to market, they are all the exact same underlying architecture, the exact same set of group of people focus in the example of Turing, our most recent architecture that we just launched across the board. It's very common for you to actually walk inside and talk to our engineers and ask them what are you working on? And they will say, "Turing. "No, no. What business?" And they said, "I'm just working on Turing, the overall architecture." We spent that time then overall working on making the platform specific for the overall markets that we're doing. Even within Turing, you will see a tremendous focus on ray tracing, which we'll talk about a little bit later. You will see our focus in terms of our Tensor Cores, which are very specific use in terms of for AI, for overall inferencing. So we continue to build on top of that architecture, the customization that is necessary for the markets that we want to approach. This leverage model allows us to go to market in such much more of an efficient manner than anybody else can do because we have a house full of overall engineers that can dedicate themselves consistently to this overall platform and concentrate on probably the most important piece about our platform, which is the software layer.

What is unique across our entire platform is the software is exactly the same. Over 10 years ago, we began that work to allow the GPUs to be programmable. That programmability was the building of CUDA and building on now Version 10 of our overall CUDA platform that allows us both that underlying programmability. But also allowing components, libraries and very specific parts of CUDA focused on different parts of our industry. And that is essentially how our models worked. It allows us both the value added in terms of the platforms we sell and it allows us to equally leverage our overall OpEx effectively to approach these markets.

Q - Atif Malik {BIO 15866921 <GO>}

Great. You mentioned CUDA. And I think the last number you've shared is 850,000 CUDA developers. And you guys have talked about 3 to five years ahead in terms of software. Can you just talk about the stickiness of CUDA? Some of your customers are in multiple end markets. Is there a case where you are generating revenues across multiple end markets because of the stickiness of CUDA?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. So actually, our number in terms of our developers that are focused on our platform has now even reached the levels of 1 million. So we've surpassed 1 million developers across all of our different platforms, whether that be focused on gaming or focused on data center or even in terms of in automotive. CUDA is that key factor underneath that, that brings that developers close in. And there is absolutely a case with developers that they really want to do the work that other developers are also working on. It is very cost effective for them in terms of an investment that they can

branch together on to a new area of AI, a new area on the platform in terms of both working together in terms of developing on top of it. And that's where we have continued to see more and more developers, more and more learning overall CUDA, adding suggestions, features that we already incorporate in CUDA on a regular cadence. We're probably pushing out every year -- every half a year a new version of CUDA to overall adopt this. So because of that ease of use, because of the widespread aspect of the overall developers, that again continues to drive people to the overall platform.

Q - Atif Malik {BIO 15866921 <GO>}

Great. We'll start with your data center discussion first. At your Analyst Day in March, you updated your data center opportunity to $50 billion from $30 billion with an even greater opportunity coming from the addition of inferencing. The benefits of GPUs in training are very well understood. Can you just talk about the opportunity you have within inferencing and the benefits that the GPU garners versus an FPGA or a CPU?

A - Colette M. Kress {BIO 18297352 <GO>}

Sure. So looking back at our Analyst Day and our discussion in terms of the $50 billion opportunity, the data center is a very complex area and a lot of things are occurring, primarily focused on accelerated computing and the use of an overall GPU is really what builds up to our overall $50 billion TAM. But breaking that down, the $50 billion TAM can initially start with one of our underlying pins that we've been a part of for more than 10 years, our high-performance computing or essentially the supercomputing. This is an area where you see a significant amount of influence from overall government, building some of the biggest and greatest overall supercomputers around the world. We continue to take the top spots in overall high-performance computing. We are in 5 of the 7 top supercomputers in the world. And we are in the fastest and largest performing supercomputer here in the overall U.S. Our focus in terms of that industry, we still believe it's a $10 billion industry as we go forward, continuing to do a continuation of high-performance computing. But also it will continue to merge to be influenced by AI as we build out the use of overall accelerated in there. It's a transformation that is necessary. The more compute that is going to be necessary for high-performance computing, needs to take place in an accelerated manner. The GPU is extremely well positioned for that overall market.

The next $20 billion industry focuses in terms of what we've seen in terms of our hyperscales or what we refer to as our consumer Internet companies or essentially another tier of these large Internet service providers and the overall applications that they are doing. This is an industry that moved very fast and very quick to adopt the use of overall AI for many of the applications that we see. It started off in the early days in terms of focused on image detection, video encoding and it's advanced tremendously to the natural language processing, essentially doing search with voice commands. But also thinking about all the recommendation engines.

As you know, your news that comes to you every single morning is smarter and smarter. There are ideas in terms of your next restaurant is also smarter and smarter. That is all based on the technology of AI and using overall GPUs in many of those

cases. This is a combination of both training as well as inference. There's a tremendous amount of work in terms of training these large data sets, these deep Internets. But also a significant amount of incoming data that also needs to go through inferencing and process quite quickly. So there is definitely an inferencing focus in terms of there.

Our last $20 billion market really focuses on the industry and the overall edge. As we've talked about the overall hyperscales, you can talk about the enterprises and their overall amount of data that they are needing to process. Process efficiently with overall compute structures that allow the acceleration of their work and their speed. This again can incorporate overall training of their data as well as the overall inference. It's important when we talk about the edge, we're also talking about autonomous types of machines, machines that are not connected necessarily to the data center. But will also incur a significant amount of inferencing on the spot that needs to take place in there. We believe inference again spans across all of these different markets. And there's probably the majority or a significant portion of our overall TAM as we see forward. This in the past has primarily been a CPU market. But we are continuing from both a performance of the platform. But also a significant amount in our TensorRT software platform, able to address this market day-by-day and increasing to where it's probably a material part of our revenue today.

Q - Atif Malik {BIO 15866921 <GO>}

Great. And can you just talk about your momentum on V100 cards and instances? AWS is up and running, Google is also up and running, Microsoft is already running. Just talk about V100 momentum on the inference side?

A - Colette M. Kress {BIO 18297352 <GO>}

Sure. So V100 is our largest performance overall platform that we have. It is really a training machine through and through. It is an important piece in every single cloud across the universe. Nearly, all hyperscales have already qualified and already have instances available in terms of in their cloud instances. It's an important piece both for the cloud. But also for internal use in many of them on their applications. We have continued to expand the V100 offering, or Volta offering, not only to an overall Tesla Board. But also in terms of our DGX complete system. Our DGX system first version incorporated 8 GPUs together inside a single overall configuration. And then now we are working toward shipping our DGX-2s, which are 16 GPUs together along with our own NVSwitch that helps in terms of the overall networking and connectivity of all these things together. These are great opportunities both for hyperscales. But also for overall enterprises to quickly get on board, work in terms of their core competencies in terms of the types of workloads that they want to address with a complete system, including the software that allows them to get started. So our deployments are going quite well as you have seen in terms of results today and we look forward to shipping DGX-2.

Q - Atif Malik {BIO 15866921 <GO>}

Great. How does the launch of Turing impact growth in inferencing? You haven't shared much about the Turing impact on your data center. But the number I picked

up probably launch was a 10x improvement in inferencing. Can you talk more about the impact of Turing?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. so Turing will probably get into in terms of its aspect of gaming and pro viz. But we also announced with our Turing architecture, as we do across all of our platforms to incorporate Tensor Cores in this case. In the case of the Tensor Cores, again we are focusing on the deep learning capabilities. We're focused on using those Tensor Cores for our overall inferencing. And yes, we did comment that overall Pascal architecture. We're talking about a 10x improvement just from the overall performance of the overall hardware there, again added with our TensorRT 4 from the software. It is a very, very powerful overall compelling desire to the overall due inferencing. But stay tuned, we haven't announced that product yet.

Q - Atif Malik {BIO 15866921 <GO>}

Great. Edge computing is an emerging paradigm, which uses local computing to enable analytics at the source of the data rather than relying on cloud computing to relay data from IoT devices. Is this an attractive market for NVIDIA? I know you guys have talked about the robotics market. And can you talk about your competitive advantages in this market?

A - Colette M. Kress {BIO 18297352 <GO>}

Sure. It is a very important part, as we discussed overall IoT devices. There are IoT devices that can also be truly autonomous machines. The autonomous machines we will continue to focus on with our overall platforms are very similar to some of the things that we've seen in the data center. But truly in an overall inferencing stage. This can be robotics, this can be drones, this can be significant machinery in industrial, very common areas in terms of where we might see this. But additionally, we may focus on other types of IoT devices. This is where you can take one of our SOC overall platforms. But also use our overall DLA type of open sourced capabilities for deep learning to incorporate that into an IoT. So we think broad and wide in terms of the overall edge computing. The way you should think about this is autonomous cars as probably a very specific example of inferencing in an overall machine. When you think about the overall processing power that is going to be needed for the automotive autonomous driving, it is significant. And so we have now the capabilities to create a performance level supercomputer at the low wattage that is necessary for autonomous driving. And we have those now that those can start working in terms of incorporating and in terms of the cars.

Q - Atif Malik {BIO 15866921 <GO>}

Okay. And some of your partners and web scale have aspirations to design their own chips for certain workloads and applications. Can you talk about the merits of GPU versus an ASIC? I think there's a bit of confusion. You guys are moving towards an ASIC type core competency. But just talk about the merits of GPU versus ASIC for machine learning?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. There is always a lot of discussion of potential startups that may be working on creating our custom ASIC. Sometimes when we go back, I think, a lot of us think of the overall GPU is a conglomerate of a lot of different ASICs altogether, which allows both the flexibility in terms of the types of workloads that you want to work on. But it also leverages that consistent software platform across. The difference in terms of an ASIC, the programmability generally is not there. It may be a fixed function that says that by the time it's complete -- completed in design, it's done. It's not going back and revising that for all of the new types of evolution that you see in the overall AI world, which I think is really important. When you look over the last 4 to five years on how fast this industry has moved and if we go back in terms of the things that we were talking about four years and how that has advanced with again the million types of overall development group. But also just because of the speed of AI in terms of what they feel they can do, it may be challenging to do a fixed ASIC for all of those different things. I'm not saying that it can't be leveraged in some cases quite effectively. But our power of the flexibility of the GPU and powered by the overall software for the specifics of the industry has probably been the approach that we have found very successful for the material or majority of what we see out there in terms of AI.

Q - Atif Malik {BIO 15866921 <GO>}

Great. And moving to gaming, in August, you announced the new GeForce RTX gaming platform on the new Turing architecture. These new cards are significant step-up in both in terms of technology and price. Can you talk about the improvements in the Turing architecture for gamers and developers?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. So we get to talk about ray tracing. So ray tracing is what we brought together to overall gaming. A lot of people ask what is ray tracing? Ray tracing for those that have that type of background understand that, that's probably the holy grail of overall graphics, meaning over all this time what we have been doing is in terms of simulating the overall use of light, the use in creating the types of shadows that you need to make things look overly realistic. Now what we are doing with overall ray tracing is doing that true light illumination across all of our overall products. So this is a place that can leverage in terms of gaming. We have brought it in terms of market by discussing already with ecosystem, back with the overall developer community, back in the early spring. And the excitement was extremely high, both from Microsoft in terms of the DirectX API that they could bring to market. But also so many of the overall gaming engines bringing to light what they can do in terms of the games, in terms of the future. Then just a couple of weeks ago, now we've announced the overall cards for overall gaming. The cards will come out. We'll start with the ray tracing cards. We have the 2080 Ti, the 2080 and the 2070 overall coming to market. This is a major leap in terms of something that people probably weren't expecting for another 10 to 15 years. The games will look different. There will be a moment that you may pause whether or not that is a film or whether or not that is a game. So we're extremely excited to bring this overall technology first to the market and as widespread as we are. So these cards will be available shortly within the quarter. But we're very excited about the excitement both in the ecosystem with the developers as well as what this brings to the overall gamers.

Q - Atif Malik {BIO 15866921 <GO>}

It sounds like a game changer, no pun intended. But if you just talk about your opportunities with ray tracing, with movie studios in Hollywood and help us kind of quantify how big of an opportunity could that be incrementally?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. So that's an important piece as well. Ray tracing is 4 games. But there is also a tremendous amount of industry focused in terms of product building as well as what we see in terms of the film industry. When you look back in terms of the history and you think about all those special effect types of films that you have watched, pretty much every film has some form of special effects into them. But those that have won the awards, those that all the programmers spend on are built upon, on in video GPUs, you are now getting to the point with ray tracing that the photorealistic capabilities or the cinematic view of those types of films going forward will again have the opportunity to improve. So interestingly, when we announced the Turing architecture, we announced first our overall Quadro cards that will be used specifically for this industry. But there's a second effect in terms of the overall capabilities of both creating the cinematic overall photos. But you changed the overall process of how they build that overall film. That film is built in terms of frame by frame. And often it's a layer within the overall frame that must be built. We have an opportunity to again redo what they do in terms of the rendering part of that. When they made to put all these frames together, that is generally a piece of work that could take more than a day to actually accomplish. And we could probably shorten that amount of period of time to less than an hour or a couple hours to do so. A tremendous improvement in terms of just the overall production of the overall films. This can be taken to films, this can be taken to catalogs, this can be taken to overall photo shopping. Now we look at the overall film industry, is a very, very large industry. Probably about a $250 billion industry. And we think there is probably within that industry maybe about $1 million or more overall rendering CPU configurations out there that we could move to overall GPUs. So it's a great opportunity with us as we make the overall films better. But we shorten up the overall process.

Q - Atif Malik {BIO 15866921 <GO>}

Good. Just in terms of performance, right out-of-the-box 2070, '80 card versus 1070, 1080 card, what kind of performance improvement we'll get initially on existing games? Then with ray tracing, it's a whole different comparison.

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. We always like the effect that just says even as you upgrade to any new card, your games automatically get better, okay. What you were playing the day before and what you're now playing will absolutely improve. We will likely see a 2x improvement without even dealing with overall ray tracing on your existing games in terms of the existing performance. That is probably one of our largest leaps in terms of architecture-to-architecture, in terms of what we are doing in Turing. When you get into ray tracing, you are now talking about a 6x improvement. So a substantial improvement in terms of performance also using the ray tracing.

Q - Atif Malik {BIO 15866921 <GO>}

Great. And what are your expectations for demand for Turing versus Pascal heading into the holiday season, especially given the price difference between the 2 cards?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. When you think about our overall pricing and when you think about our overall cards, we will be selling probably for the holiday season both our Turing and our Pascal overall architecture. Remember, Turing is a leap forward in terms of overall capabilities. The performance improvement is much greater than the overall price. What that means is, you are getting for your dollar spend tremendous more improvement. Then I think from a value perspective is essentially how we approach our pricing. We want the overall gamer to feel yes, I'm getting more performance, I'm getting more overall quality of my experience for every generation we see. It's still to be seen in terms of what that mix is, how we'll see this. But we're excited in terms of the portfolio of games that are coming out for ray tracing, all the other types of games and the beginning of that holiday season is in front of us.

Q - Atif Malik {BIO 15866921 <GO>}

And what are some of the titles for the games that you're most excited about? Maybe I should ask this question to your son, who is a gaming enthusiast.

A - Colette M. Kress {BIO 18297352 <GO>}

My sons are gaming enthusiasts in terms of that. So we talked about that as we spoke at our overall earnings call. Battlefield V is well anticipated that I'm sure will be here soon and back up on track. But there is a long list in terms of games that are coming out for this season.

Q - Atif Malik {BIO 15866921 <GO>}

Great. Moving on to the automotive segment. At your Analyst Day, Jensen talked about that he believes every vehicle will be autonomous in the future and the autonomous vehicles represent $60 billion TAM by 2035. Can you just talk about NVIDIA's approach with virtual miles? And how that approach is different from competition like Intel, Mobileye?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. So when you think about the market for autonomous driving, there's many different ways that this will all be put together when we think about every single type of car. I think historically, we looked that this is only a passenger car or essentially a high-end type of car, type of market. But yes, that is still a very important part and a key component of it, which you can also think about the use of autonomous driving as it really relates to a more taxi approach or a robotaxi is what we are hearing and referring to it. That says, in a confined area whether that be a square mile or larger, you could actually map out that situation and get it to where it is fully autonomous, that is, no steering wheel, no driver, no brake pedal. It essentially has been able to move around that area. You will see and probably see the more outside more

robotaxis in your area as they are testing those overall surrounding areas, as we also work in terms of high-end Level 2, Level 3 in terms of those markets. This also doesn't even address the ability in terms of cargo and trucks in terms of what we can do in terms of AI and improve their overall productivity of transporting goods across and using autonomous in terms of those. So it's a broad and wide overall market. When we think about the use of overall virtual, what we're discussing there is to say in order for you to create a safe environment for an autonomous car, you are testing data over and over and over again. What becomes challenging is the ability to derive-tested data that you see actually out in the environment. We have the ability leveraging our work and our knowledge in terms of gaming to simulate that. We can simulate that on our own platform. We can find all the various scenarios that you may encounter without it occurring in terms of in-real life and be able to speed up the overall development cycle of how we're going to take these autonomous cars to the market. This works only on our platform because we have connected with our constellation offering an ability to do all of that simulated data as well. So different than our competitors, both from having a platform of today that they can also develop on as well as an ability to do simulated types of scenarios to assure the safety, which is their #1 concern that they bring out to market.

Q - Atif Malik {BIO 15866921 <GO>}

Great. Colette, what are the some of the milestones we need to hit to see acceleration in your auto sales? Is it Rainbow coming out. What are the things that we need to see to start painting a trajectory for your auto sales growth?

A - Colette M. Kress {BIO 18297352 <GO>}

Well it's still -- when we think about autonomous driving going forward, as we've discussed, there is a significant amount of development work. It is a significant amount of compute. They have the platforms that they are testing against and really trying to determine what is the best path in terms of market. So we're probably still in the medium term in terms of when that gets to market. In the meantime, you have seen us already continuing to focus on that high-end infotainment side, really talking about AI inside of the cockpit. So for right now what you'll likely see is more of, such as our Daimler agreement right now, where they have produced cars on the road, leveraging our technology for those high-end overall infotainment incorporating AI as well as their future SUVs that will come out. You'll also see us in terms of time-to-time different development contracts as we work with companies in terms of omni software and leveraging our platforms for those future autonomous cars. Longer term, we'll see the production value boards and platforms that will incorporate into the cars, the trucks and robotaxis.

Q - Atif Malik {BIO 15866921 <GO>}

Okay. And can you help us map the ASP or your dollar opportunity over time as you go from an autopilot to driverless robotaxi? What kind of ASPs we should expect from your products?

A - Colette M. Kress {BIO 18297352 <GO>}

Sure. So when you think about what we're building for autonomous driving, it's not a footprint that every single configuration for each one of these cars will be the same. They are all custom configurations. Redundancy is an important piece, overall amount of compute that lasts in the car over many, many years is also extremely important. So you see custom configurations that are being built. You may have a case where you have a single SOC and a GPU, you could have 2 SOCs, 2 GPUs or any other types of overall combinations. We can see by the time we move to overall robotaxis that we can extend an overall ASPs that could get over $1,000 or into the thousands in terms of there. That's likely some of that, that we'll produce, also incorporating the key components of the software that is also included with those platforms.

Q - Atif Malik {BIO 15866921 <GO>}

Okay. And China tariffs, trade wars are hot topic with investors. Can you just talk about the impact on your business from the current tariffs? And are you seeing or hearing about any kind of concerns in China, which is a big part of the gaming market?

A - Colette M. Kress {BIO 18297352 <GO>}

Well I think every industry, every company around the world is being affected of what may be perceived as just a small piece of the market. But everybody has direct and indirect impact from what is currently occurring in terms of on the tariffs. Again, our focus is on compiling together what we need to do. We will make sure that we follow all laws in all type of countries. And I think that's the best that we can do right now.

Q - Atif Malik {BIO 15866921 <GO>}

Great. Let me pause here and see if any questions in the audience.

Q - Unidentified Participant

Thank you. Regarding automotive, recently Elon Musk has elaborated on custom hardware known as hardware free. And he stated that Tesla has developed the world's most advanced computer for autonomous driving, even better than NVIDIA. So please comment on that? And how do you see competitive edge over Tesla? And second, on competition on -- and is something 7-nanometer (inaudible) some customers. So could you tell me what's your advantage in 10 nanometers versus 7 nanometers and when can we expect 7 nanometers from NVIDIA?

A - Colette M. Kress {BIO 18297352 <GO>}

Okay. So first talking about an autonomous platform. Our relationship, we were very pleased with in terms of with Tesla and helping them bring the very first autopilot to the road. We, with our overall architecture more than three years ago, brought the very first one to the market, extremely successful partnership in terms of there. Their decision to try and build their own is their decision, probably a little different than, I would say, most of the other auto manufacturers knowing the importance of top-end computing platform for their overall cars. The overall comparisons, they may be comparing about something that is three years old. And as you know, we have such

a significant workforce that is always pushing out tremendous platforms each and every quarter, even every month in some cases. So comparison in terms of what we have today, I think we have probably one of the highest performing platforms today. But again, we are very pleased to be a part of overall Tesla at the time. Now our future generations as we think about no changes in terms of there. We make no changes, of course, just like all do and we carefully consider in terms of those overall changes. Our performance improvement that we have with Turing without overall change in the overall mode is absolutely phenomenal, meaning the overall GPU has so many abilities through the overall architecture and design to concentrate on performance. So there's always a right time to think about when those right changes will be going forward. We haven't announced in terms of any new architectures coming down the pipeline. We always like to keep that a surprise. But stay tuned, we'll talk about that.

Q - Unidentified Participant

I think at the Investor Day, you had talked about in the past five years shareholder returns of 70% or 75% of free cash flow, can't remember the exact number. If we look at 2018 -- fiscal 2018, you're going to be only 25%, 30% of free cash flow. How should we think about that going forward? That's first question. Should we think about kind of 70% or 75% number going forward and even if we did, how can -- can you just talk about the rationale behind that when most of the major tech companies -- you would still be building cash and most other major tech companies that are reducing their net cash positions. You're increasing under that type of scenarios. Can you just talk about that?

A - Colette M. Kress {BIO 18297352 <GO>}

Sure. So what we discussed at our overall Analyst Day was our uses of cash. Our #1 use of cash in terms of overall cash is to obviously focus in terms of investment in the business. We have tremendously large markets in front of us. We want to make sure we are investing effectively for that. But then it comes to the case in terms of the free cash flow and the use. Our capital investments to support our business are also important for us to understand. We do have to think about -- we sell supercomputers, we actually work on supercomputers as well. So we have that as well in terms of our engineers in terms of building that out. Then that comes to distributing that overall cash. We'll look at M&A from time to time. As we know historically, we have essentially been more in terms of in the small and medium case in terms of what we buy to add on to our existing overall platform and bolt those on, that's generally been our position. Then it gets to returning cash to overall shareholders. I think we've kept up quite well. You're right. Our life to date since restarting back in '13 has been over 75%. We feel very good about that return. And we will continue to look in terms of that cash to say how do we increase that. It's not something that you can keep up with every single day. So of course, this is still a extremely important part of our capital return program is to return that. And so you'll see us continue to do that going forward. Now in terms of our overall cash levels, our cash levels for our overall market cap are within reason in terms of that to have. But again, we are going to focus on capital return for the long term.

Q - Unidentified Participant

Yes. Interested in the margin profile of Turing versus Pascal, I guess, there is maybe a bigger (inaudible) maybe some more cost that also has a healthy step-up in ASP. So what's the right way to think about that, whether that's sort of headwind or tailwind?

A - Colette M. Kress {BIO 18297352 <GO>}

Yes. When you think about our overall pricing, our pricing is again going back to our conversations really focused on the value that we can overall deliver. We always have a mix of overall gross margin across our overall platforms within gaming. Mix is probably our largest contributor to gross margin. How that mix weighs out not only within gaming. But also our mix within the other platforms that we have in our portfolio. So in the same manner, we have different overall gross margins across our portfolio of gaming. That will continue same in terms of what we've seen before. So depending on which of those platforms see different, overall volumes will affect and drive our overall gross margins.

Q - Atif Malik {BIO 15866921 <GO>}

We're almost out of time, Colette, thank you for coming to the Citi Conference.

A - Colette M. Kress {BIO 18297352 <GO>}

Thank you.