

Q4 2017 Earnings Call

Company Participants

- Arnab Chanda, VP of IR
- Colette Kress, EVP & CFO
- Jen-Hsun Huang, President and CEO

Other Participants

- Atif Malik, Analyst
- C.J. Muse, Analyst
- Craig Ellis, Analyst
- Joe Moore, Analyst
- Mark Lipacis, Analyst
- Matt Ramsay, Analyst
- Raji Gill, Analyst
- Romit Shah, Analyst
- Stephen Chin, Analyst
- Steve Smigie, Analyst
- Toshiya Hari, Analyst
- Vivek Arya, Analyst

Presentation

Operator

Good afternoon, my name is Victoria. And I'm your conference operator for today. Welcome to NVIDIA's financial results conference call.

(Operator Instructions)

Thank you, I'll now turn the call over to Arnab Chanda, Vice President of Investor Relations, to begin your conference.

Arnab Chanda {BIO 1846976 <GO>}

Thank you. Good afternoon, everyone. Welcome to NVIDIA's Conference Call for the Fourth Quarter and FY17. With me on the call today from NVIDIA are Jen-Hsun Huang, President and Chief Executive Officer; and Colette Kress, Executive Vice President and Chief Financial Officer.

FINAL

Bloomberg Transcript

FINAL

I'd like to remind you that our call is being webcast live on NVIDIA's Investor Relations website. It's also being recorded. You can hear a replay by telephone until the 16th of February, 2017. The webcast will be available for replay up until next quarter's conference call to discuss Q1 financial results. The content of today's call is NVIDIA's property. It cannot be replayed, reproduced, or transcribed without our prior written consent.

During this call, we may make forward-looking statements based on current expectations. These are subject to a number of significant risks and uncertainties. And our actual results may differ materially. For a discussion of factors that could affect our future financial results and business, please refer to the disclosure in today's earnings release, our most recent forms 10-K. And 10-Q. And the reports that we may file on Form 8-K with the Securities and Exchange Commission.

All our statements are made as of today, the 9th of February, 2017, based on information currently available to us. Except as required by law, we assume no outlook obligation to update any such statements.

During this call, we will discuss non GAAP financial measures. You can find a reconciliation of these non-GAAP financial measures to GAAP financial measures in our CFO commentary, which is posted on our website. With that, let me turn the call over to Colette.

Colette Kress {BIO 18297352 <GO>}

Thanks, Arnab. We had a stellar Q4 and FY17, with records in all of our financial metrics -- revenue, gross margin, operating margins. And EPS. Growth was driven primarily by data center tripling, with a rapid adoption of AI worldwide. Quarterly revenue reached \$2.17 billion, up 55% from a year earlier. And up 8% sequentially. And above our outlook of \$2.1 billion.

FY17 revenue was just over \$6.9 billion, up 38%. And nearly \$2 billion more than FY16. Growth for the quarter and fiscal year was broad-based, with record revenue in each of our four platforms -- gaming, professional visualization, data center. And automotive. Our full-year performance demonstrates the success of our GPU platform-based business model.

From a reporting segment perspective, Q4 GPU revenue grew 57% to \$1.85 billion from a year earlier. Tegra processor revenue was up 64% to \$257 million.

Let's start with our gaming platform. Q4 gaming record revenue was a record \$1.35 billion, rising 66% year on year. And up 8% from Q3. Gamers continue to upgrade to our new Pascal-based GPUs. Adding to our gaming lineup, we launched GTX 1050-class GPUs for notebooks, bringing e-sports and VR capabilities to mobile at great value. The GTX 1050 and 1050-TI were featured in more than 30 new models launched at last month's consumer electronics show.

To enhance the gaming experience, we announced G-Sync HDR, a technology that enables displays which are brighter and more vibrant than any other gaming monitor. Our partners have launched more than 60 G-Sync-capable monitors and laptops, enabling smooth play without screen tear artifacts. E-sports too continues to attract new gamers. Major tournaments with multi-million dollar purses are drawing enormous audiences.

This last quarter, Dota 2 held its first major tournament of the season in Boston. Tickets sold out in minutes. The prize pool reached 3 million. And millions of gamers watched online.

Moving to professional visualization, Quadro revenue grew 11% from a year ago to a record \$225 million, driven by demand for high-end real time renderings and mobile workstations. We recently launched a family of Pascal-based GPUs designed for mobile workstations, which leading OEMs are embracing.

Earlier this week, we introduced Quadro GP100, which creates a new super-computing workstation. This new type of workstation enables engineers, designers. And artists to take advantage of new technologies of photo-realism, fluid simulation. And deep learning.

Next, data center. Revenue more than tripled from a year ago. And was up 23% sequentially to \$296 million. Growth was driven by AI, cloud service providers deploying GPU instances, high-performance computing, grid graphics virtualization. And our DGX AI super-computing appliance.

AI is transforming industries worldwide. The first adopters were hyper-scale companies like Microsoft, Facebook. And Google, which use deep learnings to provide billions to customers with AI services that utilizes image recognition and voice processing. The next area of growth will occur as enterprise in such fields as health care, retail, transportation. And finance embrace deep learning on GPUs.

At November's SC16 super-computing conference, Microsoft announced that its GPU-accelerated Microsoft Cognitive Toolkit is available both in Azure Cloud and on-premises with our DGX-1 AI super-computer. In a series of related announcements at SC16, we described our plans to join the Cancer Moonshot project, in conjunction with the National Cancer Institute, the US Department of Energy. And several national labs.

To help build predictive models and guide treatment under those projects, we are collaborating on a new AI framework called CANDLE, the Cancer Distributed Learning Environment. To support this work, we unveiled our new own super-computer, the NVIDIA DGX Saturn 5, which joins together 124 DGX-1 systems. It's currently the world's 28th fastest super-computer. And the number one system in energy efficiency.

Our grid graphics virtualization business doubled year on year, driven by strong growth in the education, automotive. And energy sectors. We are excited to be hosting our eighth annual GPU Technology Conference here in Silicon Valley from May 8 to May 11. This will be the year's most important event for AI and accelerating computing. And we expect it

to be our largest GTC yet, attended by thousands of application developers, scientists and academics, as well as entrepreneurs and corporate executives.

Finally, in automotive, revenue grew to a record \$128 million, up 38% year over year. At Jen-Hsun's CES opening keynote, we demonstrated our leadership position in self-driving vehicles. With a growing list of industry players adopting our AI car platform, we also showcased AI Copilot, a technology that will recognize a driver and their preferences, monitor their alertness, understand natural spoken language. And provide alerts in dangerous situations.

One of the highlights of CES was the demonstration of our own autonomous car, dubbed BB8. More than 500 passengers took rides in the back seat without a driver behind the wheel. We announced a number of new partnerships at the show, among them were collaborations with Bosch, the world's largest automotive supplier. And ZF, Europe's leading supplier for the truck industry. Both center on developing AI car computers with Drive PX2 technology.

We also announced that we're working on cloud-to-car mapping collaborations with HERE, focused on the US and Europe. And Zen Ren, focused on Japan. These complement partnerships announced in Q3 with Europe's TomTom. And China's Baidu. Our mapping partnerships now span to all geographies.

Jen-Hsun was joined on the CES stage by Audi of America's President Scott Keogh. They announced the extension of our decade-long partnership to deliver cars with level four autonomy starting in 2020, powered by Drive PX technology. Audi will deliver level three autonomy in its A-8 luxury sedan later this year through its V-fast system, powered by NVIDIA. We also shared news at CES of our partnership with Mercedes-Benz to collaborate on a car that will be available by year's end.

During the quarter, Tesla began delivering a new autopilot system powered by the NVIDIA Drive PX2 platform in every new Model S and Model X, to be followed by the Model 3. Tesla's cars will be capable of fully autonomous operation via future software updates.

In addition, Volvo started turning over the keys to initial customers of its Drive Me program. Its XC90 SUVs equipped with Drive PX2 are capable of fully autonomous operation on designated roads in Volvo's hometown of Gothenburg, Sweden. With NVIDIA's powering the market's only self-driving cars and partnerships with leading automakers, tier-one suppliers. And mapping companies, we feel very confident in our position as the transportation industry moves to autonomous vehicles.

Next, our OEM and IP business was \$176 million, down 11% year on year.

Now turning to the rest of the income statement for Q4. Gross margins were at record levels, with GAAP gross margins at 60% and non-GAAP at 60.2%. These reflect the success of our platform approach, as well as strong demand for G-Force gaming GPUs and deep learning.

FINAL

GAAP operating expenses were \$570 million. Non-GAAP operating expenses were \$498 million, up 12% from a year earlier, reflecting head-count-related costs for our AI growth initiatives, as well as investments in sales and marketing. We are investing into huge market opportunities: AI, self-driving cars, cloud computing. And gaming. Thus we expect our operating expense growth rate to be in the high teens over the next several quarters.

GAAP operating income was \$733 million. And non GAAP operating income was \$809 million. Both more than doubled from a year earlier. Our GAAP tax rate was 10%. And our non-GAAP was 13%. These rates were lower than expected, primarily due to a decrease in the amount of earnings subject to US tax. GAAP EPS was \$0.99; non-GAAP EPS was \$1.13.

In FY17, we returned \$1 billion to shareholders through dividends and share repurchases, in line with our intentions. For FY18, we intend to return \$1.25 billion to shareholders through dividends and share repurchases.

Now turning to the outlook for the First Quarter of FY18, we expect revenue to be \$1.9 billion, plus or minus 2%. At the mid-point, this represents 46% growth over the prior year. We expect data center to grow sequentially.

Our GAAP and non-GAAP gross margins are expected to be 59.5% and 59.7%, respectively, plus or minus 50 basis points. This guidance assumes that our licensing agreement with Intel ends at March. And does not renew. GAAP operating expenses are expected to be approximately \$603 million. Non-GAAP operating expenses are expected to be approximately \$520 million.

GAAP OI&E is expected to be an expense of approximately \$20 million, including additional charges from the early conversions of convertible notes. Non-GAAP OI&E is expected to be an expense of approximately \$4 million. GAAP and non-GAAP tax rates for the First Quarter of FY18 are both expected to be 17%, plus or minus 1%, excluding any discrete items.

With that I'm going to turn it back for the Operator so we can open up for questions. Please limit your questions to just one. Operator, let's start with the questions.

Questions And Answers

Operator

Certainly. Your first question comes from the line of C.J. Muse with Evercore. C.J., your line is open.

Q - C.J. Muse

Can you hear me? Yes, my apologies, stuck on a plane here. Great results. I guess I was hoping to get a little more color on the data center side. Now that we've completed a full FY17, would love to get some clarity on the different moving parts and contributions there, then looking into FY18, how you see the growth unfolding thereafter? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Yes, C.J. First of all, thanks a lot. Well the single biggest mover would have to be data center. When you look back on last year and when you look forward, there's a lot of reasons why data center business overall grew 3X, grew by a factor of three. I would expect that to happen, to continue.

There's several elements of our data center business. There's the high-performance computing part, there's the AI part; there's grid, which is graphics virtualization; there's the cloud computing, which is providing our GPU platform up in the cloud for start-ups and enterprises and all kinds of external customers to be able to access in the cloud, as well as a brand new AI super-computing appliance that we created last year, for anybody who would like to engage deep learning and AI. But don't have the skills or don't have the resources or don't have the desire to build their own high-performance computing cluster. We integrated all of that with all of the complicated software stacks into an appliance that we maintain over the cloud. We call that DGX-1.

These pieces -- AI, high-performance computing, cloud computing, grid. And DGX -- all in contribution contributed to our growth in data center quite substantially. My sense is that as we look forward to next year, we're going to continue to see that major trend.

Of course, gaming was a very large and important factor. And my expectation is that gaming is going to continue to do that. Then longer term, our position in self-driving cars I think is becoming more and more clear to people over time. I expect that self-driving cars will be available on the road starting this year with early movers. And no later than 2020 for level four by the majors. And you might even see some of them pull into 2019. Those are some of the things we're looking forward to.

Operator

Your next question is from Vivek Arya with Bank of America.

Q - Vivek Arya {BIO 6781604 <GO>}

Thanks. I actually had one question for Jen-Hsun and one clarification for Colette. Jen-Hsun, where are we in the gaming cycle? It's been very strong the last few years. What proportion of your base do you think has a credit to Pascal? Where does that usually peak before you launch your next-generation products?

Then for Colette, just inventory dollars and base ticked up. If you could give us some comment on that? Then on OpEx activity, you did a very good job last year. But this time you're saying OpEx will go up mid-teens. Do you still think there is operating leverage in the model? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Well let's see. We typically assume that we have an installed base of \$200 million G-Force gamers. And we've upgraded about two quarters of them -- as in two operating quarters out of four years. It takes about three to four years to upgrade the entire installed base.

We started ramping Pascal, as you know, a few quarters ago. Our data would suggest that the upgrade cycle is going well. And we have plenty to go.

A - Colette Kress {BIO 18297352 <GO>}

Thanks, Vivek. On your question on inventory, as you know, in many of our businesses we are still carrying significant architectures. And a broad list of different products for those architectures across. We feel comfortable with our level of inventory as we look forward into FY18 and our sales going forward.

Your second question was regarding OpEx. And comparing it to where we finished in 2017 and moving into FY18. We do have some great opportunities, large businesses for us to go capture the overall TAMs for. And we are going to be continuing to invest in the data center, specifically in AI, self-driving cars, as well as gaming. Rather than a focus on what the specific operating margin is, we're going to focus primarily just on growing the overall TAM. And capturing that TAM on the top line.

Operator

Your next question comes from the line of Mark Lipacis -- I apologize -- from Jefferies.

Q - Mark Lipacis {BIO 2380059 <GO>}

Thanks for taking my question. Question back on the data center. The growth was impressive. And I'm wondering you mentioned that the hyper-scale players really have embraced the products first. I'm wondering if you could share with us to the extent that you think that they're embracing it for their own use, or to the extent that they're deploying it for services such as machine learning as a service and enterprises are really tapping into this, also, through the hyper-scale guys.

I'm wondering if you could help -- you mentioned that the enterprise is where you expect to see embracing the technology next in health care, retail, transport, finance. I'm wondering if you could share with us how you feel about that visibility, where you're getting that visibility from? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

On hyper-scale, you're absolutely right that there's what we call internal use for deep learning. And then there's the hosting GPU in the cloud for external high-performance computing use, which includes deep learning. Inside the hyper-scalers, the early adopters are moving obviously very fast. And -- but everybody has to follow. Deep learning has proven to be too effective. You guys -- everybody knows now that every hyper-scaler in the world is investing very heavily in deep learning. My expectation is that over the next coming years, deep learning and AI would become the essential tool by which they do their computing.

Now when they host it in the cloud, people on the cloud use it for a variety of applications. One of the reasons why the NVIDIA GPU is such a great platform is because of its broad utility. We've been working on GPU computing now for coming up on 12

years. And industry after industry, our GPU computing architecture has been embraced for high-performance computing, for data processing, for deep learning and such.

When somebody hosted up in the cloud -- for example, Amazon putting our GPUs up in the cloud -- that instance has the ability to do molecular dynamics to deep learning training to deep learning inferencing. Companies could use it for off-loading their computation, to start-ups being able to build their Company and build their application. And then host it for hundreds of millions of people to use. I think the hyper-scalers are going to continue to adopt GPU, both for internal consumption and cloud hosting for some time to come. We're just in the beginning of that cycle. And that's one of the reasons why we have a fair amount of enthusiasm around the growth here.

You mentioned enterprise. Enterprise has all woken to the power of AI. And everybody understands that they have a treasure trove of data that they would like to find a way to discover insight from. In the case of real applications that we're engaging now, you could just imagine that in the transportation industry, in car companies creating self-driving cars, one car company after another needs to take all of their road data and start to train their neuro-networks for their future self-driving cars. They use our DGX or Tesla GPUs to train the networks, which is then used to run their cars running on Drive PX. That's one application example.

Another application example, which is quite significant, it's going to be the future of processing all of the HD maps in the world. You guys might have seen that we announced at GTC this API SDK called MapWorks. MapWorks takes video information, video information that is recorded from a car. And reconstructs the three-dimensional terrain information from that live video.

It has to do computer vision, 3-D reconstruction, has to determine and detect where the lanes are, the signs are, the lights are. And even some interesting 3-D features, maybe buildings and curbs and such. It would do that automatically. And we need to process that for the world, for the planet. You could just imagine how much video is being recorded today. And how much data is being generated. And how much inferencing, computer vision. And 3-D reconstruction that has to be done. And our GPUs are really quite perfect for it.

MapWorks runs on top of our GPUs. And we're working with just about every mapping company in the world today to adopt, MapWorks and to do HD processing for their maps. That's another example.

Medical imaging companies all over the world have recognized the importance of deep learning in their ability to detect cancer and retinopathy. And the list of examples goes on and on. All the different modalities have now recognized the importance of deep learning. And you're going to start to see one medical imaging company after another.

The list of examples just keep on going. The fact of the matter is at this point, deep learning and AI has really become how future software development's going to be done

for a large number of industries. And that's the enthusiasm we're seeing around the world.

Operator

Your next question comes from the line of Atif Malik with Citigroup.

Q - Atif Malik {BIO 15866921 <GO>}

Hi, thanks for taking my question. And congratulations to the team on great results and guide. My first question is for Jen-Hsun. Jen-Hsun, on the adoption of VR for gaming, if I look at the price points of the headset and the PC, a little bit high for wider adoption. Could the use of GPU in the cloud like you guys are introducing with G-Force now be a way for the price points on VR to come down? Then I have a follow-up for Colette.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

The first year of VR has sold several hundred thousand units. And many hundreds of thousands of units. Our VR Works SDK, which allows us to process graphics in very low latency, dealing with all of the computer vision, processing. And whether it's lens warping and such, has been -- has delivered really excellent results.

The early VR is really targeted at early adopters. I think the focus of ensuring an excellent experience that surprises people, that delight people by Oculus and by Valve and by Epic and by Vive, by ourselves, by the industry, has really been a good focus. I think that we've delivered on the promise of a great experience.

The thing that we have to do now is we have to make the headsets easier to use with fewer cables. We have to make it lighter, we have to make it cheaper. Those are all things that the industry's working on. As the applications continue to come on line, you're going to see that they're going to meet themselves and find success. I think the experience is very clear that VR is exciting.

However, remember that we are also in the VR, we also brought VR to computer-aided design. And to professional applications. In this particular area, the cost is just simply not an issue. In fact, many of the applications previously were power walls or caves, VR caves, that cost hundreds of thousands of dollars. And now you could put that same experience if not even better on the desk of designers and creators.

I think that you're going to find that creative use and professional use of VR is going to grow quite rapidly. Just recently we announced a brand new Quadro P5000 with VR, the world's first VR notebook that went to market with HP and Dell. And they're doing terrifically. I would think about VR in the context of both professional applications as well as consumer applications; but I think the first year was absolutely a great success.

Operator

Your next question comes from the line of Romit Shah with Nomura.

Q - Romit Shah {BIO 16865852 <GO>}

Yes. Thank you. First of all, congratulations on a strong FY17. If I may, Jen-Hsun, the revenue beat this quarter wasn't as big as we've seen in the last several periods. And most of it came from data center. I totally understand that when your gaming business expands as much as it has it becomes harder to beat expectations by the same margin. But I was wondering if you could just spend some time talking about gaming demand. And how you think it was during the holiday season?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Well the global PC gaming market is still vibrant and growing. The number of e-sports gamers around the world is growing. You guys know that Overwatch is a home run. Activision Blizzard's Overwatch is raging all over Asia. And e-sports fans all over the world are picking it up. It's graphically very intensive, without a 1050 class and above, it's simply a non-starter. To really enjoy it, you need at least a 1060. This last quarter we launched a 1050 and a 1050 TI all over the world. And we're seeing terrific success out of that.

My expectation going into next year is that Overwatch is going to continue to spread all over the world. It really basically just started. It started in the west. And it's now moving into the east, where the largest e-sports markets are. Overwatch is going to be a huge success. League of Legends is going to continue to be a huge success. My expectation is that the e-sports, along with AAA titles that are coming out this year, is going to keep PC gaming continuing to grow.

I quite frankly thought Q4 was pretty terrific. We had a record quarter, we had a record year. I don't remember the last time that a large business the size of ours. And surely the size of a data center business, grew by factor of three. I think we're in a great position going into next year.

Operator

Your next question comes from the line of Raji Gill with Needham & Company.

Q - Raji Gill

Yes. Thanks. Jen-Hsun, can you talk a little bit about the evolution of artificial intelligence. And make a distinction between artificial intelligence versus machine learning versus deep learning? They're different kind of categorizations and implementations of those different sub-segments. So I wanted to get a sense from you how NVIDIA's end-to-end computing platform dominates machine learning relative to say the competition? Then I have a question on the gross margins, if I could, as well?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Yes, first of all, thanks for your question. The way to think about that is deep learning is a breakthrough technique in the category of machine learning. And machine learning is an essential tool to enable AI, to achieve AI. If a computer can't learn. And if you can't learn continuously and adapt with the environment, there's no way to ever achieve artificial intelligence.

FINAL

Learning, as you know, is a foundational part of intelligence. Deep learning is a breakthrough technique where the software can write software by itself, by learning from a large quantity of data. Prior to deep learning, other techniques like expert systems and role-based systems and hand-engineered features, where engineers would write algorithms to figure out how to detect a cat. And then they would figure out how to write another algorithm to detect a car. You could imagine how difficult that is. And how imperfect that is. It basically kind of works. But it doesn't work good enough, well enough, to be useful. Then deep learning came along.

The reason why deep learning took a long time to come along is because it's singular handicap is that it requires an enormous amount of data to train the network. And it requires an enormous amount of computation. That's why a lot of people credit the work that we've done with our programmable GPUs and our GPU computing platform. And the early collaboration with deep learning AI researchers as the Big Bang, if you will -- that catalyst that made modern AI possible. We made it possible to crunch through an enormous amount of data to train these very deep neuro networks.

The reason why deep learning has just swept the world, it started with convolution neuro networks. But reinforcement networks. And time sequence networks. And all kinds of interesting adversarial networks. The list of the types of networks -- there are 100 networks being created a week. Papers are coming out of everywhere. The reason why is because deep learning has proven to be quite robust. It is incredibly useful. And this tool has at the moment found no boundaries of problems that it's figured out how to solve.

I think that the traditional methods of machine learning are still going to be useful if the absolute precision of the prediction or classification is not necessarily super-important. For example, if you wanted to understand the sentiment of consumers on a particular new product that you sent, whether the sentiment is exactly right. So long as you understand the basic trend and you largely understand the sentiment, I think people would consider that information to be useful.

However, if you're using machine learning for cancer detection, obviously we need to have a level of precision that is quite high. Whether it's in health care or financial services or high-performance computing. And in some areas where for example, ad-supported internet search, small differences in accuracy could make a very large difference in financial results for the advertiser. And for the people hosting the service. In all these cases, deep learning has found a great utility. And that's one of the reasons why we're seeing so much growth. Obviously for self-driving cars, being kind of right is not a good idea. And we like to be exactly right.

Operator

Your next question comes from the line of Matt Ramsay with Canaccord.

Q - Matt Ramsay {BIO 17978411 <GO>}

Thank you very much. Jen-Hsun, you guys obviously have won some business with your automotive super-computer at Tesla in recent periods. I was curious if you could comment

Bloomberg Transcript

on some of the application porting and moving of features from the previous architecture on to your architecture. And how that's gone. And what you guys have learned through that process. And how it might be applied to some of your future partnerships? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

First of all, you know that we are a full-stack platform. The way we think about all of our platforms is from the application all the way back to the fundamental architecture and the semiconductor device.

In the case of Drive PX, we created the architecture, optimized for neural net, for sensor fusion, for high-speed processing; the semiconductor design, in the case of Drive PX2 called Parker, Tegra Parke; the system software for high-speed

sensor fusion and moving data all the way around the car -- the better you do that, the lower cost the system will be; the neural networks on top of that, that sits on top of our deep learning SDK, called KU DNN and Tensor RT, basically frameworks of AI; then on top of that, the actual algorithms for figuring out how to use that information from perception to localization to action planning.

Then on top of that, we have an API and an SDK that is integrated into map makers. We integrate into every single map, HD map service in the world, from HERE to TomTom to Zen Ren in Japan, to Baidu in China. This entire stack is a ton of software.

But your question specifically has to do with the perception layer. And that perception layer quite frankly is just a small part of the self-driving car problem. The reason for that is because in the final analysis you want to detect lanes. You've got video coming in, you want to detect lanes. You have video coming in, you want to detect a car in front of you. All we have to do -- it's not trivial. But it's also not monumental -- we have to detect and sense the lanes and the cars. And we train our networks to do so.

As you know very well now, the deep neural net has the ability to detect objects far better than any human-engineered computer vision algorithms prior to deep learning. That's one of the reasons why Tesla and others have jumped on top of the deep learning approach. And abandoned traditional hand-engineered computer vision approaches. Anyway, the answer to your question is that by working on self-driving cars end to end, we realized that this is much more than computer vision, that the self-driving car platform is a stack of software and algorithms that's quite complex. And now we've had a lot of experience doing so.

Then recently at CES, we announced partnerships with Audi, which we announced that we will have level four self-driving cars on the road by 2020. We announced a partnership with Daimler, we announced a partnership with ZF and Bosch, two of the world's top tier-one suppliers. We also announced partnerships with all of the mapping companies. If you put all that stuff together, we have the processor, we have the tier-one partnerships for the integration of the systems, we have all the software on top of it, the deep learning networks, the car partnerships, of course. And integrated into maps around the world. All

that entire stack, when you put them all together, should allow us to have self-driving cars on the road.

Operator

Your next question comes from the line of Joe Moore with Morgan Stanley.

Q - Joe Moore {BIO 17644779 <GO>}

Great, thank you for taking the question. I wonder if you could talk a little bit about the entrance market? Where are you in terms of hyper-scale adoption for specialized inference-type solutions. And how big do you think that market counts going to be? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Yes. The inference market is going to be very large. It's -- as you know very well, in the future almost every computing device will have inferencing on it. A thermostat will have inferencing on it, a bicycle lock will have inferencing on it, cameras will have inferencing on it. And self-driving cars would have a large amount of inferencing on it. Robots, vacuum cleaners, you name it, smart microphones, smart speakers, all the way into the data center. I believe that long term there will be a trillion devices that has inferencing connected to edge computing devices near them, connected to cloud computing devices, cloud computing servers. That's basically architecture.

The largest inferencing platform will likely be ARM devices. I think that goes without saying. ARM will likely be running inferencing networks, one bit (ex-nor), eight bit. And even some floating point. It just depends on what level of accuracy do you want to achieve, what level of perception do you want to achieve. And how fast you want to perceive it. The inferencing market is going to be quite large.

We're going to focus in markets where the inferencing precision, the inferencing -- the perception scenario. And the performance by which you have to do it, is mission critical. Of course self-driving cars is a perfect example of that. Robots, manufacturing robots will be another example of that. In the future, you're going to see GTC -- if you had a chance to see that. We're working with AI City partners all over the world for end-to-end video analytics. That requires very high throughput, a lot of computation. The examples go on and on, all the way back into the data center.

In the data center, there's several areas where inferencing is quite vital. I mentioned one earlier, just mapping the earth. Mapping the earth at the street level, mapping the earth in HD in three-dimensional level for self-driving cars, that process is going to require -- just a pile of GPUs running continuously as we continuously update the information that needs to be mapped.

There's an inferencing which is called off-line inferencing where you have to re-train a network after you've deployed it. You would likely re-train and re-categorize, re-classify the data using the same servers that you used for training. Even the training servers will

be used for inferencing. Then lastly, all of the nodes in cloud will be inferencing nodes in the future. I've said before that I believe that every single node in the cloud data center will have inferencing capability and accelerated inferencing capability in the future. And I continue to believe that. These are all opportunities for us.

Operator

Your next question comes from the line of Charles Wong from Goldman Sachs.

Q - Toshiya Hari {BIO 6770302 <GO>}

Hello, can you hear me?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Sure.

Q - Toshiya Hari {BIO 6770302 <GO>}

Hi, this is Toshiya from Goldman. Thanks for taking the question. And congrats on the results. I had a question on gross margins. I think you're guiding Q1 gross margins only mildly below levels you saw in fiscal Q4, despite the royalty stream from Intel rolling over. I'm guessing improvement in mix and data center and parts of gaming are driving this. But A, is that the right way to think about the puts and takes going into Q1? B, if that is indeed the case, should we expect gross margins to edge higher in future quarters and future years as data center becomes a bigger percentage of your business?

A - Colette Kress {BIO 18297352 <GO>}

Yes. This is Colette. Let me see if I can help answer that. You're correct in terms of how to look at that in Q1. The delta from Q4 to Q1 is we only have a partial part of recognition from the Intel. And that stops in the middle of March. As we move forward, as well, going into Q2, we will also have the absence of what we had in Q1 moving to Q2. I'm not here to give guidance on Q2 because we just give guidance out one quarter. But keep that in mind. There's a partial amount of Intel still left in Q1. And then it depletes in Q2.

If you think about our overall model, our overall business model, it has moved to higher-end, value-added platforms. And that's what we're selling. Our goal is absolutely to continue to concentrate on providing those higher-value platforms. That gives us the opportunity for gross margin as we make those investments in terms of an OpEx. We'll see what that kind of mix looks like as we go into Q2. But just to leave you with a understanding of Intel is probably what we can do here, okay?

Operator

Your next question comes from the line of Stephen Chin from UBS.

Q - Stephen Chin {BIO 5565507 <GO>}

Hi, thanks for taking my questions. First one is on the data center segment. Just given the expected sequential growth in that business here in the April quarter, can you talk about what products are helping to drive that? Is it possibly the DGX-1 computer box, or is it more GPUs for training purposes at the hyper-scale cloud data center?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

It would have to be Tesla processors using the cloud. There are several SKUs of Tesla processors. There's the Tesla processors used for high-performance computing. And it has FP64, FP32, ECC. It's designed. As Kuda of course and has been optimized for molecular dynamics, astrophysics, quantum chemistry, flow dynamics, the list goes on and on. The vast majority of the world's high-performance super-computing applications, imaging applications, 3-D reconstruction applications, has been ported on to our GPUs over the course of the last decade in sum. That's a very large part of our Tesla business.

Then of course we introduced on top of the architecture our deep learning stack. Our deep learning stack starts with KU DNN, the numerics kernels -- a lot of algorithms inside them to be optimized for numerical processing of all kinds of different precisions. It's integrated into frameworks of different kinds. There's so many different frameworks from Tensor RT to Cafe, to Torch to Tiano to MXNet, to CNTK; the work that we did with Microsoft, which is really excellent -- scaling it up from one GPU to many GPUs across multiple racks. That's our deep learning stack. And that's also very important.

Then the third is grid. Grid is a completely different stack. It's the world's first graphics virtualization stack, fully integrated into Citrix, integrated to VMware. Every single workstation and PC application has been verified, tested. And has the ability to be streamed from a data center.

Then last year, starting -- I think we announced it and we started shipping it in August -- our DGX-1, the world's first AI super-computer appliance, which integrates a whole bunch more software of all different types. And has the ability to -- we introduced our first NVIDIA docker. It containerizes applications. It makes it possible for you to have a whole bunch of users use one DGX. They could all be running different frameworks because most environments are heterogeneous. That's DGX-1. And it's got an exciting pipeline ahead of it.

It's really designed for companies and work groups who don't want to build their own super-computer like the hyper-scalers. And aren't quite ready to move into the cloud because they have too much data to move to the cloud. Everybody basically can easily buy a DGX-1 that's fully integrated, fully supported. And get to work on deep learning right away.

Each one of these are all part of our data center business. But the largest, because it's been around the longest, is our Tesla business; but they're all growing, every single one of them.

Operator

Your next question comes from the line of Steve Smigie with Raymond James.

Q - Steve Smigie {BIO 5514423 <GO>}

Great. Thanks a lot for the time. Just a quick question in the auto market. At CES, you had some solutions you were demonstrating that showed pretty significant declines in the size of what was being offered. You really shrunk it down a lot, yet still having great performance. If you think out to the level-four solution that you talk about for 2020, how small can you ultimately make that? It seems like you could be relative to the size of the car pretty small. Just curious if you would comment on that. And what impact having the system in the car makes on it?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

We currently have -- Drive PX today is a one-chip solution for level three. It can have -- and with two chips, two processors, you could achieve level four. With many processors, you could achieve level five today. Some people are using many processors to develop their level five. And some people are using a couple of processors to develop their level four.

Our next generation -- so that's all based on a Pascal generation. Our next generation, the processor's called Xavier. We announced that recently. Xavier basically takes four processors and shrink it into one. We'll be able to achieve level four with one processor. That's the easiest way to think about it. While achieve level three with one processor today, next year we'll achieve level four with one processor. And with several processors, you could achieve level five.

I think that the number of processors is really interesting, because we need to do the processing of sensor fusion. And we've got to do perception, we have to do localization, we have to do driving. There's a lot of functional safety aspects to it, fail over functionality, all kinds of black box recorders, all kinds of different functionality that goes into the processor. I think it's really quite interesting.

In the final analysis, what's really hard -- and this is one of the reasons why our positioning in the autonomous driving market is becoming more and more clear -- is that in the final analysis, it's really a software problem. And it's an end-to-end software problem. It goes all the way from processing in the perception processing in the car, to AI processing to helping you drive, connected to HD clouds for HD map processing all over the world. This end-to-end stack of software is really quite a large undertaking.

I just don't know where anybody else is currently doing that, with the exception of one or two companies. I think that's really where the great complexity is. We have the ability to see and to optimize across the entire range.

Now the other thing that we announced at CS that's worth mentioning is that we believe in the future, level four means that you will have autopilot capability, hands-free autopilot capability in many scenarios; however, it's unlikely to ensure and to guarantee that in every scenario that you can achieve level four. It's just not practical for some time.

However, during those circumstances, we believe that the car should still have an AI, that the car should be monitoring what's happening outside. And it should be monitoring the driver. When it's not driving for you, it's looking out for you. We call that the AI co-pilot. Whereas AI autopilot achieves level four driving, AI co-pilot looks out for you in the event that it doesn't have the confidence to drive on your behalf.

I believe that's a really big breakthrough. And we're just seeing incredible excitement about it around the industry, because I think it just makes a lot of sense. The combination of the two systems allows us to achieve or build a better car.

Operator

Your next question comes from the line of Craig Ellis with B. Riley & Company.

Q - Craig Ellis {BIO 1870408 <GO>}

Thanks for sneaking me in. And congratulations on the very good execution. Jen-Hsun, I wanted to come back to the gaming platform. You've now got the business running at a \$5 billion annualized run rate. So congratulations on the growth there. I think investors look at that as a business that's been built on the strength of a vibrant enthusiast market. But at CES, you announced the G-Force Now offering, which really allows you to tap into the more casual potential gamer. The question is, what will G-Force now do incrementally for the opportunity that you have with your gaming platform?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Yes. Appreciate that. I think first of all, the PC gaming market is growing because of a dynamic that nobody ever expected -- a dynamic that nobody ever expected 20 years ago. That's basically how video games went from being a game to becoming a sport. Not only is it a sport, it's a social sport. In order to play some of these modern e-sports games, it's a five on five. So you kind of need four other friends. As a result, in order to enjoy, to be part of this phenomenon that's sweeping the world, that it's rather sticky.

That's one of the reasons why Activision Blizzard's doing so well. That's one of the reasons why Tensen's doing so well. These two companies have benefited from tremendously from the e-sport dynamic. And we're seeing it all over the world. Although it's free to play for some people, of course you need to have a reasonably good computer to run it. And that's one of the reasons why you need G-Force in your PC. So that you can enjoy these sports.

When it's also a sport, nobody likes to lose. And surely nobody likes to blame their equipment when they do lose. Having G-Force allows -- gives you confidence and gives you an edge. For a lot of gamers, it's just a world standard.

I think that number one, e-sport is one of the reasons why gaming continues to grow. I think at this point it's fair to say that even though it's now the second most watched spectator sport on the planet behind Super Bowl, it is also the second highest paid winning sport behind football. It is -- it will soon be the largest sport in the world. I can't

imagine too many young people long term not coming into the sport somehow, as this sport continues to expand in genres. That's one of the core reasons.

Now aside for -- you asked the question about G-Force Now, which I really appreciate. The simple way to think about that is this. There are many computers in the world that simply don't have the ability to enjoy video games, whether it's extremely thin and light notebooks, Apple Macs, ChromeBooks, the integrated graphics that don't have very good capabilities. I think that it's the reasonable thing to do is to put the technology in the cloud. It took us some five years to make this possible, to put the technology in the cloud and stream the video game experience with very low latency to the computer, like Netflix does. We're basically turning the PC into a virtualized gaming experience. And putting that in the cloud.

I don't know exactly how big it's going to be yet. But our aspiration is that we would reach the parts of the market where they're casual, or they just want to have another way, another device where they can game, or somebody would like to come into the gaming world and isn't quite ready to invest the time in building a computer or buying into a G-Force PC yet. I'm anxious to learn from it. When I learn more about G-Force Now, I'll be more than happy to share it.

Operator

Unfortunately, that is all the time we have for questions. Do you have any closing remarks?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

I want to thank all of you guys for following us. We had a record year, a record quarter. And most importantly, we're at the beginning of the AI computing revolution. This is a new form of computing, a new way of computing, where parallel data processing is vital to success and GPU computing that we've been nurturing for the last decade in sum is really the perfect computing approach.

We're seeing tremendous growth and exciting growth in the data center market. Data center now represents -- had grew at 3X over year over year. And it's on its way to become a very significant business for us. Gaming is a significant business for us. And longer term, self-driving cars is going to be a really exciting growth opportunity.

The thing that has really changed about our Company, what really defines how our Company goes to market today is really the platform approach, that instead of just building a chip that is industry standard, we created software stacks on top of it to serve vertical markets that we believe will be exciting long term that we can serve. We find ourselves incredibly well-positioned now in gaming, in AI. And in self-driving cars. I want to thank all of you guys for following NVIDIA. And have a great year.

Operator

This concludes today's conference call. You may now disconnect.

FINAL

This transcript may not be 100 percent accurate and may contain misspellings and other inaccuracies. This transcript is provided "as is", without express or implied warranties of any kind. Bloomberg retains all rights to this transcript and provides it solely for your personal, non-commercial use. Bloomberg, its suppliers and third-party agents shall have no liability for errors in this transcript or for lost profits, losses, or direct, indirect, incidental, consequential, special or punitive damages in connection with the furnishing, performance or use of such transcript. Neither the information nor any opinion expressed in this transcript constitutes a solicitation of the purchase or sale of securities or commodities. Any opinion expressed in the transcript does not necessarily reflect the views of Bloomberg LP. © COPYRIGHT 2021, BLOOMBERG LP. All rights reserved. Any reproduction, redistribution or retransmission is expressly prohibited.

Bloomberg Transcript