Pat Gelsinger Keynote

Tigerfish®
Transcribing·Editing

203 Columbus Avenue · San Francisco 94133
toll-free 877-TIGERFISH

www.tigerfish.com

Pat Gelsinger Keynote

[Beginning of recorded material.]

Pat Gelsinger: Thank you. Thank you very much. It's a pleasure to be here at IDF, the world's greatest geekfest, and it's a great pleasure to be back in China. From the opening video, the challenge was the architecture for life, and today, we'll describe this idea of milli watts to peta FLOPS, imagining the scalability and breadth of the Intel architecture. And developers the world over, and here in Asia, able to take advantage of the incredible scalability and power of that architecture.

Gordon Moore, founder of Intel, brilliant man, and the world's nicest billionaire, observed this idea of doubling of density every two years, and Moore's Law has guided Intel and the industry with respect to the actions we have taken as a technology industry for now over four decades. Since the introduction of the IBM PC to today, where we're selling over a quarter billion computers per year, an incredible testament to the forward progress has made PCs and computing free and personal.

I love the 80386, partially because it was a chip I helped design, but also as we were about to introduce the 386, Andy Grove observed that we had already over $1 billion of software that was ready to run on that architecture. And he made this observation of the software spiral, the idea that faster hardware enabled new innovative software that then required faster hardware. And this spiral of innovation of software to hardware and software compatibility, how it offers investment protection and capabilities and building on the strength of the Intel architecture.

Bob Metcalf, well known for the design of Ethernet for computer networks, one of the founders of 3Com corporation, partnered with Intel and Xerox to form the Ethernet. And Metcalf's Law identified that the usefulness of a network was built on the square of the number of users of that network. And the rise of the Internet today is a profound testament to the impact and benefit of Metcalf's Law.

David Reed, a computer scientist educated at MIT and well-known for his contributions in computer network observed that Reed's Law, the utility of networks, in terms of its social effects, actually squares exponentially with the number of users as you add the two-person networks, three-person, four-person social groups that emerge. And, in fact, he says that Metcalf understated the value of the network. It wasn't N square, it was actually two to the N groupings possible. And, if anything, we see in the dramatic rise of the use of the Internet today for social networking, we see this incredible power of Reed's Law at work today.

Putting it all together we see that the scaling, through Moore's Law, the power of software compatibility through the software spiral. We see that the scaling of computer networks, and finally the benefits of social networking. And it's the sum of those, the power, the function of all of those coming together that defines the value of the Intel architecture. And over $100 billion worth of Intel architecture software was sold in 2007 alone, millions of developers worldwide. This technology, this momentum has no parallel anywhere that we could find in the technology world. But interestingly we did find a parallel that might help us to look at the Intel architecture today.

In Chinese literature, and every child in China knows the story of the Monkey King, how the Monkey King can leap great distances, he can shrink and turn into any object that he wishes. But the Monkey King has a powerful tool. The Monkey King has a golden stick. And the Monkey King can use that stick to scale up for many tasks. It can be as large as a pillar that holds up the sky or as small as a needle that he can hide in his ear. Intel architecture is like the golden stick of the Monkey King. It scales from milliwatts up to petaflops, from the very low end up to the very high end, a powerful tool, and at the Intel Developer Forum, we make all of you Monkey Kings with the powerful tool of the Intel Architecture.

IA is the architecture that's delivered compatibility. Compatibility, scalability, is that value proposition. From handhelds through large supercomputers, all able to take advantage of the scalability and the compatibility. This power allows our customers, our partners, to scale and take advantage of that architecture for platforms across every usage. As you saw in the opening video, the incredible breadth of usages of the Intel architecture. And it's hard to imagine life today or in the future where Intel architecture has not served you in some way that may be obvious but unstated, and today we want to declare clearly that Intel architecture is the architecture for life.

In the course of our keynotes today, we will actually fill out the spectrum. Anand Chandrasekher will talk about milli watts and how Intel architecture covers that end of the spectrum. Dadi Perlmutter will talk about the large volume of platforms in the client section. And I'll begin our day by looking at workstation servers and high-end computing. And let's begin by taking a look at the top of the top, high-performance computing.

The need for high-performance computing is an insatiable demand for computing. Problems of science that computer models in size and computer computation and capability are yet to be able to touch. And this touches all areas of science and life, from manufacturing and design, oil exploration. And with oil at over $100 a barrel, the seismic data and analyzing and finding reserves has enormous economic value. Financial analysis, science and analytics, defense, education, and entertainment. An insatiable demand for performance across high-performance capabilities.

Dr. Stephen Chen made this observation and he put this graph together and presented it at 2006 Super Computing in Germany. And it plots the fastest supercomputer in the world over the last several decades. And it looks at and shows Moore's Law at work, this exponential growth of capabilities as a result. And big high-performance computers are built to solve the grand challenge problems of science with both scientific but also economic and social benefit. Today, we are in a race to cross over and build the world's first peta FLOP computer, something we forecast that we will complete before the end of this decade.

Now, high-performance computing, what do you need it for? Let's look at medical imaging. An example might be just analyzing a CT scan, one of the most powerful diagnostic tools available today, currently at a 50 teraflop computing, which is used for most of these, that takes approximately 2.5 hours to look at and analyze a single scan. At a peta FLOP, we can do that in real time. Genomics research, modeling an entire cell's metabolic and signaling networks. A particle-based model has been developed to look at this. 54,000 operations per particle per timestamp. To be able to look at a single simple model of a cell would take six days on a peta FLOP computer, an Exa FLOP computer, 1,000 times more powerful is required to allow a model of an entire cell. This, we believe, is the milestone required for personalized medicine in pharmaceuticals.

But we continue to see problems even larger than that. An Exa FLOP computer, a zeta FLOP computer we believe is what is required to accomplish full weather modeling. Today at 100 tera FLOPS, we're able to have wide ranges and poor forecasts and anybody knows how well do the weather forecasters do? Not very well today. And that's because the problem is many times harder than the computing resources that are available. NASA estimates that we need a Zeta FLOP computer to be able to accurately predict weather up to the limitations of the theoretical limit of a two-week forecast. Huge economic and social benefit for accurate weather prediction, a million times faster than the fastest computer we expect to build before the end of this decade to solve that problem.

High-performance computing, an insatiable demand for capabilities. At Intel, we have a very simple vision for high-performance computing. We want to create a world in which all major breakthroughs, discoveries are powered by Intel-based supercomputers, from exploration to production. We're applying our silicon, our boards, our system development, working with external researchers, partnerships, and enabling also the key software tools and application developments. I'm happy to announce that we formed a specific business unit aimed at delivering high-performance computing needs. Intel is committed to high-performance computing. Let's look at some of our successes.

First example is the world's third fastest machine in partnership with Intel and SGI. The New Mexico Computing Applications Center at 127 tera FLOPS. Second, the Tata Computation Research Laboratory in Pune India. And this machine, for the first time ever, has placed India in the world's top ten of supercomputers. And this is also the fastest HPC in Asia today. And they're targeting applications such as neurosimulations, molecular computational fluid dynamics, and longer-term seismic, weather, and medical imaging applications. Third, the Swedish National Defense Radio Institute aimed for defense and security purposes, the world's fifth fastest machine. But if we look on a worldwide basis, Intel has supplied four out of five CPUs in the high-performance computing in the last year. We now are 354 of the world's top 500 machines, a spectacular achievement for Intel architecture and for high-performance computing.

But we're in China. How are we doing in China? And we have -- eight of the top machines are built with Intel architecture today, and we're very proud of many of our customers and partners in China. But who is number one? And we're very excited that Intel is pleased to be working with Sinopec and Intel architecture is powering the most demanding Sinopec, a leader in petroleum and seismic research in China is built with Intel architecture. And the Sinopec Shanghai Geophysical Institute has applied this specifically to seismic data processing, delivering up to a 400 percent improvement over their prior system that they were working, a 7X increase in the tera FLOPS, so to a 40 tera FLOP computer, and are not just working to deliver the hardware, but also working with them at the software applications' tuning, porting, and optimization.

In China, though, we're working very closely with key partners and OEMs, and it's my pleasure to have one of the leaders in the area of high-performance computing inside China, Dawning, represented by Li Jun, the CEO, who will talk about our work for IA and high-performance computing in China and the exciting new products that they are bringing. Please join me in welcoming Li Jun to the stage. Good morning, thank you very much, it's very good to have you.

Li Jun: [Speaking Mandarin.]

Pat Gelsinger: And we are excited to be working with Dawning and their progress -- very happy with the success that we've seen with you today. Maybe you can describe some more of your plans and product technologies.

Li Jun: [Speaking Mandarin.]

Pat Gelsinger: Well, we're very happy with the great collaboration that we've had with Dawning. We're very excited with the introduction of the new blades and the cooperation of SSI and standards. I thank you for joining us here at the IDF stage, and you, too, are today a Monkey King with the Intel architecture.

Li Jun: Thank you, thank you very much.

Pat Gelsinger: Thank you.

Li Jun: [Speaking Mandarin.]

Pat Gelsinger: We're very excited with the progress in high-performance computing. And now we'd like to look at the mission-critical computing segments of the marketplace. With Tukwila, we continue the Itanium Processor Family momentum in mission-critical applications. Customers are converting from proprietary risk architectures to Itanium. Tukwila is two billion transistors, healthy, and delivers incredible scaleable performance. And onstage here we have one of the Tukwila wafers. With Tukwila it includes the QPI, the Quick Path Interconnect, it includes the integrated memory controller, stunning thirty megabyte on-die cache. And includes multithreading. And we expect all of these will deliver greater than 2X performance capabilities, advanced RAS capabilities, and core frequencies up to two gigahertz. And as you see by the HP quote, the health that we're already seeing with the Tukwila systems is very high. What I'd like to do is not just talk about Tukwila but, today, give you the first ever public demonstration of the Itanium Tukwila processor in action. And please join me to have Gu Fan join to the stage to demonstrate the first Tukwila Itanium system. Gu Fan?

Gufan: Hello, Pat. Nice to see you again.

Pat Gelsinger: Very good, thank you. So can you demonstrate Tukwila for us?

Gufan: Yes. So today I want to show you our latest Itanium product to appear. So this is our Tukwila four processor system. And each processor has four cores and multithreading enabled. So actually each processor has eight threads now. So let's look at the demo screen. So on the right-hand side of the demo screen, you can see there are 32 threads running together and running very well. Pat, it's a validation system with RAS, running multiple OS's and applications on this system.

And, today, we are running the Red Hat Linux 5.1 version on this system. And today we have commercial applications running here. It's a Transitive application. So Transitive application is a cross-platform solution. And it can help our customer to run their software without any code changes on any other platform. So today we have Solaris software here. One application is PVWave, another is Pamcrash, application. So this application is running very well Tukwila system without any software porting. So this is a Tukwila, first time live demo, and it's very healthy and very stable.

Pat Gelsinger: So first ever public demonstration of Tukwila.

Gufan: Yesl

Pat Gelsinger: Multiple operating systems and major high-performance computing applications running on this already.

Gufan: Yes.

Pat Gelsinger: Thank you so much, Gu Fan.

Gufan: Thanks, Pat.

Pat Gelsinger: Very good. High-performance computing, mission critical computings, mainstream servers, so let's look at the Xeon family of processors. Expandable and scaleable platforms are the workhorse of the corporate environment. The Xeon 7300 family of products, also known as Caneland, four-socket quad-core systems, so native 16-core operation, enterprise-proven reliability of architecture, and really the virtualization platform of choice. Since introducing this platform late last year, we've had tremendous response from customers. And just a short list of the customer names are given here. Tremendous enthusiasm, market share gains, key customer win-backs as a result of this platform in the marketplace.

But, wait, it gets even better. We're very excited to have the Dunnington six-core product, the first ever Intel architecture six-core product. And we have a wafer here of the Dunnington processor, 1.9 billion transistors based on the Penryn microarchitecture, 16 megabytes level 3 cache, and, best of all, it's socket compatible with the quad-core Caneland platform. So customers can immediately upgrade, drop into the same socket, from Tigerton to Dunnington and get a substantial improvement in performance without any additional validation or deployment cost coming to them. Caneland platform gets even better with the Dunnington processor.

And Dunnington was specifically designed for virtualization. Virtualization is one of the hottest trends in enterprise computing today. And virtualization was initially used for simple application examples, such as consolidation and test and deployment. Today we're seeing virtualization have a rich set of usage models, including availability/continuity, dynamic datacenter and load balancing, fault tolerance,. Virtualization, a rich technology mine of new capabilities for enterprise users. So those are the trends and applications. And we are responding with substantial increments to what we include in the hardware capabilities of our chips as shown here. VT technology for IO, flex priority, and FlexMigration are key hardware-based enhancements that we're bringing to make virtualization even better on the Intel platform.

Let's see how VMware, the leader in virtualization, is taking advantage of those capabilities. And it's my pleasure, today, to be joined by Dr. Mendel Rosenblum, and in the field of virtualization, Mendel is the guy. He's cofounder of VMware, the leader in the area of virtualization. He is well-known in the industry and is seen as one of the key technical and visionaries for the use of virtualization in the industry. Please join me in welcoming to the IDF stage Dr. Mendel Rosenblum.

Mendel Rosenblum: Thanks. Thank you, Pat. I'm very excited and honored to be here to share what we've been able to do together.

Pat Gelsinger: Well, thank you very much, Mendel. Maybe you could talk about some of our work.

Mendel Rosenblum: Yes. I want to first show you a problem that we had, that we went to Intel and they helped us solve it. Before I can do that, I want to first show you what VMware's virtual infrastructure can do. With VMware's virtual infrastructure, you create your software in virtual machines, and it greatly simplifies the deployment, allows you to rapidly deploy servers, as well as reducing cost and power saving. When a virtual machine is created and handed over to the virtual infrastructure, it automatically places it on hardware. The virtual infrastructure's job is to pack the virtual machines as efficiently as possible on the available hardware.

One of the things that our virtual infrastructure can do is actually monitor all the hardware in your datacenter and look at the load of it. If it detects an imbalance like too many virtual machines on one box are consuming too many resources, it can invoke the VMware's VMotion technology to move the virtual machines around while they're running. And we use this distributed resource scheduler to actually highly utilize the computer and even save tremendous power. Because computers that -- when we pack the virtual machines together, computers that aren't being used can be powered off.

So one of the things our customer really likes is when they get a new piece of hardware, they can drop it in the datacenter, give it to virtual infrastructure, and it will start using it with our ESX3I embedded hypervisor, hardware comes ready to plug and play in your datacenter. You simply drop it in, and we will start using VMotion technology to move virtual machines around. Customers really like the speed and simplicity of approach. One of the problems we've run into is Intel comes out with new processors which have all these new features in them, and even though they're sort of backward compatible, we cannot use VMotion to them because the software might get confused when it sees a processor with all these new features while it's currently running. And so our customers had this problem, they would buy the newest processor and then they could not actually use it as part of -- to VMotion to it. So we turned to Intel for help.

Pat Gelsinger: Yeah, in fact, when VMware approached us with this, we came up with a technology solution called FlexMigration, which compliments VMotion, allowing us to bring those pools of new resources and still be able to take advantage of the VMotion capability.

Mendel Rosenblum: Yes. With FlexMigration, now, we can take the new hardware and make it look like the old hardware so that VMotion, now, actually will be able to use it, so now when our customers actually buy the newest hardware, they can drop it in their datacenter and the extended VMotion of VMware will be able to take advantage of it.

Pat Gelsinger: Talking about it's great, Mendel, but can we show it off?

Mendel Rosenblum: Yes, we can.

Pat Gelsinger: In fact, what we have over here, we have four generations of hardware. We have DP Woodcrest hardware, we have DP Harpertown, we have MP Caneland Tigerton, and even some next generation of hardware as well.

Mendel Rosenblum: Yeah. So what we have done is actually installed VMware's virtual infrastructure on all these boxes and we have a single virtual machine on it that -- and I will show you a demo of actually the virtual machine moving through all these different generations of hardware. So it starts out now actually running on the current generation MP and migrating to the current generation dual processor configuration. Now we're migrating from the current dual processor to an older generation dual processor. And all this works now seamlessly with FlexMigration. And, finally, we actually can migrate from the older generation processor to the brand new hardware that Pat has made available to us.

Pat Gelsinger: So this just shows the deep collaboration between our two teams. This is a demonstration of the Dunnington Caneland product, the product that I was describing earlier, that we already have demonstrated the ability to actually use FlexMigration on unreleased hardware in the industry. This is just tremendous, Mendel.

Mendel Rosenblum: Yes, thank you. And this is an example -- and we now have engineering teams working together and I look forward to being at future IDFs where we can show how we're leading the industry with virtualization.

Pat Gelsinger: Mendel, this is tremendous. So we've been able to show a FlexMigration across four generations of different platforms and hardware and the value of that to our mutual customers. In this and many other things we're working together. Thank you very much.

Mendel Rosenblum: Thank you.

Pat Gelsinger: Ladies and gentlemen, Dr. Mendel Rosenblum. Now, but before I let you go, Mendel, you, too, can be a Monkey King. Thank you.

Mendel Rosenblum: Thank you, Pat.

Pat Gelsinger: So we've seen the power of virtualization. We've seen the benefits that this can bring to IT and how we're working together with the industry leaders in this area of technology. We're also working in the mainstream of the server marketplace, and this is where the Xeon 5400 platform of products, built around virtualization, energy efficiency, and performance. Since we've introduced this platform last year, tremendous success and response in the industry. And let's see how Xeon processors are being used by Lenovo and Soho to support the IT infrastructure for the Beijing Olympics. Let's roll the video now, please.

[Video plays.]

Pat Gelsinger: Intel is very pleased with the use of Xeon family for supporting the Olympics by Lenovo and Soho, supporting the Olympics both on the field and this time behind the scenes. The Intel architecture, as we've said, is highly scaleable, and, in fact, we're very happy with the recent introduction of our modular server building blocks, being able to have a blade server design with CPU blades, with integrated storage capabilities, a midplane design and backplane of redundant fans, power supplies, networking, and management modules as part of it. And this allows IT organizations of different sizes and scopes to build end-to-end solutions that best fit their needs. I'm very pleased with the relationship between Intel and the Ministry of Railways, and we're happy, today, to have MORIT Centric Chief Director Mr. Zhang Honglin to the stage to describe the cooperation between Intel and the Ministry of Railways. Please join me in welcoming, right, Dr. Zhang.

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Well, we're very happy with our collaboration with Ministry of Railway, and we're just shocked by 1.4 billion people, 3.1 billion tons of goods, very impressed with the scale of the China Railway.

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Well, very good. You know, we can leverage our joint R&D work and define optimal server solutions to replace risk architecture, and we have a lot of successes in this space with you.

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: That's very good. And we're anxious to work with OEMs to meet your needs. Now, as I understand, you have many railway stations as well, and those remote railway stations operate sort of like branch offices. And can you tell us more about how you manage the IT for those remote offices as well?

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Well, that sounds like a very good solution for a modular server building blocks that we just introduced.

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Very good. Can you tell us about your high-speed railway and the plans you have there?

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Well, I'm very excited about that, and I'll be back in China next year, and I'm looking forward to the chance where I cold ride the high-speed railway with you.

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Well, thank you very much. And, Mr. Zhang, you too are a Monkey King with us here at IDF. So thank you so much for coming and joining us today. Thank you.

Zhang Honglin: [Speaks Mandarin.]

Pat Gelsinger: Ministry of Railway, one of the single largest companies, infrastructures, and computing IT environments in the world, and the Intel architecture can scale and be appropriately suited for applications needs at many different sizes and scope. IA servers satisfying a full range of those needs.

In addition to the scalability of Intel architecture, energy efficiency is becoming an increasingly important topic on a global basis. Intel has a very comprehensive approach to our ecofriendly agenda, and let's take a look at that. As a leader in IT, we take our ecotechnology leadership role very seriously. We see that we need to build into our products and technology power management capabilities at the chip level, at the system level, an even at the datacenter level. Product design that eliminates unnecessary and unproductive power use, higher efficiency and scaling when active, and lower idle when off, and the ability to quickly move between those different states. We also work to drive standards, benchmarks, and work with governments and regulatory agencies on a worldwide basis.

As one example of this, last year Intel and Google launched the climate savers computing initiative, an organization specifically focused at driving down the requirements for energy uses in IT. In fact, the goals that came of that were by 2010 we would seek to reduce by 50 percent the energy use of computing infrastructure. This would translate into 54 million tons of carbon dioxide emission reductions. That's the equivalent of taking 11 million cars off the road or equal to planting 65,000 square kilometers of trees. The IT industry, as a whole, is taking our responsibility to create eco-efficient green computing technologies. At Intel, we love hard problems, we love to take and challenge our engineers with hard problems. And they're excited about the opportunity to build energy-efficient products and technology.

Today we're very happy to announce a partnership between Climate Savers Computing Initiative and with the Chinese Electronics Energy Saving Council. And that partnership is represented by Liu Rulin, who is here to join me on the IDF stage and describe this partnership. Dr. Liu is a former member of MII, the head of the Chinese Institute of Engineering, and the co-chair, right, of the Chinese Electronics Energy Saving Council. Please join me in welcoming Dr. Liu to IDF. Thank you.

Liu Rulin: Thank you.

Pat Gelsinger: Very good. Could you please describe CEESC to the IDF audience?

Liu Rulin: [Speaks Mandarin.]

Pat Gelsinger: How does the partnership of CEESC and climate savers help the industry in China drive for energy-efficient computing?

Liu Rulin: [Speaks Mandarin.]

Pat Gelsinger: Well, thank you very much for your leadership with CIE, your leadership with CEESC, and today we're delighted with the partnership between CEESC and climate savers to bring China as a leader in both electronics and ecotechnology. Thank you so much.

Liu Rulin: Thank you.

Pat Gelsinger: And you, too, a Monkey King with us. Thank you.

Liu Rulin: Thank you very much. [Speaks Mandarin.]

Pat Gelsinger: As we saw, China, an exploding market for the use of IT technologies and now stepping forward to be a leader in ecotechnology as well. As I mentioned at the start of this section, we see that it's critical to create industry standards for how to measure and drive energy efficiency. And in light of that, we worked with the spec committee. The SPEC committee is maybe the most reputable benchmarking committee in the industry today. To create a metric for how to measure energy efficiency. It's not about peak performance, it's not about idle power, but how a server infrastructure operates under load, and that's what the spec power benchmark is intended to measure.

We're very delighted to see this released late last year, but as you can see, we're very happy for the strength of the Intel platform's offerings, which hold all of the top spots in the spec power benchmark results. And, in fact, today, I'm delighted to announce a new number one spec power results. Congratulations to Inspur, who's utilizing our new low-power Xeon products to produce a new world record spec power benchmark results. Thank you to our partners at Inspur.

So we've discussed a broad range of our server platforms. Let's look a little bit further at the next generation of server platforms to come. We spoke at fall IDF last year about our tick tock development model, and, in detail, how that was working to deliver a cadence of new technologies to the marketplace. With the Dunnington processor described today, we're finishing the 45-nanometer Penryn generation of technologies. And we're ready to move on to the next generation, the Nehalem family of products.

Nehalem is an innovative new microarchitecture. It's design is modular, allowing us to have two-, four-, or eight-core products with a range of caches and integrated feature capabilities. Next-generation technologies like Quick Path Interconnect is part of it. Each core is simultaneous multithreaded, includes integrated memory controllers, three channels of native DDR for stunning improvements in memory bandwidth, microarchitecture improvements. And incredibly healthy. And you might recall at the fall IDF, we showed our first demonstration of Nehalem just three weeks after the silicon came back. We were able to show 16 threads running at that demonstration. But today we've made a lot of progress since then, and we're excited to show you the advancements in the health of the Nehalem platform. So please join me again in inviting Gu Fan back to the stage to show Nehalem.

Gu Fan: Hi, Pat.

Pat Gelsinger: Very good, Gu Fan. So what do you have to show us on Nehalem today?

Gu Fan: Yes, I think I want to show a demo of Nehalem today, so here are two Nehalem dual processing (DP) workstations. And I will do a a performance demo using these two workstations. So we all know today we have a lot of real-world applications that need tremendous computing power. So actually, you know, here's an example. So let's look at the screen.

Actually, this screen shows a very complex fluid dynamic model based on race cars as they are moving around the tracks. And we have modeling software to help us accurately model how the air moves around the car and how the air interact with the other cars on the track. So this is a very complex task. We transfer this to workstation together to do a mini [car solution]. And we can see from this screen we have 32 threads running here, running together, to help us to do the simulation job and work our results faster. So is that good?

Pat Gelsinger: Well, Gu Fan, you know, I showed 16 threads last year, now you're showing 32, and that's good, that's progress, -- but can't we do a little bit more than that?

Gufan: Yeah, very good question. Absolutely, Pat, we can do more. And I think that here's an example. Because Nehalem has very scaleable architecture, actually, you can build a very more scaleable system now. So let's look over here. So we have two server racks here and we have 32 Super-micro standard chasis in those two racks. And inside the Super-micro standard chassis], we have our latest Nehalem DP solution there. And we use software to tie all of the servers together in concert. And we already turned all of the servers up [to do a live demo]. So let's look at the screen. Actually, this is absolutely a live demo, and last screen shows we have 64 Nehalem processors running, 256 cores running, and 512 threads running together to help us to solve most difficult tasks here. [The NHM 512 thread cluster ran CAOS Linux, Infiscale Perceus clustering software and a LINPACK workload.] So is that fantastic to see a 512 threads Nehalem system working here?

Pat Gelsinger: Now, that's worthy of IDF, yes, that's very good. So 16, 32, 512 threads of IDF running -- and with our theme of high-performance computing, this is what it's all about. Nehalem will be a tremendous product for high-performance computing thanks to Supermicro and the demo team getting 512 threads.

Gu Fan: Yes, very cool.

Pat Gelsinger: Congratulations. Thank you very much.

Gu Fan: Thanks, Pat.

Pat Gelsinger: This just shows the health of the product as we're working, and we're happy, today, to have a much more detailed Technology Insight as part of IDF to describe many more of the technical architectural, microarchitectural and cache features of the Nehalem microarchitecture as part of IDF.

But the innovation continues. At the fall IDF, we decided a number of new instruction SSE capabilities that we were bringing, the extensions that were part of Nehalem, the extensions that will be part of Westmere in our 2009 products. Today I'm also very happy to announce, for the first time, what we call the Intel Advanced Vector Extensions, and these are an expansion -- a significant expansion, over 100 new instructions -- being added for a vector extension to the Intel architecture. These instructions are part of the Sandy Bridge architecture, which comes in 2010.

The next tock, Nehalem, Westmere, and Sandy Bridge. And the advanced vector extensions moved from 128-bit SSE to 256-bit vector extensions. Includes a rich set of data alignment capabilities, and this allows more efficient data access of sparse data structures. It also enables three operand syntaxes. Today it's sort of AX plus BX equals AX. Now, you'll be able to have AX plus BX is CX as a vector operation, giving you much greater utility of the register sets in the architecture. A major extension, and today, at IDF, we're releasing the specifications for the AVX instructions.

We're quite excited with the industry response we're getting to AVX, and partners such as Microsoft and Adobe have been defining and working with us in these major instruction set extensions. And they have applications and needs that demand the performance and the new data structures that are part of these extensions.

As we've described Intel architecture as the architecture for life, much has been written and talked about visual computing. And today I will share Intel's visual computing vision and our future plans in this area. Moore's Law gives us power and flexibility to start looking at what's possible in the future of visual computing. First, graphics that we have all come to know and love today, I have news for you. It's coming to an end. Our multi-decade old 3D graphics rendering architecture that's based on a rasterization approach is no longer scaleable and suitable for the demands of the future. And we can do much better.

We can do next-generation photorealistic rendering architectures, digital media is well underway but demands high-definition from creation to consumption. And the graphics pipeline is inefficient for the next generation workloads coming to it. We see that we need to enable a next generation programmable and unified architecture to enable breathtaking visual and behavioral realism, enabling acts that look real, that act real, and finally feel real on next generation visual computing capabilities.

We believe that that requires photorealistic 3D rendering, life-like games, natural lighting, shadows and reflections. High-definition audio and video and real-time transcode and image searching consumption. Also model-based computing, e.g., physics for fluid, fire, and cloth simulation. It's all about acquiring, analyzing, modeling, and synthesizing visual workloads of the future. All of this requires a foundation that's programmable, ubiquitous, and unified at the architectural level.

Let's look at just a few early proof points of this visual computing transition that we see underway. First there's that multicore architecture helps ensure the performance capabilities to enable games to act real, and this means things like physics, simulation, high-fidelity animation, AI capabilities as part of the game. And what you're looking at here, this might look like a set of images that were videotaped in Africa. These zebras, the grass moving, this is all done as an animation, And if you look at that, what's real? You can't tell the difference, can you? An incredible amount of realism, making it act real is needed for immersive game playing experiences of the future.

It's also about photorealism in terms of the rendering architecture, and we thank to Autodesk for providing this glimpse. Ray tracing is moving from the academia research to products, and some high-end techniques on mainstream quad core workstations today, applying rendering improvements so that visualization looks real as well. An intro of these capabilities in the mainstream gaming, as well, we believe is possible in the future. And let's take a look for some of the skeptics that might be -- of what that might look like.

We shocked the world when Intel researcher Daniel Pull showed real-time ray tracing as part of a Quake environment. What you see from Quake are scenes that were taken out of a traditional rendering architecture and now done in a ray traced architecture. This enables the programmer to focus on delivering images, capabilities, that look real. And what it means for the artist, the person working on the content, is he's allowed to be an artist focusing on content creation and no longer focusing on being a computer scientist to know what the pipeline may or may not be able to do. And to the user, they've got visually rich photorealistic rendering capabilities.

So that's what we see as the vision for visual computing of the future. But to deliver visual computing, it requires a set of capabilities. It requires high performance multithreaded CPUs. It requires high-performance memory and IO subsystems at the platform level. It requires a rich set of programming, software tools, and support, and it requires a next generation of graphics architecture, a visual computing architecture of the future. And this is what's led us to create and drive our Larrabee architecture effort. I'd like to give a glimpse of a few of the details around the Larrabee graphics architecture. It's designed as an architecture from the ground up with a focus short pipeline that allows us to be very energy efficient. It adds to that core a massive vector unit processing that allows us on-die teraflops of performance on a single piece of silicon.

Over one hundred new instructions that will deliver a vector complete instruction set as part of it. It includes arithmetic functions for both integer as well as for floating point as you'd expect. New vector memory operations and operations such as scatter gather instructions that allow you to manage a rich set of sparse data structures as next generation workloads require. And also, you know, we feed this with a next generation cache architecture and a massive amount of on-die cache bandwidth as well as off-die memory bandwidth. An architecture designed to support the combination of programmable cores, fixed functions. In my career, which is now almost 30 years long in developing next generation silicon, never, never have I been involved in a program that has garnered more enthusiasm from ISVs than the Larrabee project. There is just stunning excitement about what this will enable for the next generation of visual computing.

But it requires software. And tomorrow you'll hear from Renee James talking more about our software efforts and how Intel is uniquely positioned to deliver a visual computing platform. Foundation element is IA, ubiquitous, programmable, and unified architecture, but a comprehensive set of software tools that enable a complete solution to be developed. Compilers, tunings, performance libraries. And to expand these for tools for the Larrabee architecture for standard APIs like DirectX and open GL, as well as enabling the next generation of visual computing applications.

Today we've discussed the Intel architecture, the architecture for life, from the original desktop of the 1980s to high-performance-class machines delivering hundreds of teraflops and petaflop level computing on the horizon. Scaling all the way down to milliwatts with the Atom processor that Anand will talk about. And the Intel developer community is about taking advantage of that architectural capability, combining this with the best programmers, the best platform designers that an IA solution can fit every stage and role in life. Intel architecture is the architecture for life. And the innovation continues. In closing, you can take the breadth and scalability of the Intel architecture to scale to new users. All of you in the audience are empowered to be Monkey Kings as well. Use Intel architecture to solve problems both big and small. Thank you very much.

And passing the baton to my esteemed colleague, Dadi Perlmutter, the next Monkey King for Intel. Thank you.

[End of recorded material.]