the Exascale Report tm
A Global Perspective

Published on *The Exascale Report* (http://theexascalereport.com)

Exascale: The Beginning of the Great HPC Disruption

written by Mike Bernhardt
August 2010

“It’s not the technology of Exascale that is disruptive, but the things Exascale makes possible that will be disruptive,” says Thomas Thurston, CEO of Growth Science International, “Exascale isn’t a “disruptor,” it is a foundation upon which a million disruptions can be based. It enables others to be disruptive in ways and at a scale never before imagined.”

With the eventual arrival of exascale systems, we face a level of disruption that is unlike anything this community has ever experienced. But what are we really talking about when we throw these business pop culture terms around like Frisbees™?

The term *disruptive technologies* was coined by Clayton Christensen, and introduced in his 1995 article *Disruptive Technologies: Catching the Wave*. Subsequently he changed the phrasing from “disruptive technologies” to “disruptive innovation” in his 2003 book, *The Innovator’s Solution*, because he came to the understanding that it’s the strategy or business model that the technology enables that creates the disruptive impact — not the technology itself.

There are many other definitions in use for “disruption,” but I particularly like “An act of delaying or interrupting the continuity.” That’s what is already happening. The continuity we’ve become accustomed to in HPC development is already experiencing disruption, and we’ve barely gotten started.

To clarify, in the HPC community, we tend to use the word ‘disruptive’ in association with technologies, such as the Disruptive Technologies program at SC10. According to Thurston, “Technologies or business approaches can often be categorized as sustaining or disruptive, but in the end, it’s the impact on the ecosystem that determines the true nature of change.”

Building the Infrastructure; Creating an Ecosystem

As we examine the potential disruptive impact of exascale computing, we have to keep in mind that we’re looking at much more than bigger and faster computers or innovative new technology.

While much of the early exascale discussions will inevitably focus on system size, numbers of processors, power consumption, new programming models, etc., the real disruption is the impact exascale-levels of computation will have on bringing more capabilities, new discovery, and more efficiencies to a much larger number of people than ever before.

The creation of an infrastructure to support exascale systems is a research and development journey that will take at least a decade. With an anticipated timeframe of 2018 to 2020 for the first exascale systems, we are already seeing a growing number of initiatives, collaborations, and organizations being formed around critical R&D

paths, with many more to come. But there are several initiatives already in place, some that have been around for more than two years.

We've learned from a number of HPC industry experts, exascale is already demonstrating evidence of being a disruptive force in the HPC ecosystem — challenging what we thought we knew about building computers. But the real disruption comes in two parts: first, when we actually start using those computers and changing the way we attack complex (and otherwise intractable) challenges of science and engineering and, second, when the trickle down of technology and engineering reaches the broad, commercial computing markets and changes the way that roughly 2 billion of us access and use information.

But, along with the promise of many great innovations, the experts have also issued a warning. If we make the same mistakes we did when the first teraFLOPS systems came to market — not having the proper infrastructure and ecosystem developed to support the use of the systems — we will fail. We need to do a lot of building to prepare for exascale.

In this article, ***The Exascale Report***™ turns the spotlight on two organizations that have been exploring early exascale strategies since 2008. One is a joint collaboration between the U.S. Department of Energy's Oak Ridge National Laboratory (ORNL) and Sandia National Laboratories, and the other is an Alliance between Sandia and Los Alamos National Laboratory (LANL). One common voice representing both of these important collaborations is Sudip Dosanjh. Dosanjh is a Senior Manager at Sandia who serves as co-director of both the Institute for Advanced Architectures and Algorithms (IAA) with Jeff Nichols of ORNL, and of the Alliance for Computing at the Extreme Scale (ACES) with John Morrison of Los Alamos.

To provide a broader perspective, we've also interviewed Gilad Shainer, Chairman of the HPC Advisory Council, a non-profit group with more than 150 member organizations from around the globe, and Sumit Gupta, who runs the product management and marketing team for the Tesla High Performance GPU Computing Group at NVIDIA, a company that knows a lot about disruptive innovation.

Early Efforts Bring Insight

In March of this year, I heard Sudip Dosanjh give a presentation at the National High Performance Computing and Communications Council conference in Newport, Rhode Island. His presentation, *co-Design of Architectures and Algorithms*, gave a practical and eye-opening orientation to many of the challenges we face in exascale development. But what really got my attention was to learn that Sudip had been involved in these exascale initiatives for more than two years. They weren't even on my radar screen. And, according to Dosanjh, the planning for these two initiatives goes all the way back to 2006.

U.S. Department of Energy (DOE): In it to win it

The U.S. Department of Energy (DOE) has demonstrated early, visionary leadership by getting the ball rolling on a number of initiatives designed to provide early insight into many possible approaches, not only for new technology direction, but also for new levels of collaboration. There are approximately 60 researchers shared between Sandia, ORNL, and LANL as part of the ACES and IAA initiatives. And while today these programs are pair wise partnerships between Sandia/ORNL and Sandia/Los Alamos, we can expect to see a more formalized three-way relationship in the near future — hopefully enabling even closer collaboration and increased efficiency.

As a side note, many people do not know that the International Exascale Software Project (IESP) was also started with initial funding from the DOE and NSF (The Exascale Report, July 2010 issue).

According to Dosanjh, cooperation among DOE researchers has been excellent. "Working in close collaboration, we've put together various technology roadmaps — and have done much of the groundwork to help make the case for why the technology is needed."

Dosanjh believes the early support for exascale R&D is evidence that the DOE recognizes the absolute importance of technology leadership.

Competition Drives Collaboration

He comments, “The nation needs to attack this [exascale R&D challenge] with a sense of urgency for a number of reasons. “We have mission problems to solve in national security and energy, and certainly other countries seem to be investing heavily in HPC. We’re at a technology inflection point, and I believe the change to exascale technology is ultimately going to look as profound as the change from vector to massively parallel computing. Any time you have that big of a technology disruption, tied with other countries investing in the area, you really have to maintain focus to retain leadership. The U.S. has been leading microelectronics and supercomputing for a long time, but with this technology inflection point, I think it will take some work for us to continue that leadership.”

Dosanjh has been closely involved in the IESP, as well as the two DOE initiatives for which he serves as co-director. “We really need the global cooperation, such as we have with the IESP, for critical aspects of the software stack,” he says. “The community needs to come together on a programming model — this is a critical, short-term need that we’ve identified because we don’t believe the traditional ‘MPI everywhere’ model will get us there.”

“Building the systems will not be the most difficult part. Building a system that has infrastructure — and real applications running — and within constraints such as power is where we face the biggest challenge. If we’re successful, our goal within DOE is to enable exascale computing — real applications — national security and energy apps — running on an exascale system that consumes less than 20MW of power.”

If We Plan for it, is it Really “Disruptive”?

According to Thurston, disruptions don’t have to be a surprise. “We shouldn’t think of exascale as ‘disruptive’ *per se*, but as the watermark breakthrough that will enable unprecedented waves of disruption. A disruptive enabler. The soil from which the world’s future disruptions will be based.”

Power consumption and cooling are where we will most likely experience real disruption. “So, what we’re talking about here is one teraFLOPS for about 20 watts — about what it takes to run a light bulb,” continues Dosanjh. It’s an ambitious goal to say the least. “Generating one teraFLOPS of computational power for only 20 watts. As a disruptive force, this would be a ‘game changer’ for so many companies providing support and consulting services ranging from data center design to power supplies and regulators.”

A term that is being used a lot today is “co-design” — something we haven’t seen very often in HPC. According to Dosanjh, DOE initiatives place strong emphasis on co-design efforts.

Both IAA and ACES partner with many different organizations, and those co-design partners will change based on the requirements of specific projects. Companies such as AMD, Cray, IBM, Intel, Micron, NVIDIA, Panasas and SGI have worked closely with these DOE organizations on various co-design R&D efforts.

One of the current ACES projects is an effort to develop and deploy a 2010 production petascale supercomputer this year, codenamed Cielo. According to a paper presented at the 2010 Cray Users Group, this system will be an instantiation of Cray’s Baker architecture. Plans call for Cielo to be deployed in two phases, with the first phase right around the corner — the 4th quarter of 2010. The second phase, which will increase the size of the platform by another third is targeted for the 2nd quarter of 2011.

Today, collaboration and co-design efforts are strong among DOE researchers and vendor organizations, and the potential benefits are huge. Dosanjh states, “Ultimately, the countries that lead on both the hardware and software side — will have a competitive advantage — for not just the computing industry but for a broad range of industries.”

But not everyone believes this is enough to move the community forward.

Filling the Gap

The HPC Advisory Council was formed to fill what Gilad Shainer, chairman of the council, refers to as, “a missing link.” It addresses how to extend HPC usage and technology development to bridge the gap between HPC use and its potential and bring the capabilities of HPC to new users in more areas.

“There are several government-funded organizations that basically meet up [to address the question of how to increase adoption of HPC]. They put 2,000 pages of plans together, and that’s it — no one had really taken that to the usage model to help understand how you actually do things,” says Shainer. “This is what the HPC Advisory Council was formed to address.”

The HPC advisory council is a non-profit organization, and membership is free. Currently, the membership is made up of about half users and half vendors, with more than 150 participating member organizations from around the globe. Because of this group’s widespread representation (the participants are the U.S., Japan, China, Russia, Australia, the European Union, India, South Africa, Canada and Saudi Arabia), Shanier and his colleagues have an interesting, big picture view when it comes to exascale development.

The Council has several sub-groups. One of the newest is the HPC|Scale group. Headed by Richard Graham from ORNL, and Shainer, this group includes multiple participants from a large variety of user organizations and vendors. It is a good example of user/vendor collaboration. This sub-group’s mission is to explore scalability issues and potential optimization technologies believed necessary for development of exascale systems. One project already underway is optimization of atmospheric simulations codes, initiated at the end of 2009 with the Jülich supercomputing center in Germany and NCAR in the U.S. According to Shainer, “We started earlier this year to extend this to exascale and look at ideas in that area.”

Collaboration Drives Competition

“Right now, there’s a good level of global cooperation, but what you have to understand is that as of today global doesn’t really mean worldwide,” says Shainer. “For example, on one hand you see close cooperation between the U.S. and Europe, on the other you see organizations that are more U.S. focused or some that are more European focused — with each wanting their own vendors to participate. There is already a sense of competition among countries.”

Shainer continues, “China for example is going their own direction — more or less — and targeting increased development within China. The HPC Advisory Council has a new paper, *Toward Exascale Computing*, that has just been published. It includes graphs showing where China was five years ago and where they are today. According to Shainer, “It’s a huge performance jump — a giant step. If this pace of progress is maintained, then the next number one system in the world is going to be in China.”

“There definitely needs to be cooperation…collaboration — on both the software side and hardware side of exascale research,” Shainer adds. “On the software side there definitely needs to be cooperation across multiple places. On the hardware side, there needs to be more cooperation between the different vendors —and we definitely see some of that. But moving towards our goals, we’re probably going to see more competition evolving. You will see less information being distributed and shared between organizations and vendors, and more things will become secret and handled under NDA. Moving forward we will definitely see more competition in this area.”

Shainer agrees with Dosanjh regarding the inevitability of reaching exascale. “It’s much more than just bringing the first system to market. Right now there is nothing we see that would prevent us from getting to exascale — in just the sense of a system. The problems identified are all solvable. In fact, at the most recent SciDAC meeting, nothing stood out as being a barrier that would keep us from reaching exascale. Sure, we face some pretty big challenges like power and memory issues — but they are all solvable.”

Exascale Reality Check — Can We Make These Systems Affordable?

Many industry spokespeople agree that the challenges are (or should be) solvable. But are they solvable in a way that will make exascale systems affordable to more than just one or two of the world’s largest research organizations?

We spoke with Sumit Gupta, the senior marketing manager of NVIDIA's Tesla High Performance Graphics Processing Unit (GPU) Computing Group to get his company's perspective on preparing for exascale. Most people will agree that NVIDIA has been a disruptive force in the HPC community, and the general-purpose GPU (GPGPU) has radically changed the landscape of technical computing.

NVIDIA has created an entire ecosystem around their hardware, including a very popular technology gathering, the GPU Technology Conference, being held next month at the San Jose Convention Center. Last year at the conference we saw a surge in new companies and new collaborations — all enabled in some fashion by NVIDIA's hardware and software. This year, with nearly five times as many papers submitted, we can expect to find the topic of preparing for exascale in a number of the presentations.

"From our point of view, we believe that heterogeneous systems are critical and represent the only way forward for exascale systems, the only viable strategy that one can consider today," said Gupta.

"We're never going to get there if we don't cooperate, and the main cooperation is going to be with the government agencies, the processor providers, the system providers, the applications and tool developers."

Gupta continued, "Many of the petaFLOPS class systems being built now and over the next two years are using NVIDIA GPUs. This whole notion of heterogeneous computing has taken off very fast. The research has been going for quite awhile but it is now much more interesting because today petaFLOPS are easier to acquire. The next step will be incremental — building a 10 petaFLOPS system and so forth — with the focus being on application scaling."

Every person I've interviewed to date has echoed Gupta's comments. Heterogeneous systems and some combination of CPUs and GPUs will be required to reach exascale. Most have zeroed in on the technology side. But in addition to technology and engineering research, clearly a demonstrated strength for NVIDIA, the team there is keenly focused on a very important aspect that not many people have been talking about, and one that Gupta feels is essential if exascale deployment is to be successful.

According to Gupta, "I think one of the things we believe in right now — the only way to make exascale accessible to many people — is to enable the consumer market to drive the HPC processor. In other words, any government can always pay to go build one exascale supercomputer, but if you really want to make it affordable and practical for many people, research organizations, academia, government labs, industry, to have access to exascale systems, you need the consumer market to drive the HPC processor."

"Let me break that down," he says. "One of the trends obvious to most people today is that the x86 CPU now dominates the supercomputing TOP500, and the reason that happened is the economics of x86 architectures is actually being driven by the consumer market. My laptop, your laptop, everyone's has an x86 CPU in it. Intel tacks on an SSE side unit that allows them to target the HPC market, and the whole premise of NVIDIA's Tesla GPU strategy is exactly the same."

"The GPGPU became possible because graphic workloads started demanding the same kind of performance and features similar to the HPC workload," adds Gupta. "In fact, graphics workloads are becoming more and more like HPC workloads, so GPUs will continue to innovate for the mainstream computer market in a way that's good for HPC."

"If you look at the Fermi architecture, that's a half billion dollar investment just to get to market," adds Gupta. "To do that just for HPC — if you were going to go off and build a processor just for exascale — you'd have a problem with the economics in a big way. For example, for Intel or AMD or NVIDIA — for any company — to spend half a billion dollars to build a processor that would only be useful for HPC, with that processor only getting deployed in a few systems — would mean the processor cost would become prohibitive. The only way to make it affordable is if that same investment can also be used for the consumer product line. That's ours strategy and Intel's exact strategy."

Gupta points out that there are only a few companies in the world who can drive the processor development for exascale. "The investment is not just in the silicon — the computer architecture team, and the implementation team — But also having a huge software team is what I think is the enabler. HPC has a high barrier to entry."

Gupta believes the reason why GPUs have been successful is because they are easy to program, and because NVIDIA ensured the programming model, CUDA, was enabled in C, C++ and Fortran — something they identified as critical to adoption.

Many people were caught a bit by surprise by the rapid acceptance of the GPGPU. I can remember the NVIDIA booth at SC08, which seemed to be the busiest booth on the exhibition floor. However, listening to the conference buzz, many people were saying that the GPGPU wouldn't catch on. Today, the disruptive effect of the GPGPU on the HPC ecosystem is well documented. NVIDIA's strategy for penetrating the HPC market was spot on.

Gupta says, "Exascale is still very far out and of course things may change. We could realize there is no commercial train to ride on to get there, and we may have to do something different. But today it seems that if we follow this trend — if NVIDIA stays on the graphics commercial train — it will give us a GPU that will be an exascale GPU."

Like Dosanjh and Shainer, Gupta has the same perspective on bringing an exascale system to market in terms of it being much more than just the hardware, and emphasizes a big part of building an affordable exascale system will be power. Gupta comments, "Jaguar is 7MW. That's a lot of power. That's a small town. If you multiply that by 1,000 times, that's probably Los Angeles. So how do you power something like that? In fact, how many countries can power something like that?" The challenge of building an exascale system that will run at the target of 20MW is a daunting task to say the least. [*According to the Los Angeles Department of Water and Power website, the actual multiplier is closer to 400.*]

He continued, "The system Dawning built in China is much more power-efficient — it scales a little better — but still not there. Processor technology, new GPUs, and new CPUs will buy us more — but we are going to need the time to get there."

Gupta believes the first exascale system will be fielded in the U.S. "This is about a number of things, but mostly about need. A country has to believe that exascale is fundamental to the future of its economy and fundamental to the future of science. This is expensive, not just to build, but to run, and if you have a system this big, you better have a huge pool of scientists to actually use it. I believe the U.S. is that country."

Competition Drives More Competition

Well, after that comment, I contacted our friend in China, Mr. Zheng, (introduced in the July 2010 issue of The Exascale Report) in my article, *Can the Exascale Effort Survive the Need for Global Cooperation?* He is still adamant about remaining anonymous for fear of retribution from his employer. He points out that China's Minister of Science and Technology, Wan Gang, was in Washington, D.C. meeting with the White House Office of Science and Technology Policy to discuss the recent attention being drawn to China's policy of "indigenous innovation," aspects of which are perceived by many as anti-foreign, but also clearly designed to boost China's technology leadership.

"We have researchers involved in pretty much every exascale initiative where we can participate and are not blocked because of political barriers," said Zheng. "At this point, we believe our nation has a distinct advantage and could very well emerge as the frontrunner of exascale development and deployment because of both our unified desire and our government's strong sense of importance being placed on technology leadership. I have no doubt we will be one of the first — if not the first — to demonstrate a working exascale system based on hardware, operating system and benchmarks. I'm well aware of the growing sentiment that we need so much more than just a benchmark — and we are working to align all the necessary resources, including funding, to ensure we produce much more than just a great benchmark number."

He continues, "Most people, in most countries, would be shocked if they had visibility to the massive effort we already have in place. I personally think your perspective on exascale as a disrupting innovation or influence in HPC is accurate. I believe the availability of exascale systems will have a disruptive effect on every nation, on millions of people, and certainly on scientific exploration and discovery. We are already talking about resulting industries that might spin out from exascale."

Dealing with the Great HPC Disruption: It Takes a Lot of Energy

More than any aspect of computer technology, concerns seem to run deepest when it comes to the anticipated power consumption of an exascale system. According to a recent paper titled, *On the Path to Exascale* published in the International Journal of Distributed Systems and Technologies, April-June 2010, and authored by thirteen (yes, 13) members of the IAA, "The architectural challenges for reaching exascale are dominated by power, memory, interconnection networks and resilience."

Most technologists feel that power is the one area in which significant breakthroughs will be the most difficult to achieve, and if (or when) accomplished will have the most disruptive effect on the HPC ecosystem and numerous other markets.

Mass distribution of electricity as part of the standard infrastructure of many nations of the world was a huge breakthrough that enabled unprecedented waves of disruption. As we learned to harness electricity, revolutionary products and technologies were developed creating a disruptive effect on numerous industries worldwide. Moving forward, as we tackle the energy challenges related to exascale, new innovations in the area of power and energy consumption will have a tremendous direct and trickle-down impact on all of us.

Computing technology aside, energy continues to be one of the biggest challenges we face. The development of exascale systems will only magnify this challenge. But the energy challenge has been an ongoing topic of discussion — with impact well beyond the HPC-related energy challenge.

In 2005, John Marburger was the Director, Office of Science and Technology Policy under the Executive Office of the President, (he currently serves as the Interim Vice President for Research at Stony Brook University). Marburger raised a number of issues that deal with the very semantics we are wrestling with today. He framed the discussion exploring the link between then current revolutionary science and a possible revolution in energy technology. As part of this discussion, he talked about the impact of disruptive innovation and disruptive technologies, and raised a particular point that I've always liked.

Marburger stated, "We ought to be clear about what is a revolution in science or technology, and this is somewhat a matter of opinion, especially when we are trying to decide whether we are living in one."

To quote Thurston again, "No technology is, itself, inherently sustaining or disruptive. Rather, sustaining or disruptive character is given to technologies (be they products like a light bulb, or categories like electricity) due to how they take root in the marketplace and turn it on its head over time. For example, the CPU wasn't necessarily disruptive by virtue of merely existing. Rather, it became disruptive relative to printed circuit board logic because of what it ended up enabling over time — computing went from something only the elite could do (minicomputers) to something everyone could do (PCs)."

So, what are we facing? Is exascale the enabler of the great HPC disruption or simply a matter of evolution? Would it be more accurate to refer to our journey to exascale as a "planned revolution?" Are we already living in a revolution? (See the article by John West in the July 2010 issue of The Exascale Report, *Evolution or Revolution*.)

I suppose it doesn't matter if exascale is achieved by an evolutionary or a revolutionary path, the resulting changes in products, collaborations, infrastructure and ecosystem will drive new approaches to business development, new manufacturing processes, and new support models. The HPC market, as we know it today, along with many other industries, will be disrupted. It will take a lot of energy. From all of us.

Viva La Revolución!