SC09, Computing for a Changing World

Report, Riken

"*Never send a human to do a machine job*"

The SC09 Conference put together the best researches, scientists, industries and laboratories in the super-computing field in Portland Oregon last November. This conference represents the best opportunity to present and discuss the new advances in all areas of supercomputing like networking, hardware, software and I/O among many others.

The conference consisted of many events under the same roof, the Oregon Convention Center. The main event or the bounding one is the exhibition area where the principal industries and companies in the business as long with universities and laboratories can show their new discovers and research. There are many other events around this like the tutorials, masterworks, poster sessions, forums, etc.

The organization was neat and orderly developed, some unexpected changes of room happened but due to the number of attendances. The early registration could be done the day before the official start day to avoid long queues. Each participant was given a handbag with all the information about the conference plus magazines related to super computing, the batch and important guides like a floor-map or detailed schedules.

The freedom to swap between tutorials and presentations give an excellent chance to get a broad idea of more topics also.

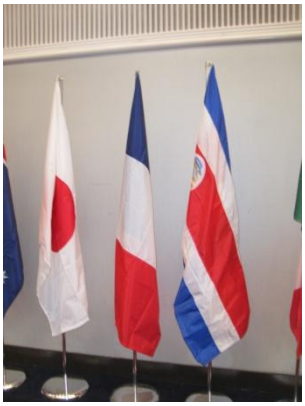

Figure No.1 Oregon Convention Center Hall

(Right Close-Up: Japan, France and Costa Rica Flags)

During the tutorials many different topics were presented from new high speed networks to I/O and from High Performance on GPU to the new OpenCL platform.

High Performance on GPU seemed to be a new important subject to analyze. Since its rapid introduction and adoption in the market the GPGPU technology seems to have come a long way. One of the tutorials explained quickly the GPU history which is one of the fastest adopted devices in supercomputing. The general purpose GPU computing or computing on the graphics card gives the ability to program the GPU as a general device meaning not only for graphics. The advance of this approach is that GPU GFLOPs are several orders of magnitude bigger than that of ordinary CPUs. Therefore the power inside a GPU can be used to solve numerical problems or any general problem now. If this performance is put together like in a cluster then this power is increased dramatically.

Therefore the interest of many new companies to explore this technology as an alternative to super-computing as dedicated clusters or heterogeneous CPU-GPU clusters.

In the High Performance Computing with GPU several ideas and techniques were discussed to improve even more the performance than a programmer can get from the GPUs. For instance fusing kernels as much as possible to minimize redundant access to memory or avoid if-clause statements. Another interesting GPU advantage presented at the tutorial is the initial-to-program time versus improve-benefit of using GPU compared to other technologies. Using results taken from MIT's 6.963 course they found that after a 10-hour studying and programming with CUDA they obtained a 330x speed up while on the time they obtained only a 9x speed up using C/SSE and even worst using Matlab. This clearly showed how CUDA gave easy access to GPU programming with great results.

In the exhibition-booth-part of the conference nVIDIA gave presentations about their new product named Fermi which is the next-generation GPU devices. These new devices promises even higher FLOPs than current Tesla GPUs by increasing the number of stream processors among other improvements. These cards will speed up double precision computing which up until now has been not even close as fast as single precision. They promise an 8-times speed up which is good but still not good enough is we compare the old single precision with new single precision.

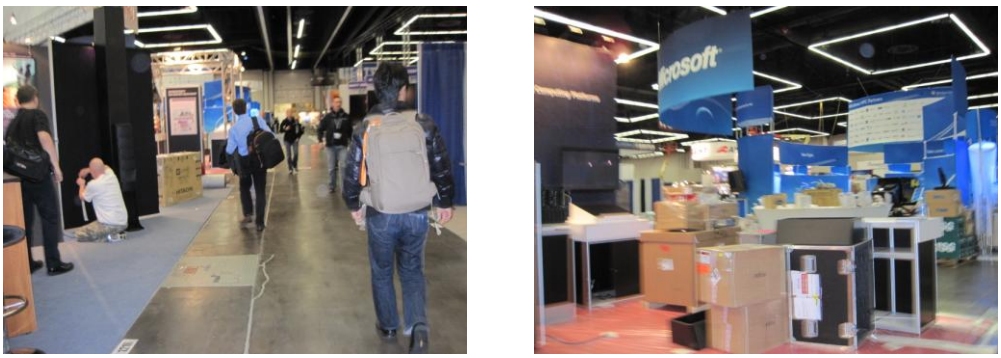

Figure No.2 Booth Preparation

A disadvantage using GPU computing however is that the cards cannot share data directly between them therefore one must use a bridge to send this information, usually the CPU. Clearly this is an undesired additional overhead that

would be better to avoid, nevertheless there is no way to avoid it with the present hardware. Fortunately some companies are already working a way around the problem to share information directly between GPUs.

Until now GPU computing has been possible thanks to CUDA from nVIDIA, which is a C extension language consisting of APIs designed to program at will the GPU. However this limits the portability of the program since other brands of cards are not able to compile or run general purpose programs. This solve this issue in a tutorial and other presentations a new standard was introduced called OpenCL, which is a multi-platform programming extension. Using OpenCL will give the possibility to port one program using any type of hardware. This will be clearly a great breakthrough in the current system however it brings up some questions like the performance obtained in a program optimized for a specific card in a competitor's card. Nevertheless as a joint effort to bring standardization and portability it must be noted as a to-watch innovation.

Figure No.3 Excerpt from Exposition Booth Floor

Different universities from Europe, USA, Japan and others were present at the conference presenting their research in climate, fluid dynamics, nuclear physics

and other main applications of super-computing. Super-Computing centers as well presented their innovations in hardware and applications.

A particularly interesting presentation was in the NASA booth regarding CFD applications. One of these applications is the support for vehicle assembly building, in which they modeled the fire scenario in case of an accident inside the vehicle rockets facility. Another application they are working on is related to the new space program and its launching vehicle. They have been developing numerical simulations to study the behavior of the atmosphere around the rocket during a launch as well as the movement and behavior of the fuel ejected from the booster. These researches require high accurate numerical methods creating a great complexity, different mesh sizes and unstructured grids sometimes. Additionally NASA uses its own rendering program which is outstanding to present results.

This bring us to visualization, this is another important application of super-computing. Most of the results nowdays obtained by super-computers are so vast and huge that they cannot be easily rendered. The data by itself might not be very helpful if there is not a reasonable way to interpret them therefore the importance of scientific rendering. However with the increasing data size visualizing is becoming itself in a super-computing challenge.

Very interesting efforts to do large-scale rendering were also presented at the conference. Some of the most impressive were renderings of millions of points sent over the network to a remote display where they were not only rendered but also manipulated. Research like that of Los Alamos National Laboratory named “Interactive Remote Large-Scale Data Visualization via Prioritized Multi-Resolution Streaming” presented different approaches to render large amounts of data. Some of the ideas were to render first a coarse resolution and produce a finer resolution with time or maybe if the user zoomed in a particular zone then refined that particular area. An interesting challenge discussed in the rendering presentations was the hardware versus data-set size issue, which basically means that we are

producing larger and larger amounts of data and with current displays is impossible to achieve that extremely high resolution. Therefore the solution must be found either improving the displays or taking advantage of this and for instance render only a part of the dataset and not the complete domain which anyway won't be correctly displayed.

The plenary speech on wednesday was given by Doctor Leroy Hood, one of the eleven scientists who, in 1985, initiated the Human Genome Project. His talk was under the Bio-Computing aspect of the conference. He explained in some detailed the work he has been doing plus his work for the future.

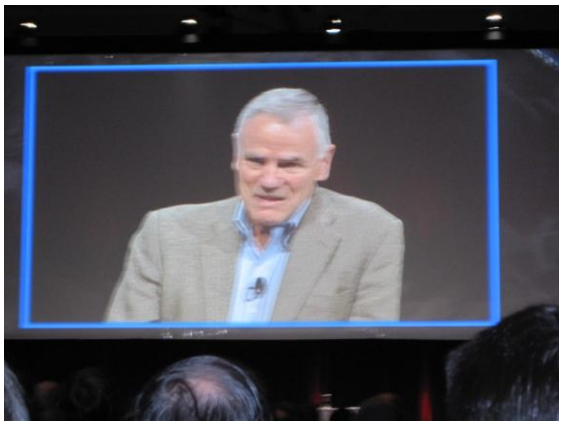

Figure No.4 Doctor Leroy Hood Plenary Speech at SC09

Clearly decoding gene information is a remarkably intensive task for any machine therefore the importance of super-computing to analyze these results. Nevertheless Doctor Hood is not only interested in this but currently he is interesting into cancer research. In this research he is interested into developing a cell-by-cell grow of the malign tissue as opposite to today-used approach where the cancer tissue is treated as a whole. In this kind of proposal we are talking about thousands and millions of millions of cells reproducing and interacting between each other, this is a new area of super-computing where all the resources will be needed to understand correctly the behavior of cancer grow and thus come up with a cure.

During the conference the new Top 500 list was released introducing Oak Ridge National Laboratory's super-computer called *Jaguar* as the new number one in the world. It has over 240.000 processors with a 1.75 Petaflop/s performance setting a new road in Peta-scale computing.

Figure No.5 Top500 November 2009 List

On thursday the main event was the keynote speech delivered by Nobel Prize Laurette and former USA Vice-President Al Gore. Al Gore was a key factor in the development of the internet back in the 70s and 80s while serving as a US senator and since that time he also consumed his time devoted to global climate activism. Known for his Oscar-winning documentary "An Inconvenient Truth" Al Gore presented an interesting speech trying to merge super-computing with climate change prevention.

Figure No. 6 Former Vice-President Al Gore giving Keynote Speech
(Photograph Credit: ©SC09)

Therefore the main point of his interesting talk was about the application of the super-computing infrastructure to climate modeling in order to understand and prevent natural disasters produced by pollution.

Finally one last remark but present throughout the conference was the Exa-Scale Super-Computing which is the next natural step for this field. After reaching and passing the Peta-Scale barrier all eyes are now in the next challenge.

However the next step might not be as easy as just ensemble more processors to increase the performance of a super-computing. There are several and new problems to solve before thinking about going into Exa-Scale. According to presentations during the conference, panels and reports some of the main issues to tackle are: lack of cooperation, the industries, research laboratories and universities are not really working together to achieve this goal. Therefore much resource is wasted instead of combining and sharing knowledge. Some suggestion to solve this problem is to create a global network to share information. Another interesting fact is that up until now scientists have enjoyed automatic improvement of their programs with each release of a new and more powerful processor however this is no longer the case and in order to achieve a better performance it is necessary to develop efficient code. This is going to be specially true for Exa-Scale computing since the program should be as efficient as possible not to waste the huge amount of hardware available. One sort of new issue raising with exa-scale computing is fault control and error prevention. With currents technologies the influence of a hardware failure in the code is not really critical but with the new hardware the presence of errors will grow and techniques for checking/restart and ecc for instance are going to become more and more important.

Finally maybe the two main issues when thinking on a exa-scale computing facility are not precisely how to find more powerful processors or higher speed networks but space and energy. In fact even current supercomputers require a large area to be ensemble in, in the future with exa-scale computers will require even larger areas what creates a big trouble with industries and laboratories.

Second and also as important as the previous issue is the energy. The energy to power up and keep running these monstrous machines will be very considerable and it is very important to analyze if the electrical system would support such machines.

In more relaxed terms, the conference took place in the city of Portland Oregon a very photogenic and peaceful city. People were warm and kind and it had a great environment. The last night the organization prepared a reception at the Portland Center for the Performing Arts that was very entertaining and relaxing.

Figure No.7 Portland City and the Portland Center for the Performing Arts

Once again I want to deeply thank Riken for giving me this great opportunity to explore the world of a Super-Computing conference as their reporter.

Marlon Arce Acuna