http://www.top500.org/blog/2009/05/20/top_trends_high_performance_computing

The Top Trends in High Performance Computing

Wed, 2009-05-20 21:32

Horst Gietl, Executive Consultant, Prometeus GmbH, Germany

Hans Meuer, ISC'09 General Chair, University of Mannheim & Prometeus GmbH, Germany

1. Introduction

If we calculate the performance predictions for the TOP500 list, then we will see a 100 Petaflops system most likely in the year 2016. In the past we had a performance increase within 11 years from Gigaflops (Cray2 in 1986), via Teraflops (Intel ASCI Red in 1997) up to the Petaflops (IBM Roadrunner in 2008) by a factor of 1000. Despite these performance improvements the HPC arena will more and more be defined indirectly via the highly compute-intensive applications. They are coming from a variety of areas, involving quantum mechanical physics, weather forecasting, climate research, molecular modeling (computing the structures and properties of chemical compounds, biological macromolecules, polymers, and crystals), physical simulations (such as simulation of airplanes in wind tunnels and research into nuclear fusion), cryptanalysis, and improved seismic processing for oil exploration for continued supply. For most of these applications detailed results may only be achieved with systems in the Petaflops range. And hopefully Exascale Systems will be seen first in 2019.

But what are the trends in the nearest future which will form the basis for all these performance predictions. The major challenges to all processor requirements for HPC systems now and in the future will be: low cost, low power consumption, availability of support for parallel programming, and efficient porting of existing codes.

2. Many-/Multi-Core Systems within large clusters

Recent trends in high-performance computing (HPC) systems have shown that future increases in performance, will only be achieved through increases in system scale, i.e., using a larger number of components and not by improvements in single-processor performance.

The fact that future single CPU-chips need higher Gigahertz rates, resulting in higher energy consumption, developing more heat and bringing single CPU-chip to their physical limits was the real stimulus for the multi-core processor technology. Going from Terascale to Petascale HPC systems and beyond means that the number of components (cores, interconnect, storage) within such a system will grow enormously. In the near future we will see clustered Multi-Core systems with core numbers in the range of one-hundred thousand to one million and more. It is obvious that these highly parallel systems will raise questions about parallel software development and especially to the fault tolerance and reliability of the hardware components.

But the real problem which faces every Multi-Core system is the limited bandwidth for the memory access. All cores on a Multi-Core chip are in competition for the memory access. Concerning the memory access, a good solution would be one Byte per Flop and Core. If this solution would be available even one hundred cores per chip would not be a problem. With Nehalem from Intel we will go into this direction by half a Byte per Flop and Core. And Intel recently displayed its currently-in-the-lab, future-oriented prototype 80 core chip with an expected computation performance of one teraflops.

Further developments would be dynamic Multi-Core chips for a variable adaptation of the hardware architecture to the requirements of a parallel program and even Multi-Core chips with vector extensions (again) are seen in the pipeline.

As a consequence to all this Multi-Core technology, parallel computing is clearly becoming a MUST. Parallel computation could dramatically increase the speed, efficiency and performance of HPC systems but going from 1 to more Central Processing Units (or Cores) might be not an easy task.

3. Parallel Computing is a MUST

The growing number of computational components within the hardware architecture means big efforts for the parallelization of application programs. The shift to the parallel paradigm will be a tough task because there are a lot of programs that have been developed under the premise of a single-core technology.

Parallel computer programs are more difficult to write than sequential ones, because concurrency issue will introduce different new classes of potential software bugs, of which race conditions are the most common. Communication and synchronization between the different subtasks are typically one of the greatest challenges to obtain good parallel program performance. It is a well known fact that parallelization tools are far behind the possibilities offered by the HPC hardware. Various programming techniques are already in use, like data locality; that means that the data are partitioned into blocks that fit into the CPU's local memory. The parallelization has to be the focus point for every new program development that should run on a Multi-Core system. And porting existing sequential industrial codes is still an open question.

The idea to place more and more cores on a single chip will only work if there is heavy investment into parallel programming; otherwise Supercomputers are not getting faster for a lot of applications. And as already said under point 2., one limiting factor for performance increase in the future will also be the memory bandwidth.

4. GPUs, Competition to Multi-Core and also an Add-On

A serious competitor for the multi-core CPU is represented by graphical processing units (GPUs), which are graphic cards used for scientific computing. There are four basic things about GPUs. They are fast and will get a lot faster. They are cheap, measured on a performance-per-dollar basis. They use less power than CPUs when compared on a performance-per-watt basis. But the fourth thing is their limitations.

GPUs are only good for tasks that perform some type of number crunching. The GPU was designed specifically to process graphics, and that means processing streams of data. Graphics chips may simply be seen as massive Multi-Cores, where in high end versions up to 800 (Shader-) units are running in parallel, within a GPU cluster there could be more than 3000.

The large potential of the GPU can be shown by the following numbers. The fastest graphics chips from AMD and Nvidia are already in the Teraflops range whereas normal Multi-Core chips are slowly touching this border.

The real problem with GPUs is that they may not be programmed as it is for usual x86-; Sparc- or Power-CPUs. That's the reason that Nvidia GPUs offer the support of the CUDA (compute unified device architecture) library that provides a set of user-level subroutines and allows the GPU to be programmed with standard C or Fortran without the need to use a graphics specific API.

Nevertheless with the Supercomputer »Tsubame« from the institute of Technology in Tokyo we have the first system in the TOP500 list that is running »Tesla«-Graphics-Chip from Nvidia. The system-cluster consists of 170 Tesla-S1070-systems resulting in 170 Teraflops –theoretically. In practice the system reaches 77,48 Teraflops, which means number 29 in the ranking of the TOP500 list (November 2008).

For the nearest future scenario of HPC systems we expect that the hardware architecture will be a combination of specialized CPU and GPU type cores.

But now to the most challenging problem for HPC is the energy consumption. In the future all (Chip manufacturer, data center manager) have to tackle the problems of energy-efficiency and energy consumption.

5. Green IT and energy consumption.

The energy consumption of the giants in the Teraflops range and beyond is enormous. Remember the hot summer in the USA where some of the supercomputer centers had to shut down their HPC facility because of a lack of power. It is well-known fact that the energy consumption of HPC data centers will double in the next 4 to 5 years, if the current trend continues. A straight forward extrapolation for Exaflops systems shows that they will be somewhere in the range of hundreds of Mega-Watt. That's one of the reasons that the managers of SC centers are critically observing further SC developments especially under the aspects of power supply. And you can be sure that at every supercomputer conference there will be session about energy consumption of HPC systems.

Frank Baetke from Hewlett-Packard brought the problem to a point. "In the past the classical SC was designed regardless of its energy consumption. If the clock rate of the Multi-Core CPUs will be reduced by 20% only, then the energy consumption is reduced to 50% compared to a system running at full clock speed."

HPC manufacturers and data centres will more concentrate themselves on energy efficiency. Therefore the development of HPC systems that reduces the energy consumption (50 to 70 percent of the power is normally used to cool the equipment) is absolutely necessary.

But perhaps we have already crossed the critical border of energy consumption in the HPC range and the further imminent growing energy consumption will be the limiting factor for future HPC applications. There is an open question how to solve the future performance requirements without increasing the CO_2 emissions and thus speeding up the global warming.

One idea is that the wasted heat of the processors is by no more blown into the air, but will be removed by a cooling liquid. In this way the liquid will be warmed up and is sent to a heat exchanger thus warming up offices and rooms. First environment tests are underway.

And the Green-It trend also has some implications for data storage systems.

6. Data volumes will explode.

The growth rate of data volumes is tremendous. With the large Multi-Core systems it is not unusual that Terabytes will be produced within hours (CERN).

Especially under the aspects of Green IT it is also necessary that the power consumption of the primary storage components will be reduced by a remarkable factor. In the next years we will see new developments in the area of solid state disk drives (SSDs). They have no mechanical components and in comparison to usual hard disk drives they are faster, more stable, can shift to very low power mode and clearly have a better performance-to-power ratio, which is one of their biggest advantages. If they are getting cheap enough they will be the best alternative to normal hard disk drives for installation in HPC systems under various aspects.

With all this in mind HPC is becoming more mainstream.

7. HPC will expand to small/midsized companies.

The flexible structure of the clustered Multi-Core systems, especially if they are based on Blade-technology offers a broad spectrum of HPC systems for installation at small / midsized and enterprise companies. It is obvious that data centres may start with small HPC systems at low prices and expand to larger systems, depending on their budget and application needs. In addition ISVs with their software packages are more and more jumping onto the parallel bandwagon, thus making it easier for companies to integrate HPC systems within their environment. And last but not least Microsoft is offering an alternative to Linux with their Windows HPC Server thus driving HPC adoption and provides supercomputer performance to companies that might otherwise not be able to afford it because of operating system issues.

A new buzzword is flying around “containers” filled with a totally equipped HPC systems. They are coming into the field of vision as a solution for the future and are designed to deliver compute and storage power exactly when it is needed. Clearly they will be an ultimate modular turn-key solution for “plug and play” that can add SC performance in a fraction of the time it would take to install a traditional data centre. The only question is what happens to the customers if the container will be picked up by a truck?

8. What’s on the horizon?

We will see systems that will set a new Flops/Watt Standard“. In 2011 the Sequoia-System developed and manufactured by IBM will be installed at LLNL and will start running in 2012. It will have 1,6 Millions of Power-Processors and 1,6 Pbytes of main memory leading to a peak performance of more than 20 Pflops and will be the number One in the Top500-List. From a technological point of view the system is further development of the BlueGene/P-Architecture, where chips with 16 cores (45nm-Technology) will be integrated with a newly developed switching-technology based on optical communication. As a whole Sequoia will only need 6 MW resulting in the extraordinary value of 3,000 MFflops/Watt which means an improvement of 7against BlueGene/P and an improvement of 5 against the Cell-based Roadrunner.

9. Cloud Computing

For 25 years, Sun has been guided by the vision that "the network is the computer." Cloud computing in 2009 at least could have the potential that this vision will become truer than ever. Let’s wait and see. At the International

Supercomputing Conference ISC'09 in Hamburg, 23 June – 26 June, www.isc09.org , the highlight session will be "HPC & Cloud Computing – Synergy or Competition ?", featuring speakers from Google, Amazon, Yahoo, Microsoft, IBM, HP and Sun Microsystems. In more than 4 hours these experts will present and discuss whether cloud computing will continue to impact the way IT infrastructure is designed and delivered to meet the varied needs of the Web, of business and especially of HPC users.