Ter@tec, Europe's first technical park dedicated to high-performance simulation.

© www.imaconcept.tv

On the road to

After the petaflops, the next challenge will be exaflops (performing a quintillion operations per second). There will be many obstacles to overcome.

Léo Gerat,
scientific journalist

The petaflops threshold (a quadrillion operations per second) was reached in 2008 by the IBM Roadrunner, eleven years after the teraflops threshold (1997: ASCI Red, Intel). Now scientists are setting their sights on a quintillion operations per second. Demand for this has already been expressed by governmental administrations such as the DOE (Department of Energy), in the United States, as well as certain scientific sectors such as climate modelling.

Exaflops within the next ten years? This might as well be infinity. How can this factor of a thousand be achieved? There are few voices offering an answer at this stage, but many more asking the questions. A report commissioned by Darpa[1], a team headed by Peter Kogge (University of Notre-Dame, Indiana), entitled Exascale Computing Study, set out four major 'challenges'. The first is the question of energy. Everybody knows it; this is the fundamental problem.

Roadrunner consumes 2.5 megawatts (MW), which already seems an achievement when you consider that Jaguar (Cray) is hot on its heels in terms of processing power, but consumes an almighty 7 MW. If the energy bill were to be multiplied by a thousand to achieve the exaflops level, we would be looking at gigawatts, which is roughly the output of a large power station. This is preposterous!

We might, at a push, consider a one-off computing installation consuming a tenth of a gigawatt. But 125 MW for an exaflops for example, would represent an increase in energy efficiency of a factor of 20 compared with Roadrunner, which is not to be sniffed at.

We really need to make serious headway in this area. This means first and foremost looking at the microchips that would supply the 'flops'. However, the cost of developing a new generation of microprocessors is now so great that supercomputer manufacturers can no longer envisage specially commissioning them for their own use.

According to William Jalby, team manager of the Prism laboratory's Arpa team at the University of Versailles-Saint-Quentin-en-Yvelines, *"this is a problem we are going to have to live with. Only mass-produced 'retail' components offer a satisfactory power-price ratio."*

Today's supercomputers use microprocessors designed for other uses. Not long ago, these were microchips designed for large computers, or servers. But more recently, we have started to look into more modest and cheaper microchips.

Enter the IBM BlueGene - number one in the Top 500 in 2004 - which was already using a modest microchip originally designed for embedded systems. The same was true for 2008's champion, Roadrunner, another IBM jewel, which this time uses a microchip designed for a games console, the Sony Playstation.

"The world of high-performance scientific computing has become accustomed to borrowing microchips which were not designed for it", states Jean-François Lavignon, R&D partnerships director at Bull. *"We saw a good example of this in France with this hybrid prototype that Bull built for Cines (National Computing Centre for Higher Education), that uses nVidia's Tesla T10 240-core microchip, designed for graphics cards."*

This is why it is often said that the next generation of supercomputers could be built using microchips designed for mobile consumer devices. At Santa Clara in California, a team of researchers at the Lawrence Livermore Laboratory and Stanford University found the

(1) **Defense Advanced Research Projects Agency (American).**

exaflops

© www.imaconcept.tv

The Ter@tec Campus at Bruyères-le-Châtel will have 1,000 people working in 15,000m² of laboratories and offices by 2011.

technology that should enable them to build an exaflopic computer, capable of simulating the earth's climate by fragmenting the atmosphere into 20 billion cubes.

Xtensa, a mobile microchip designed by Tensilica, has 32 cores and consumes only 0.09 W. According to these researcher's calculations, the processing power per watt obtained is four times what is offered by the microchip in IBM's BlueGene and 100 times higher than an Intel Core 2 microchip, designed for laptop computers.

The second challenge identified by Darpa is memory and storage. Feeding millions of starving cores will be no easy feat. New memory technology would be more than welcome. Steve Scott, technology director at Cray, is optimistic on this subject.

We're going to have to literally re-invent a new way of programming

As for long-term storage, flash memory (used in mobile devices) is expensive but very modest, and will probably play a major role. *"We can expect machines with a layer of flash memory built-in between the central memory and the disc memory in the future"*, claims Steve Scott. This opinion is shared by Steve Pawlowski, technology director at Intel, a major player in the flash memory market.

The third gamble identified by the Darpa is entitled "concurrence and location". *"We are going to have to learn how to distribute the application over a billion cores"*, says Willian Jalby. *"This is not too much of a problem for those algorithms that work frequently with local data. But what about the others…?"* Franck Cappello, project manager of the Grand Large project at Inria Saclay, recalls that *"ten years ago, nobody would have predicted that multi-core technology would become a vital building block for petaflops. Now, we do not know what the building blocks for exaflops will be, or whether multi-core technology will play a positive role. It may not even be all that easy to exploit."* Because of this, we think that we will have to literally re-invent a new way of programming. We may even see some strange practices emerging. As William Jalby explains *"for example, it is sometimes better to re-compute the data that we need immediately, than to wait until it reaches a distant microchip."*

Energy management itself would also have a great influence on programming. Serge Petiton, head of the MAP team at the Lille Fundamental Computing Laboratory explains that "while the results from two parallel computations are needed in order to do something else, it is sometimes more logical to slow down the faster of the two, to save watts."

The fourth and final challenge of exascale identified by Darpa is reliability. According to Jean-François Lavignon *"this is a key question, one which will force us to re-invent the way in which we design computers."* *"The complexity of these machines is such that the average time between two failures will probably be hours, if not worse"*, claims Jean-Pierre Panziera, engineer at Silicon Graphics (SGI).

Franck Cappello feels that "failures will become normal events. We must therefore invent technologies which play them down, allowing us to continue running each software program and therefore applications, while we carry out repairs locally."

© www.imaconcept.tv

Ter@tec, architect's drawing.