

Monday, March 29, 2004

Tera Computing

In February 2004, **Pat Gelsinger** - the CTO of **Intel** - presented some great inputs concerning the future developments in the field of processor during the last **Intel Developer Forum**.

Do we really need the increasing power of processors? - yes

Based on the experience made in the past, the increase in processor power is always used by / surpassed by innovation and new technologies. The offer is somewhat creating the demand, as with Microsoft products.

Incredible volume of digital data

The volume of digital data is already today surpassing the volume of non-digital data (study from Berkeley in 2003). Almost all new data are digital ones. Some figures to illustrate this:

- 1 petabytes = 1 PB = 10^{15} bytes = 1'000 TB = 1'000'000 GB
- 1 exabytes = 1 EB = 1'000 PB
- Internet represents today about 530 PB
- The WWW itself: 170 PB
- For saving one year of telephone calls, you need 17 EB
- For saving your personal lifetime medical record, you need about 4 EB

Not enough?

Personally, I have:

- More than 8'000 digital pictures and more than 80 GB of personal digital movies (in 2 years). And I'm sure, a lot of people have really more than that!!
- I have at home more than 630 GB HD.

=> This huge amount of digital data - "Era of Tera" - requests high-efficient IT computing.

Architecture of high-efficient computing

Till today, the existing systems are constrained to use different algorithms and architectures to be able to be "efficient". On the opposite, the new requests due to the "Era of Tera" can be defined by three general characteristics:

- Teraflops of processing
- High data bandwidth
- Efficient execution and / or adaptation to variable workload requirements

=> These new common characteristics will allow a common definition of one new unique architecture which will be able to compute the Tera-requests.

Performance development

How can we define performance in the field of processor?


$$\text{Performance} = \text{Speed} + \text{Architectural Efficiency}$$

In the past, there was a clear concentration on speed increase, although some architectural improvements also appeared (hyper-threading, MMX technology, etc.). Some major obstacles are appearing now with frequencies increases and size of transistors decreases:

- power will be a limiting factor
- memory latency (the difference between the frequency speed of memory and processor is permanently increasing)
- scalar performance is more and more an issue: instructions per clock trend down with the increase of frequency

Consequently:

- performance increase has mainly to come from architectural innovations (rather than frequency increase)
- or disruptive change in the processor technology (eg.: **photonic computing**)

Conclusion

"A massively multicore architecture in which each core has multiple threads of execution with minimal memory latency, resistance-capacitance interconnect delay, and controlled thermal activity is needed to deliver teraflop performance".