

GPGPU: hype or revolution?

Gilles Civario – ICHEC

gilles.civario@ichec.ie

GPGPU?

- General Purpose Graphic Processing Unit
 - Basically a graphic card, used for computing
 - Seen as an accelerator: you still need the CPU
- Needs some special programming paradigms
 - Nvidia: CUDA
 - AMD/ATI: Brook+
 - Intel: Ct, but well, you know...
 - And for portability: OpenCL, HMPP

Faster than a CPU?

- Actually NO
 - A Nvidia Tesla card is “only” 1.44GHz
- But
 - 240 SP cores FMA + ADD capable: 1 TFlop/s
 - 30 DP cores FMA capable: 86 GFlop/s
 - 4GB of GDDR3 onboard: 102 GB/s
- Unity is strength

What can it do?

- A lot of things
 - Computing
 - Computing
 - Computing
- But not everything
 - Running the code
 - Communicating with other processes

Will it be successful?

- Based on commodity hardware
 - Just like clusters vs. supercomputers
- Much more powerful than a CPU
 - More cost-effective
 - More power-effective
- But not a CPU!
 - It should stay highly specialised for being successful
- Probably the only path toward Exascale

Portability?

- Hands-up those who feel concerned about portability...
 - I do!
- But is there any actual alternative to Brook+ or CUDA?
 - Maybe
 - HMPP
 - OpenCL

What about my code?

- It depends:
 - Vector-like computation: OK
 - Computation intensive on small data: OK
 - Random memory accesses: NO
 - Lots of branching: NO
 - BLAS, Lapack, FFT: maybe
 - Mixture of all this in small peaces: unlikely

Main current constrains

- Parallel algorithm
 - All is about parallel treatment carried out by tons of threads
- C language
 - For CUDA, Brook+ and OpenCL
 - But HMPP and PGI/CUDA allow Fortran
- Preferably SP
 - But DP may be also worth it

Performance pitfalls

- Data transfers between CPU and GPU memories
 - Similar to fast MPI communications
- Memory latency
 - No cache...
- Thread divergence
 - Similar payload as executing both branches

Main development areas

- Computational fluid dynamics
- Linear algebra
- Molecular dynamics
- Weather forecasting / climate modeling
- Life science
- Signal processing
- ...
- Image processing

In ICHEC

- Stoney: 12 GPGPU nodes
 - Not part of the national service
- Developments
 - Optimisation tool: JASM
 - MD: Quantum Espresso, DL_poly
 - Climate modeling: WRF
 - Technology watch and consultancy

What next?

- Nvidia Fermi (H1 2010)
 - ECC protected memory
 - Massive DP performances
 - Memory cache hierarchy
 - Multiple schedulers
 - Unified memory address space
- AMD / ATI Cypress (Now)
- Intel TeraScala (2011-2012)

Conclusion

- GPGPU are here, and that's for long
- A lot of issues still opened
 - Programming paradigms
 - Code portability
- Big room for improvement
- It may be worth trying
- Ask ICHEC

Thank you

Questions?