Linear algebra on multi-core and heterogeneous architectures

Web page:
`http://www.math.ethz.ch/~kressner/gpucomp.php`

Daniel Kressner

Harvesting the supercomputer below your desktop

- Most processors sold today have $\geq$ 2 cores (some have 3 or 4, few have 6 or more).
- Core i7-965, one of Intel's flagships, has 4 physical cores and theoretically attains 51 DP-Gflops or 102 SP-Gflops.
- NVIDIA's current flagship GeForce GTX 285 has 240 stream processors and theoretically attains 89 DP-Gflops or 1063 SP-Gflops.
- Multi-core + GPU computing has (almost) become main stream (right fig: March 2009 cover of popular German computer magazine ct).

Harvesting the supercomputer below your desktop

Motherboard with 6 GPUs
(3× GTX 295 = 6× GT 200)

	DP	SP
Core i7	51	102
GT 200	75	894
GT 200	75	894
GT 200	75	894
GT 200	75	894
GT 200	75	894
GT 200	75	894
Σ	501	5466

- 5466 SP-Gflops is 3.7% of DP peak performance of France's most powerful computer (Jade at GENCI-CINES, uses 3072 Intel Xeon Quadcore Procs).
- Price is about 3000 Euros.

Focus of these lectures: Heterogeneous CPU+GPU computing for linear algebra operations.

Outline

- Single-core processors
  - Parallelism within single cores.
  - Memory hierarchies.
  - Matrix-vector and matrix-matrix multiplies.
  - Case study: QR factorization.
- Multi-core processors
  - Programming multi-core CPUs
    - PThreads
    - OpenMP
  - Multi-threaded BLAS
  - Case study: QR factorization.
- General purpose GPU computing
  - Programming GPUs – CUDA
  - Matrix-matrix multiplies.
  - Case study: QR factorization
  - Case study: symmetric eigenvalue problems