Filippo Bistaffa

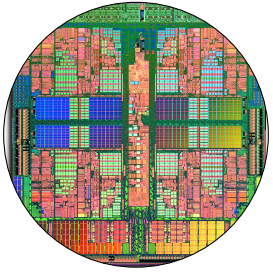
1st Crash Course on Parallelization

All examples and slides available @
https://github.com/filippobistaffa/phd-day-parallelization
(clone with `--recursive`)

14 November 2023

Types of Parallelization

Cluster

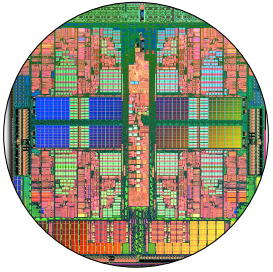


Multi-Core

GPU

Types of Parallelization

Cluster

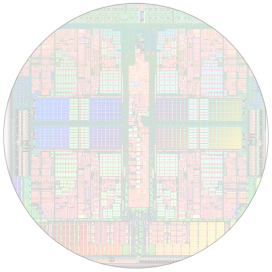


Multi-Core

GPU

Types of Parallelization

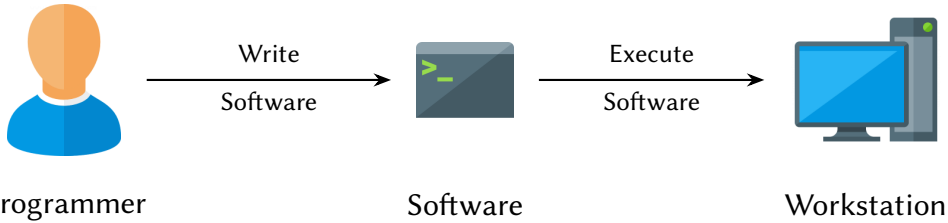
Cluster



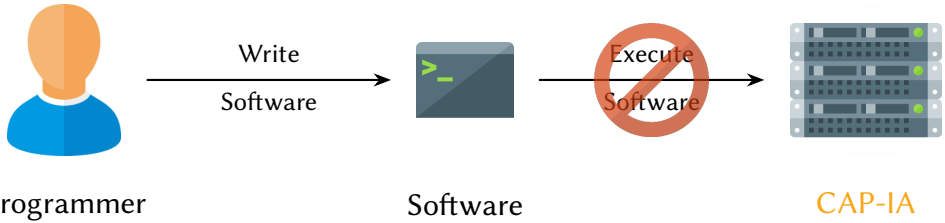
Multi-Core

GPU

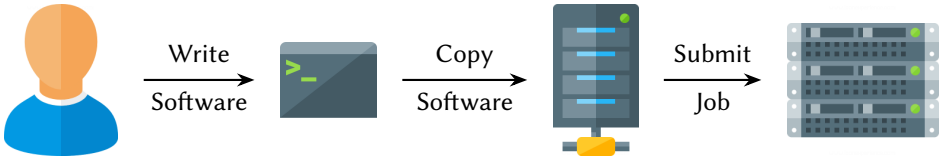
Local Software Execution



Cluster Software Execution



Cluster Software Execution

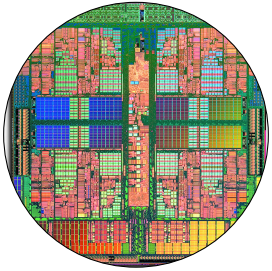


Today's Menu

- Python hello world!
  `cluster-parallel/hello.sh`
- Compile `llama.cpp`, a C++ LLM framework
  `cluster-parallel/compile-llama.sh`
- Download the "Vicuna-13B" LLM for `llama.cpp` on `beegfs`
  `cluster-parallel/download-model-beegfs.sh`
- Generate a course description with `llama.cpp`
  `cluster-parallel/description.sh`
- Compile `llama.cpp` with GPU support
  `cluster-parallel/compile-llama-gpu.sh`
- Generate a course description with `llama.cpp` with GPU support
  `cluster-parallel/description-gpu.sh`

Types of Parallelization

Cluster



Multi-Core

GPU

Today's Menu

- Setup CMake to compile C++ examples
  `cpu-parallel/CMakeLists.txt`
- Basic sequential `for` loop
  `cpu-parallel/basic-loop.cpp`
- Basic parallel `for` loop with OpenMP
  `cpu-parallel/basic-loop-openmp.cpp`
- Static *vs* dynamic scheduling
  `cpu-parallel/dynamic-scheduling.cpp`
- Parallel sum-reduction with `reduction`
  `cpu-parallel/parallel-reduction.cpp`
- Parallelized *vs* un-parallelized sections (Amdahl's law)
  `cpu-parallel/amdahl-law.cpp`

References

CAP-IA: https://informatica.doc.iiia.csic.es/cap-ia

CESGA: https://cesga-docs.gitlab.io/ft3-user-guide/overview.html

Slurm Workload Manager: https://slurm.schedmd.com/quickstart.html

Mattson *et al*: The OpenMP Common Core (Making OpenMP Simple Again)

Farber: CUDA Application Design and Development

(absolutely do not use https://www.libgen.is to download these books)