High-Performance Inverse Modeling with Reverse Monte Carlo Simulations

Abhinav Sarje[1]

Xiaoye Sherry Li | Alexander Hexemer

[1]Computational Research | Advanced Light Source

Lawrence Berkeley National Laboratory

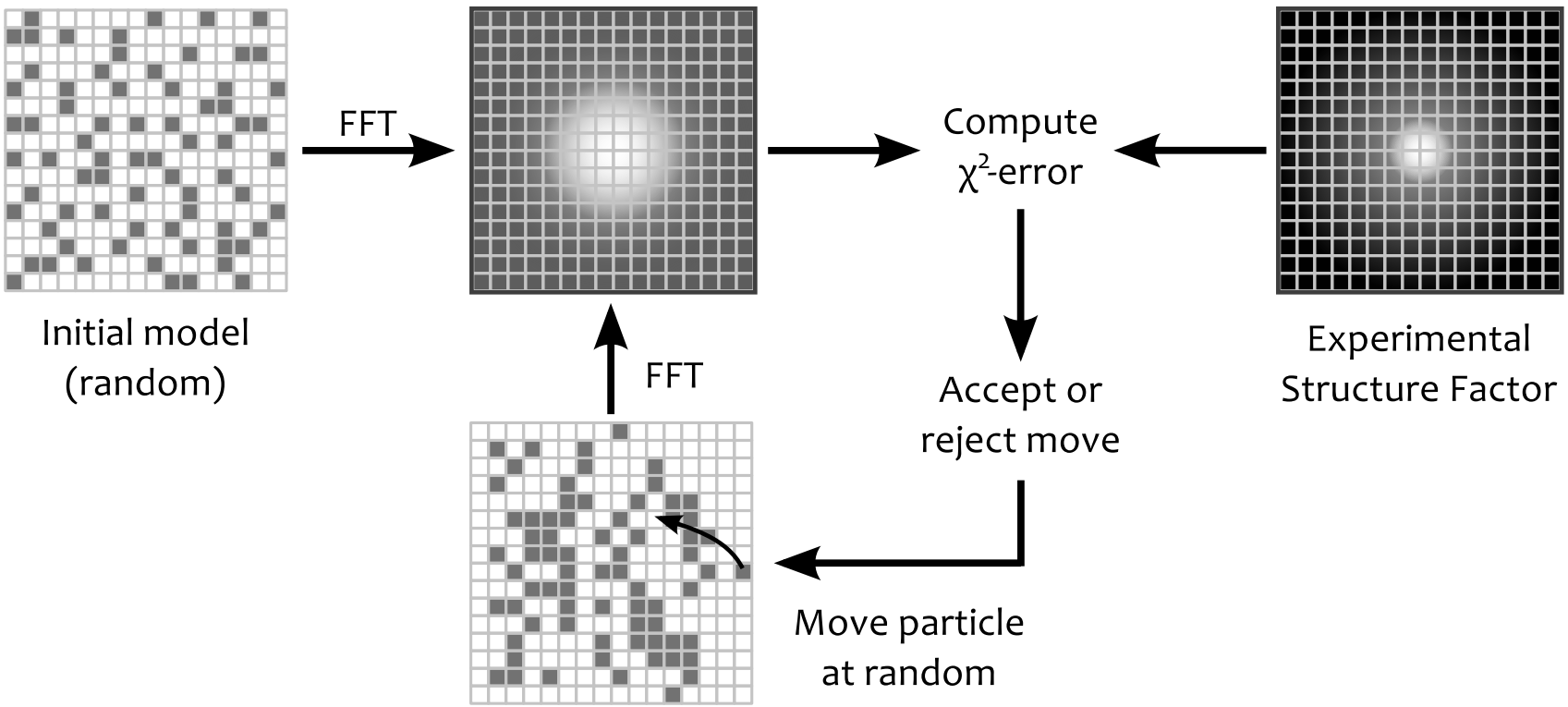
09.10.14

International Conference for Parallel Processing 2014, Minneapolis

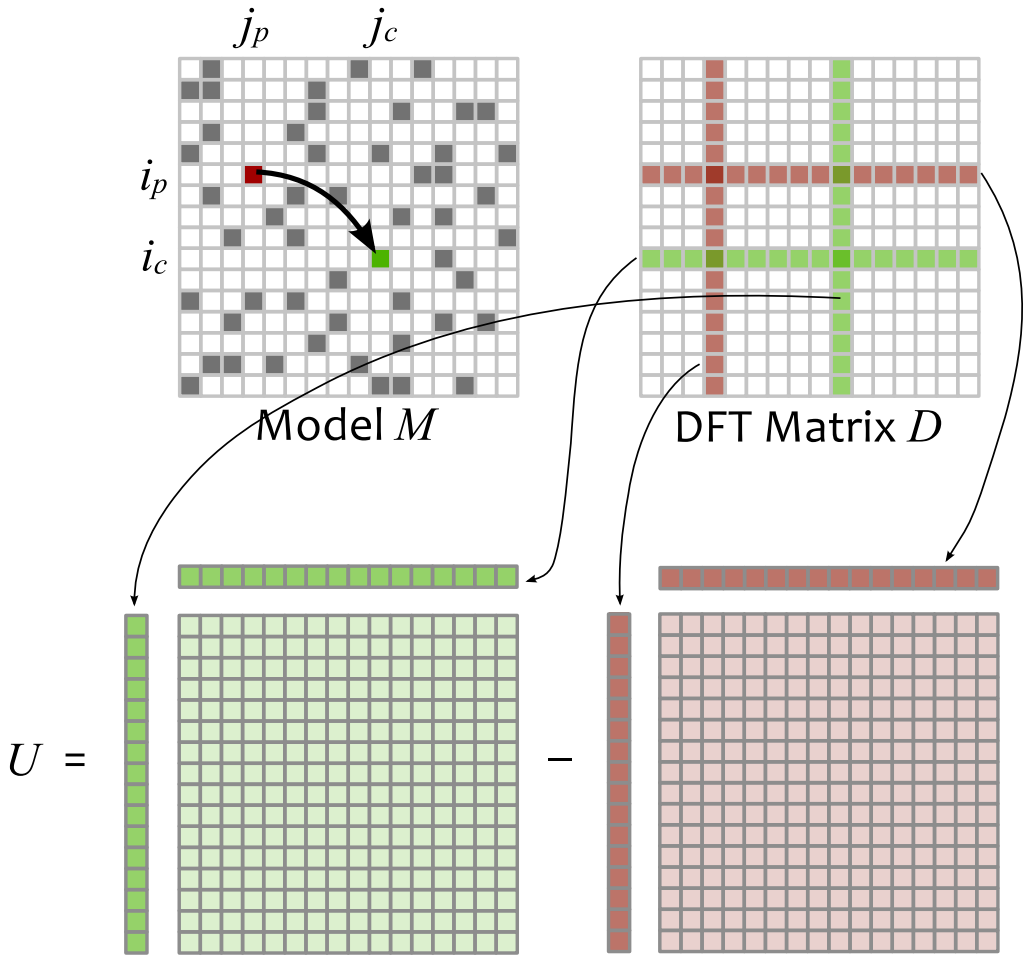
Background and Motivation

- **X-ray scattering** is an important tool for scientists to probe structural properties of materials at nano-scales.
- **SAXS, Small Angle X-ray scattering**, is a widely used type.
- SAXS is primarily used for **non-crystalline/amorphous materials**.
- Data challenge: Currently about **25-50 TB / month** of raw data from ONE beamline, and growing nearly exponentially. A Synchrotron has 10s-100 beamlines.
- Beamline scientists have not had HPC resources available. Need to bridge the gap.
- Many materials are too complex to use sophisticated modeling techniques.
- **Reverse Monte Carlo** simulation works well with SAXS data.

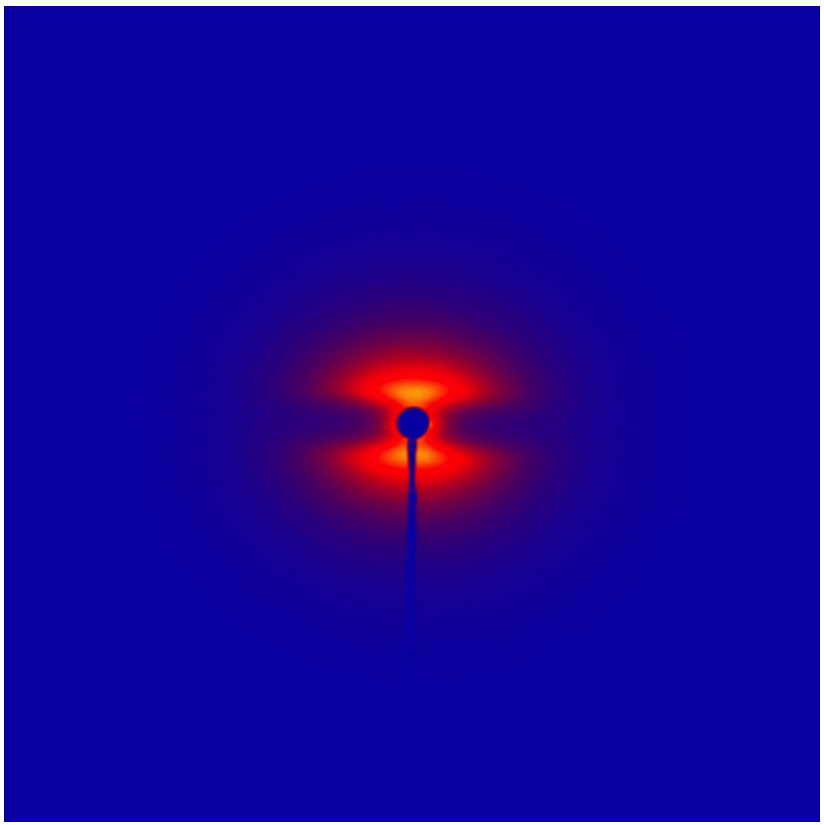
The General RMC Modeling Algorithm

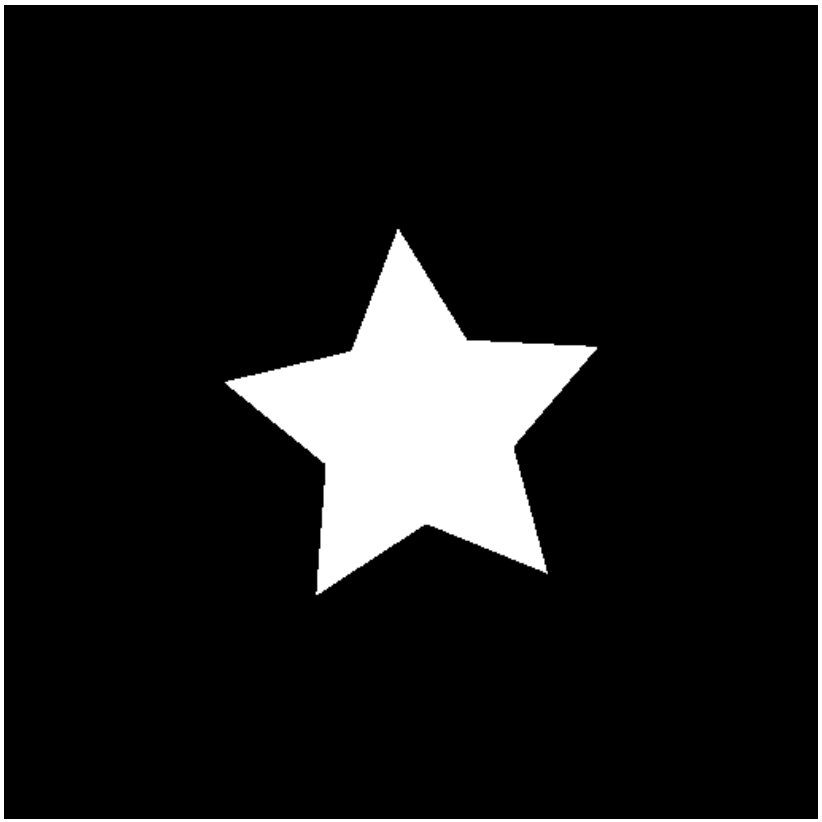


Faster Updates of Fourier Transform

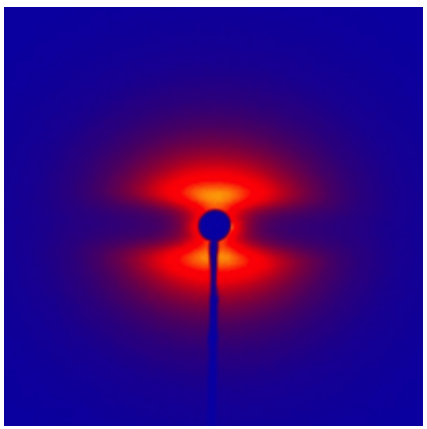


A Scaled RMC Modeling Algorithm

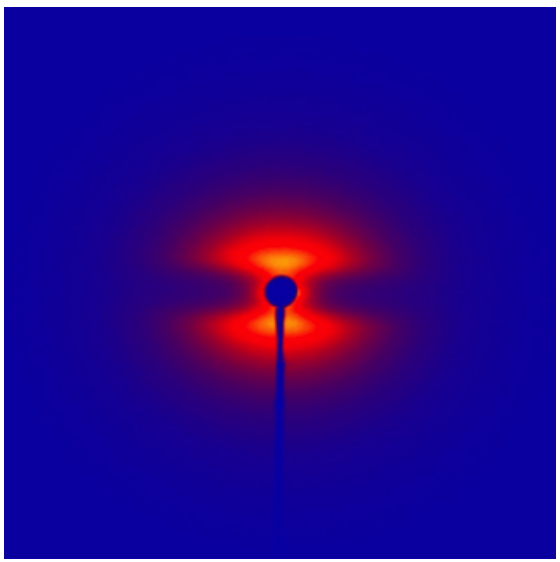




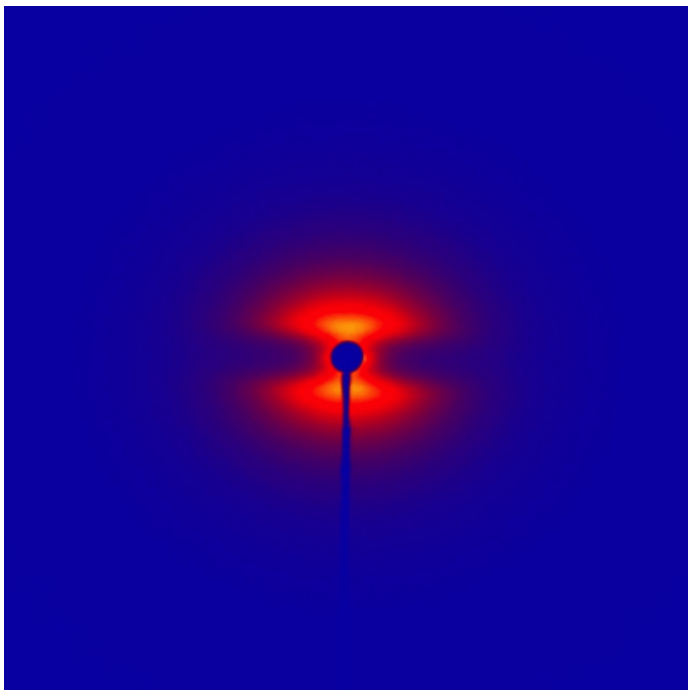
A Scaled RMC Modeling Algorithm



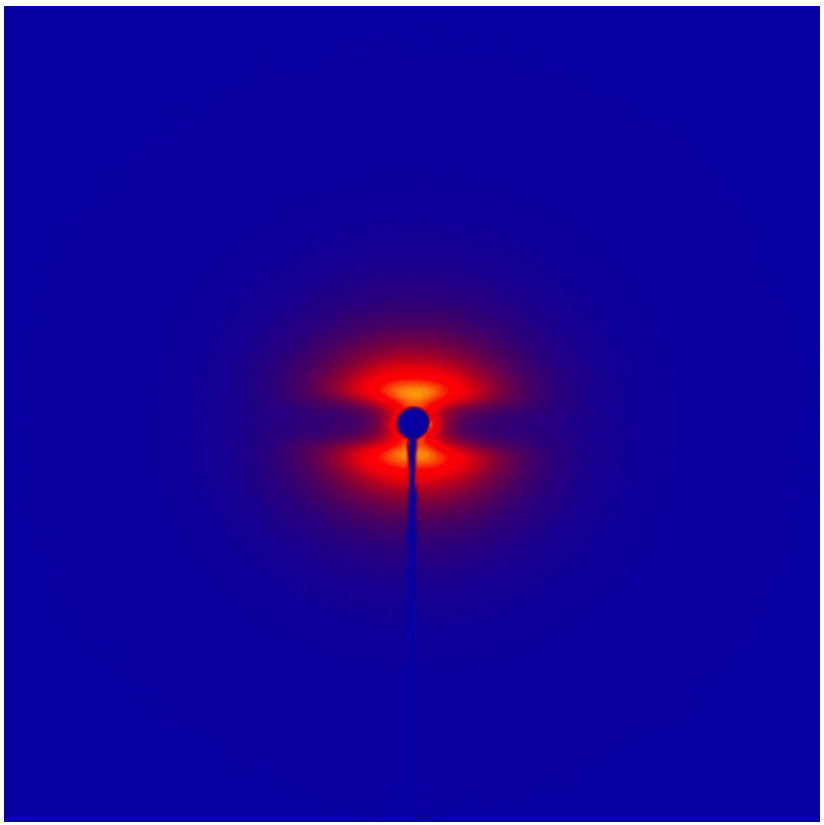
A Scaled RMC Modeling Algorithm



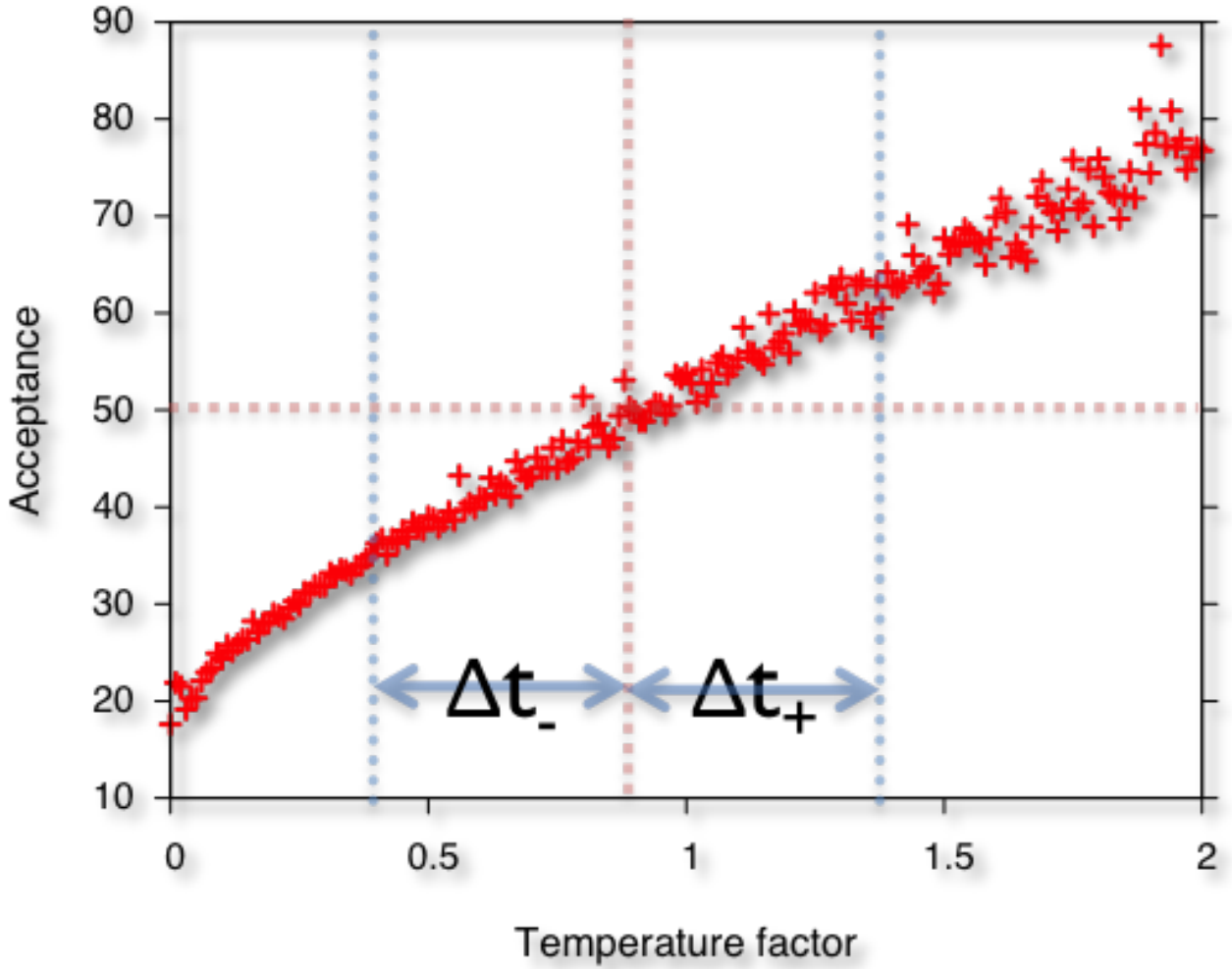
A Scaled RMC Modeling Algorithm



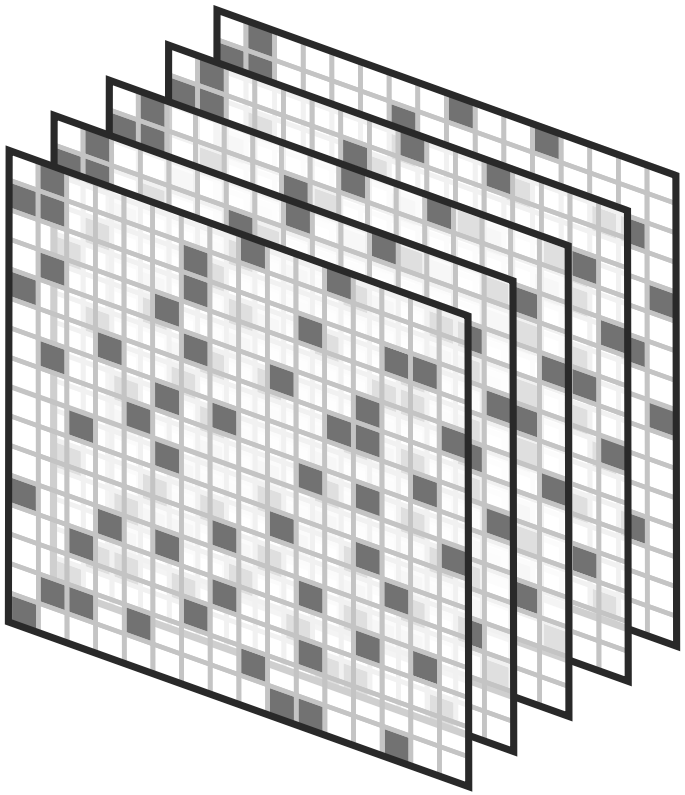
A Scaled RMC Modeling Algorithm



Autotuned Model Temperature

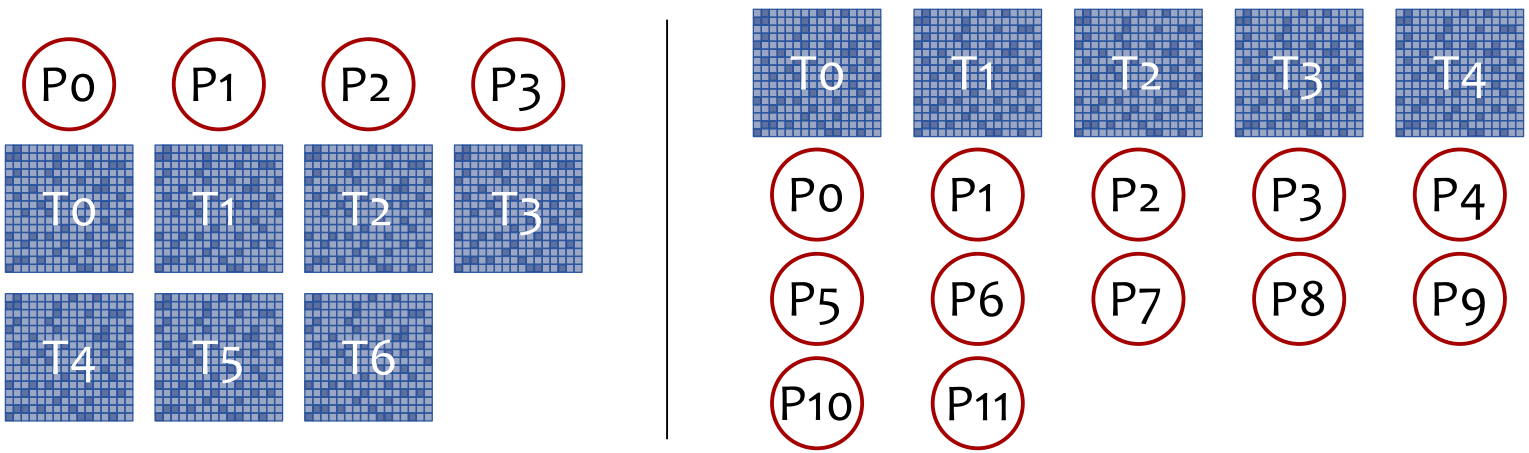


Distributed-Memory Parallelization over Multiple Tiles

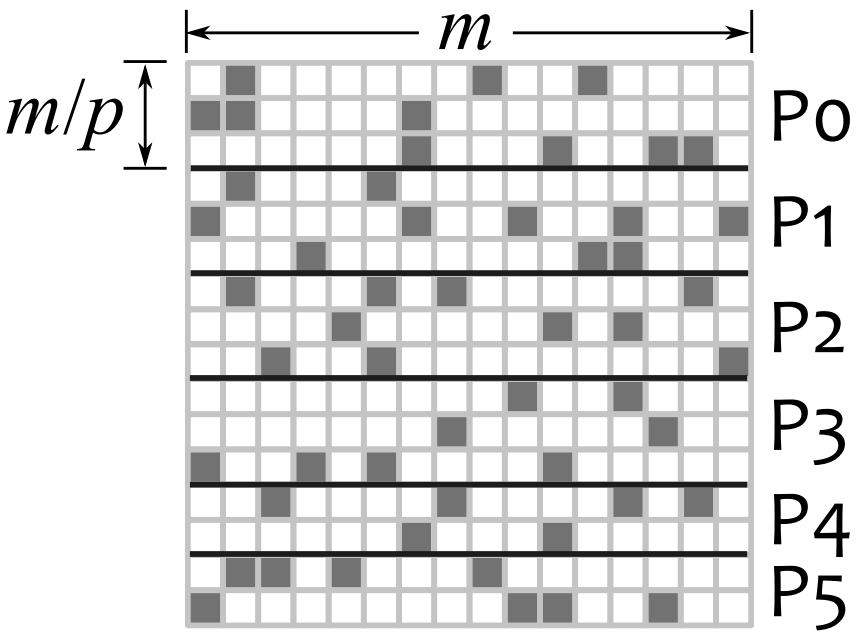


- MPI for distributed memory level.

Distributed-Memory Parallelization over Multiple Tiles



Distributed-Memory Parallelization of a Single Tile

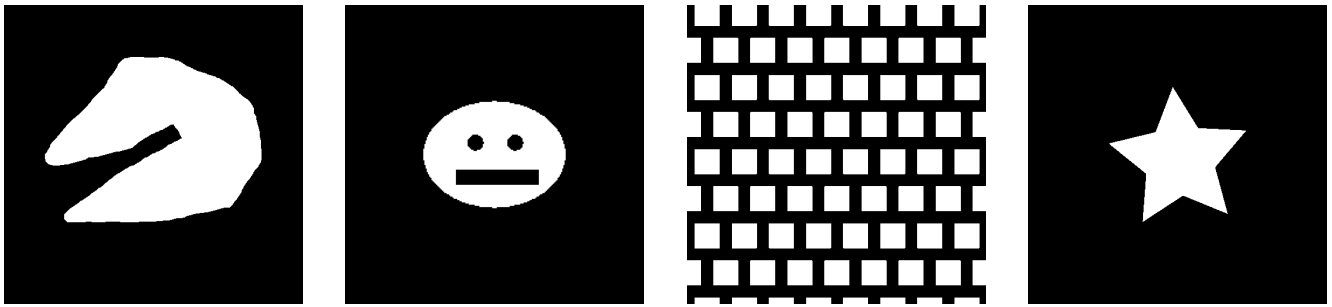


Subtile Implementations

- Two main types of kernels:
  1. Data parallel.
  2. Reduction.
- Multicore CPU implementation with OpenMP.
- Graphics Processor (GPU) implementation with Nvidia CUDA.

Validation with Model Reconstruction

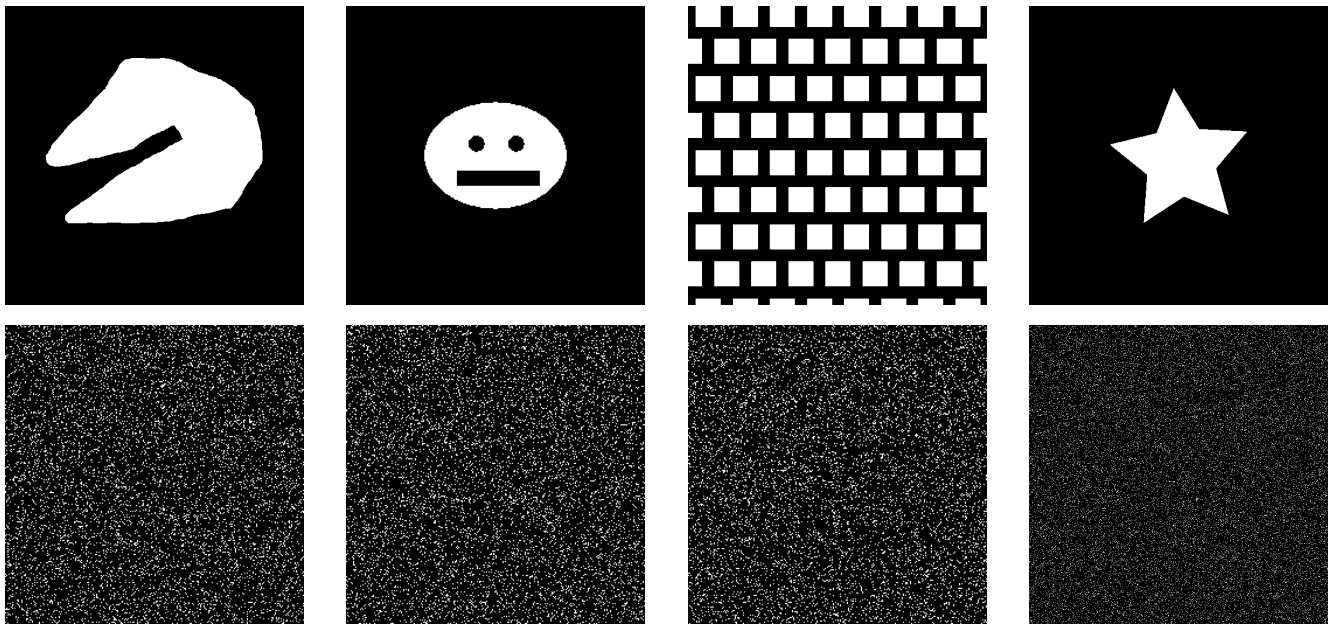
Actual Models



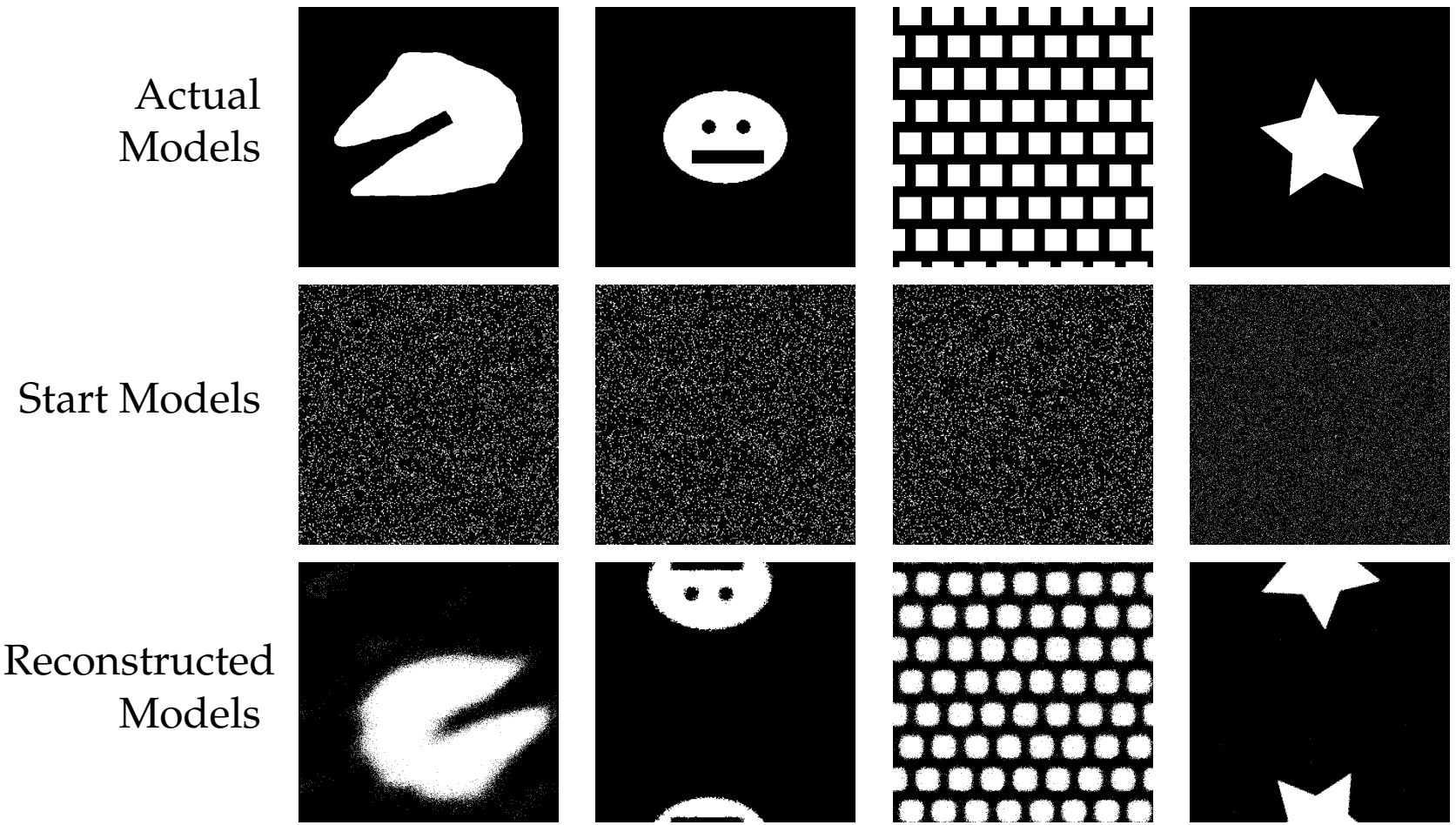
Validation with Model Reconstruction

Actual Models

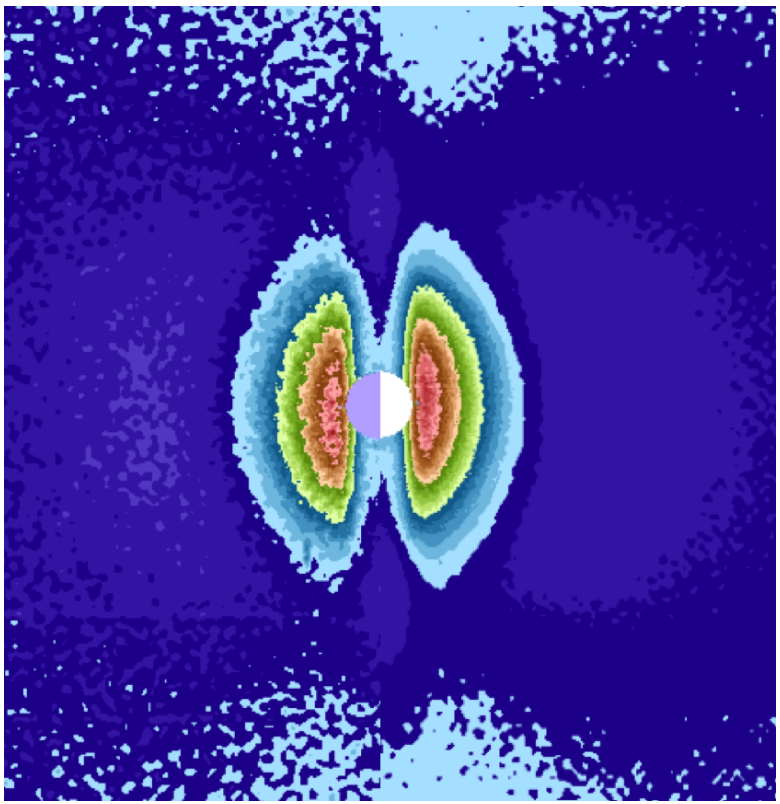
Start Models



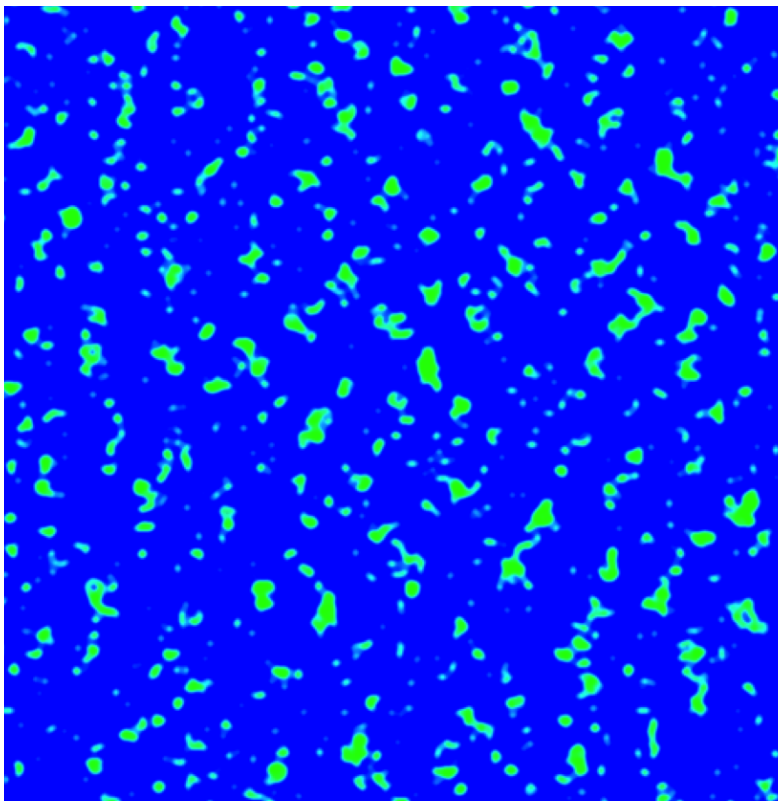
Validation with Model Reconstruction



Model Reconstruction



Simulated and Experimental Patterns



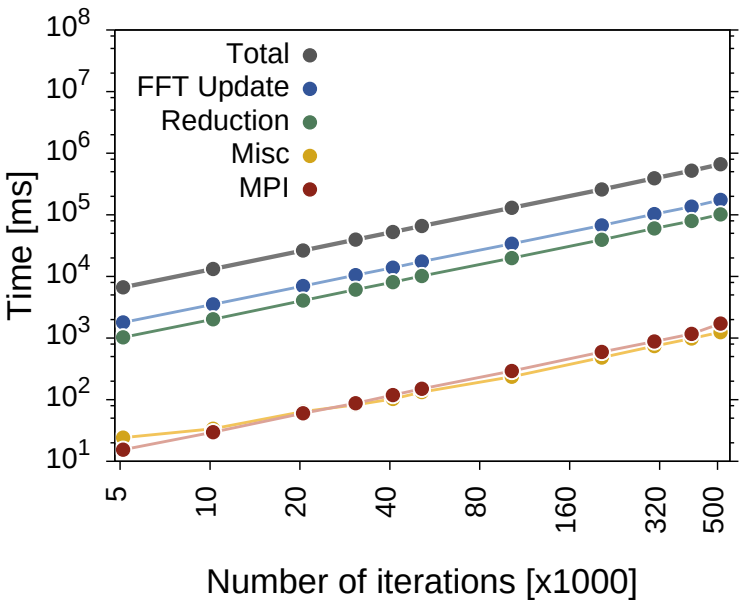
Computed Model

Experimental Performance: Environment

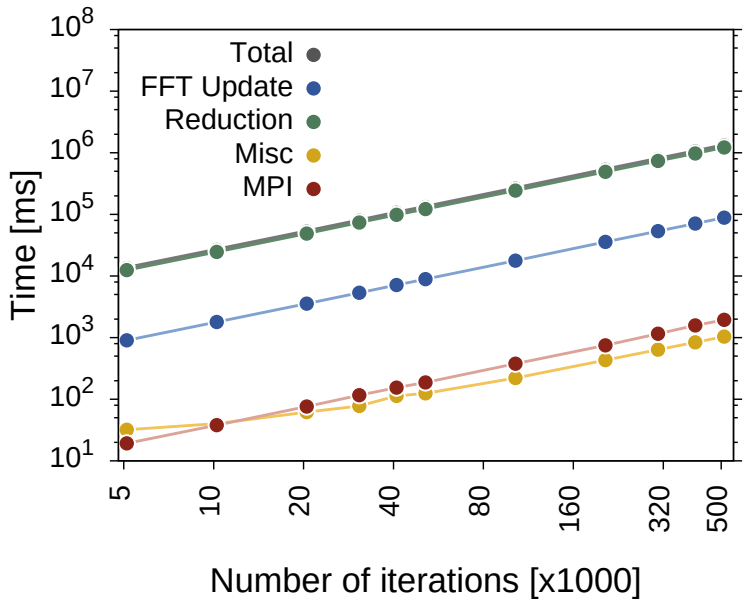
- All computations are in double precision.

1. **Single node platform:**
   - Dual-socket 2.2 GHz 8-core Intel Xeon E5-2660 (Sandy Bridge)
   - 16 cores and 32 hardware threads with 2 NUMA regions, and 64 GB RAM
   - Nvidia K20X (Kepler) graphics processor.

Experimental Performance: Number of iterations



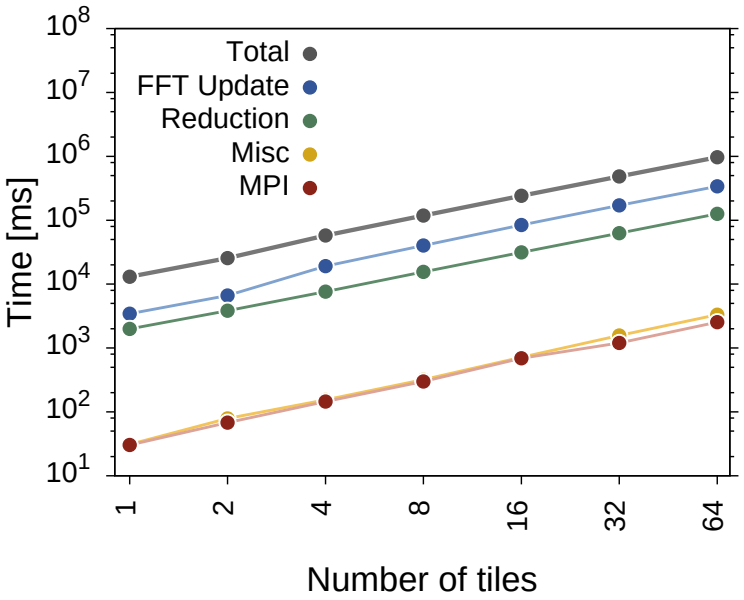
Multicore CPU



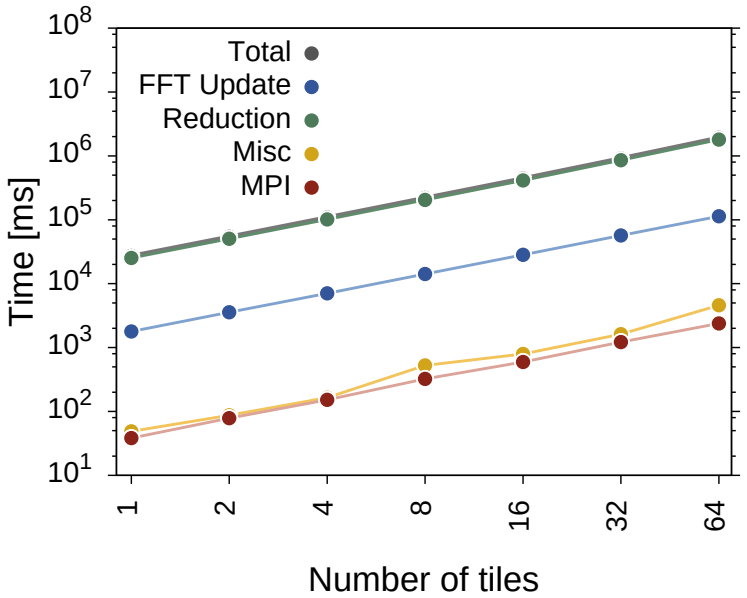
Kepler GPU

- Tile size of 512×512.

Experimental Performance: Scaling with number of tiles



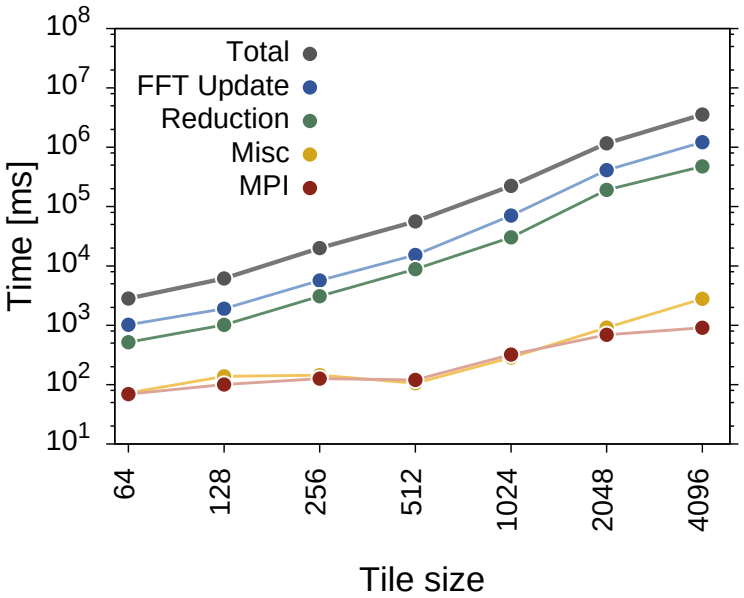
Multicore CPU



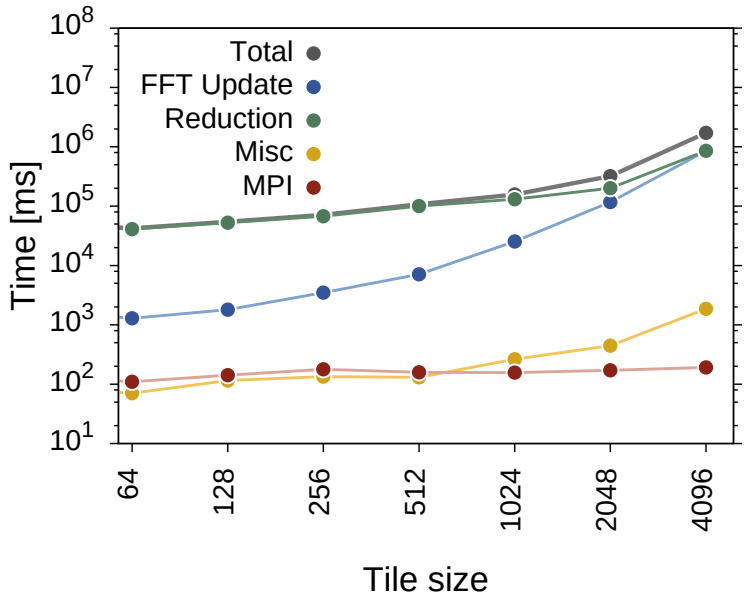
Kepler GPU

- Total of 10,000 iterations.

Experimental Performance: Scaling with tile size



Multicore CPU



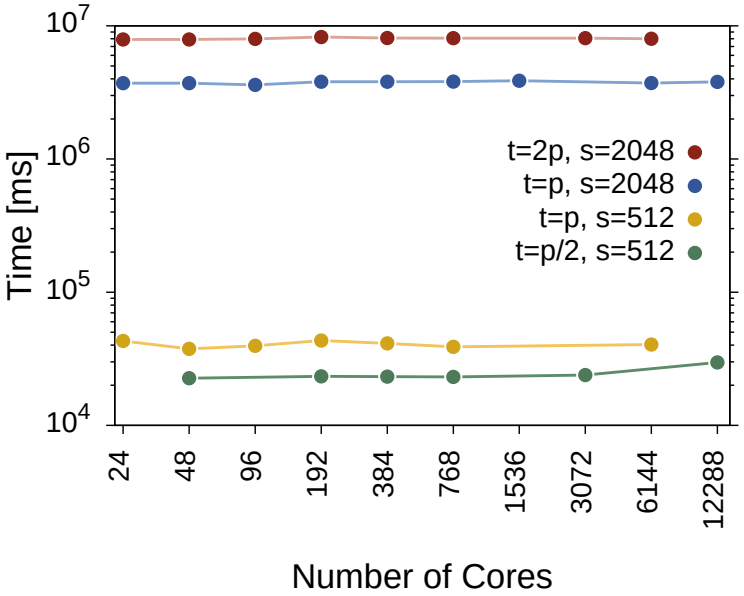
Kepler GPU

- Total of 40,000 iterations.

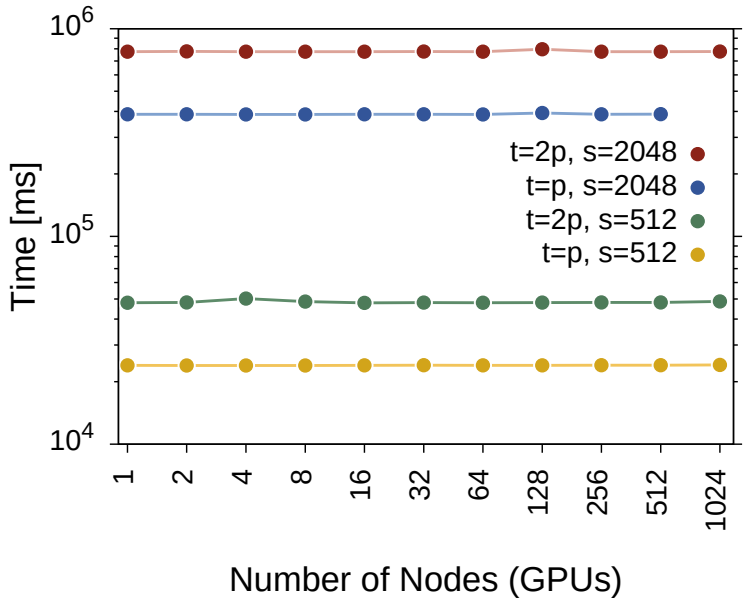
Experimental Performance: Environment

1. **Multicore CPU cluster:**
   - Cray XE6 supercomputer ('*Hopper*' at NERSC/LBNL)
   - Each node is a dual-socket 2.1 GHz 12-core AMD MagnyCours processors
   - 24 cores per node with 4 NUMA domains, and 32 GB RAM
2. **GPU cluster:**
   - Cray XK7 supercomputer ('*Titan*' at OLCF/ORNL)
   - Each node is a 2.2 GHz 16-core AMD Opteron 6274 (Interlagos)
   - 16 cores per node with 2 NUMA domains, and 32 GB RAM
   - One Nvidia K20X (Kepler) GPU on each node.

Experimental Performance: Weak scaling

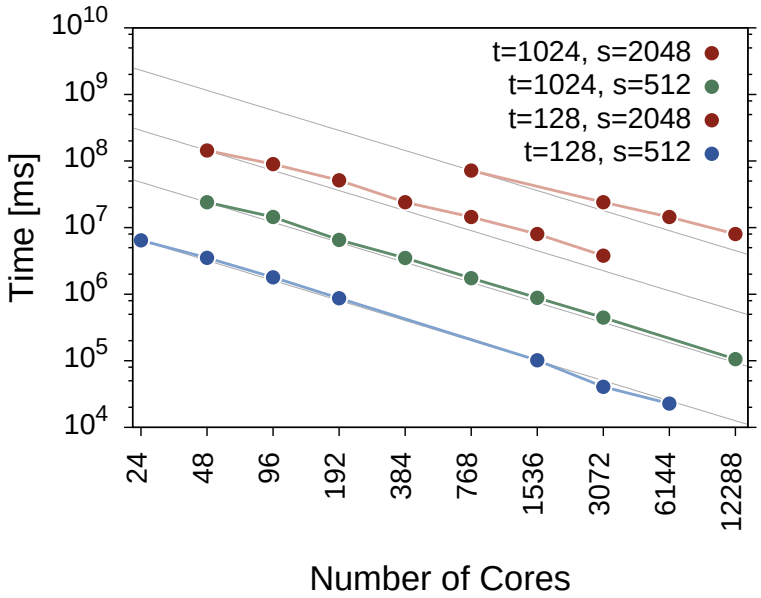


Hopper

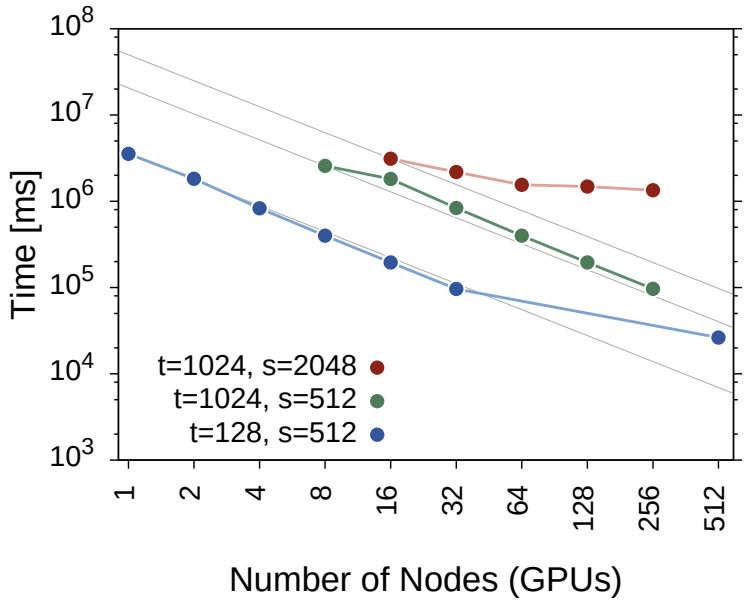


Titan

Experimental Performance: Strong scaling



Hopper



Titan

Conclusions

1. Synchrotron **beamlines need HPC** for data processing to be efficient.
2. Monte-Carlo and Reverse Monte-Carlo simulations are highly used methods in scientific computing, and same parallelization techniques can be applied.
3. **RMC works** when all other sophisticated methods fail.
4. Reduction operations are nearly one order of magnitude slower on GPUs compared to multicore CPUs. Better caching and less synchronizations on CPUs are two primary factors.
5. Even with large number of reduction kernels in the application, **at scale the overall performance on GPUs is about an order of magnitude better than on multicore CPUs**!

Future Work

1. Non-binary model reconstruction.
2. 3-D structures reconstruction.
3. Time-series fitting.
4. Realtime?

Acknowledgements

- Supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and DOE Early Career Research grant awarded to Alexander Hexemer.
- Resource usage at *National Energy Research Scientific Computing Center* (NERSC) at the Lawrence Berkeley National Laboratory, supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231.
- Resource usage at *Oak Ridge Leadership Computing Facility* (OLCF) at the Oak Ridge National Laboratory, supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Thank you!