Entangled Monte Carlo (EMC)

Poster: Tu28

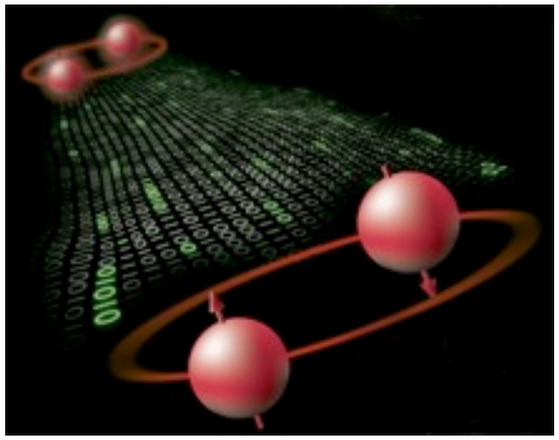
Authors: S. Jun, L. Wang, A. Bouchard-Côté
Speaker: Seong-Hwan Jun

Department of Statistics, University of British Columbia

- The sequential Monte Carlo: resampling step is critical to obtaining “good” samples. But this step makes parallelization harder.

- Entangled Monte Carlo: Scalable parallelization of sequential Monte Carlo, independent of the size of the particles.

- Name borrowed from “Quantum Entanglement” in Quantum mechanics.



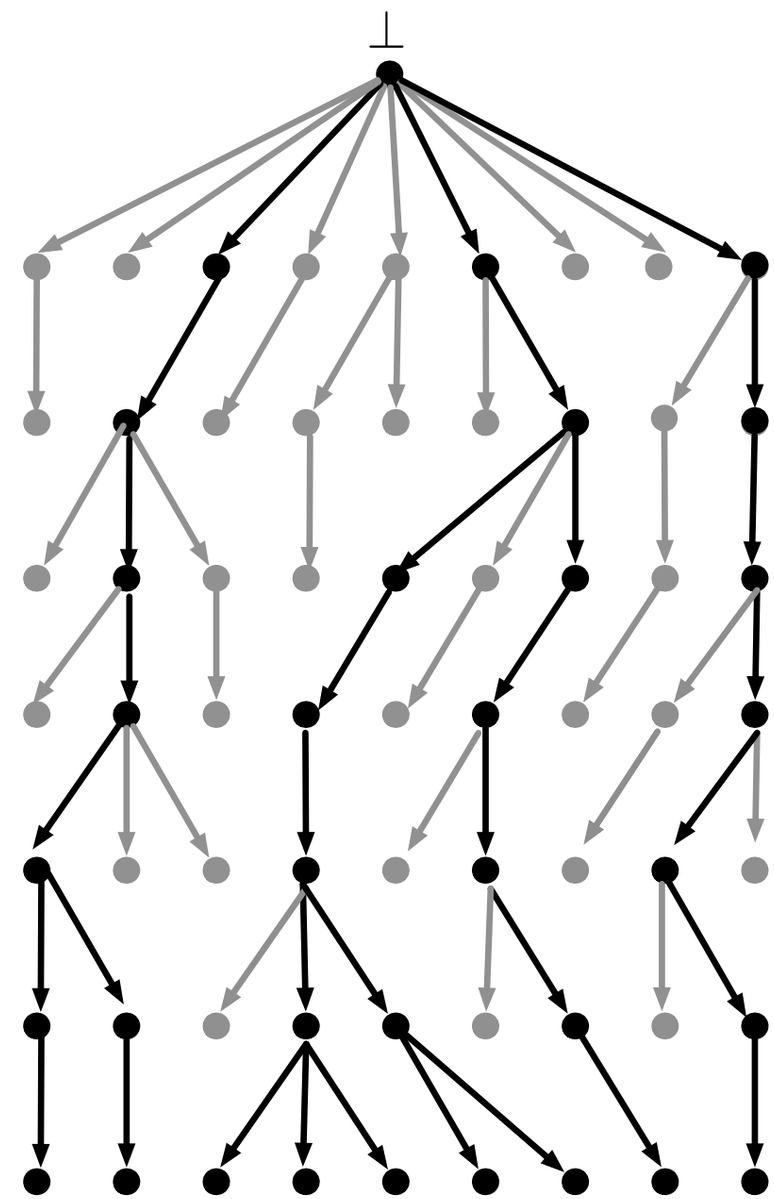
Key idea: Reconstruct the particles using the particle genealogy.

Sequential Monte Carlo & Particle genealogy

Poster: Tu28

Sampling stage of sequential Monte Carlo is "embarrassingly" parallelizable... over multiple processors.

The resampling stage induces the genealogy of the surviving particles.



Many particle genealogies die as a result of the resampling stage.

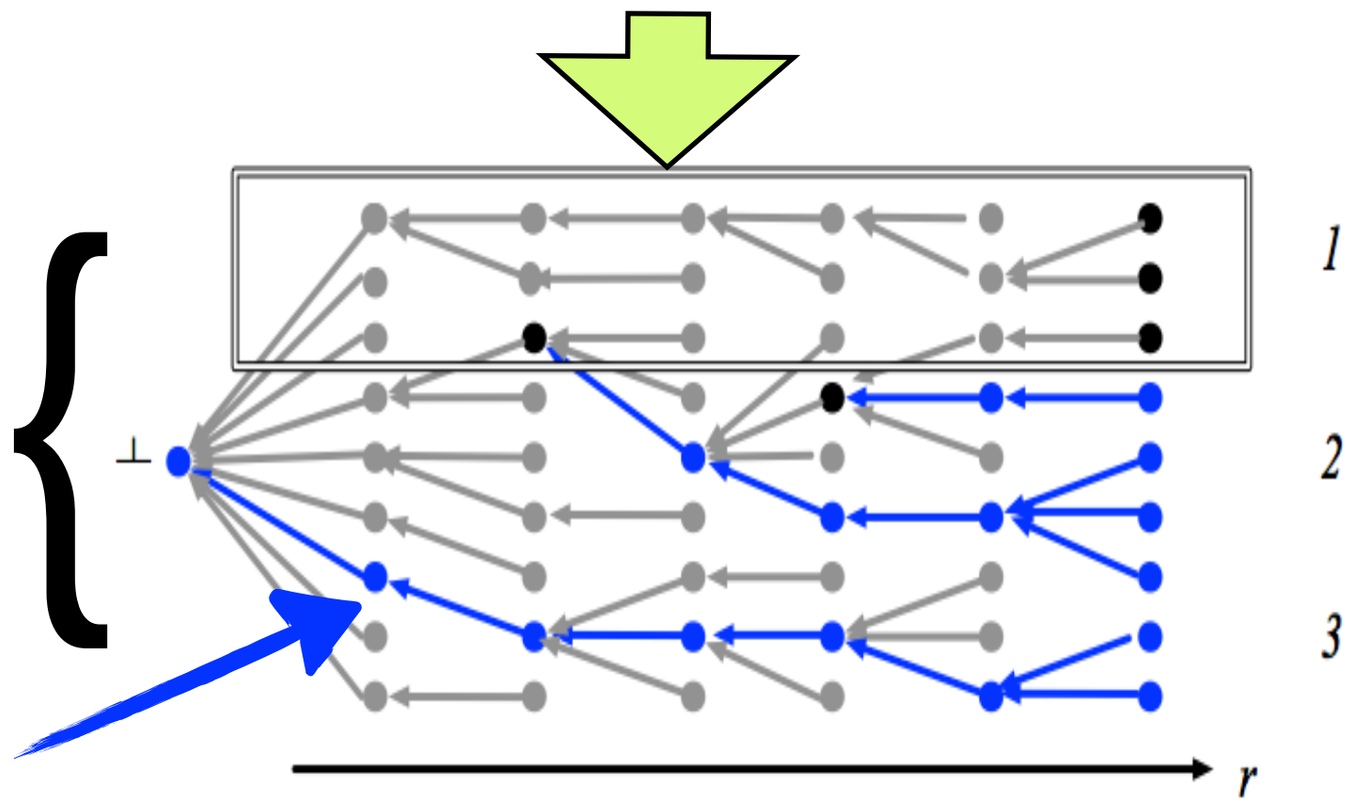
Entangled Monte Carlo & Reconstruction

Poster: Tu28

Total of N particles in the simulation but each machine generates N_k particles allocated to it.

Locally, each machine only needs to store the particles that it needs for the next stage.

The global view of the simulation is tracked by all machines.



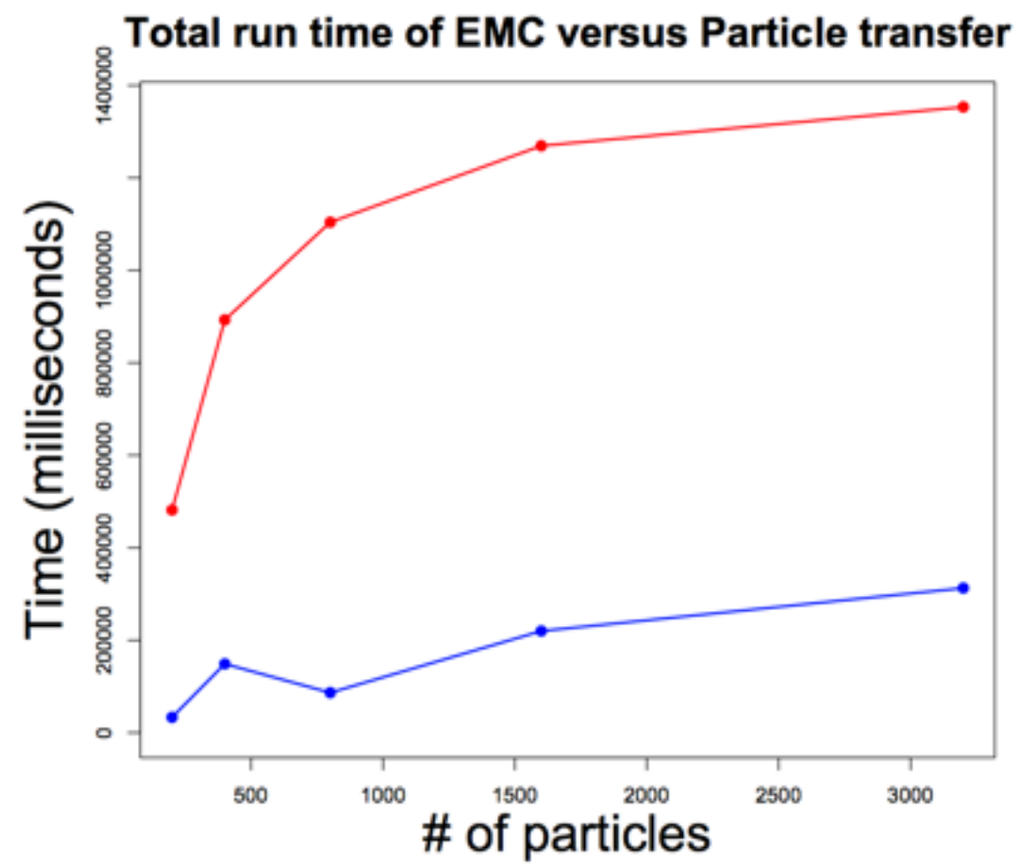
The “compact” particles (blue) are implicitly stored in machine 1 to maintain the coherence of the global simulation.

Each compact particle stores its “*stochastic map*” and the pointer to its parent.

Why it works?

Poster: Tu28

- The reconstruction rarely traces deep - Kingman's coalescent theory.
- The reconstruction relies on the CPU cycles.
- The particle transmission relies on the network protocol.
- Experiments on Bayesian phylogenetics example.
- Scalable to arbitrary number of nodes when combined with *distributed hash table* to store the particle genealogy.



Conclusion: The particles are reconstructed rather than transmitted, which allows the network communication cost to be independent of the particle size.