Kavli Institute
for Cosmological Physics
AT THE UNIVERSITY OF CHICAGO

Reducing the Size of N-Body Particle Catalogs by an Order of Magnitude

Phil Mansfield

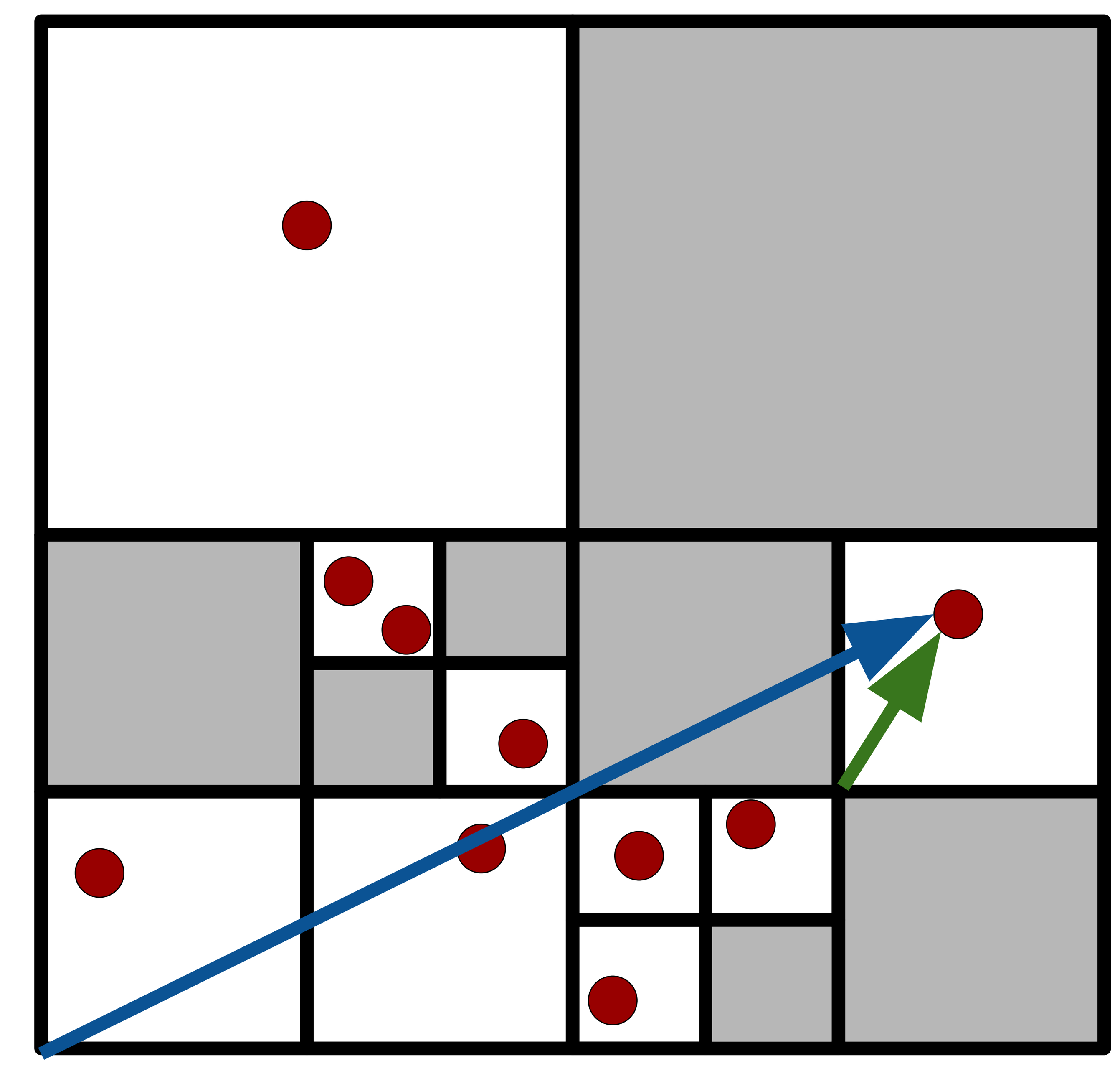
Abstract:

The size of astrophysical N-body particle catalogs is an increasingly pressing problem. The most massive simulations like Dark Sky or Blue Tides consume dozens of petabytes of disk space. The large sizes of these simulations pose significant problems for both the teams that run these simulations and for the teams which analyze subsets of them, hampering the community's ability to use these resources.
Here, I outline the **minnow** compression algorithm which allows for lossy compression of positions, velocities and IDs up to **user specified error limits**.

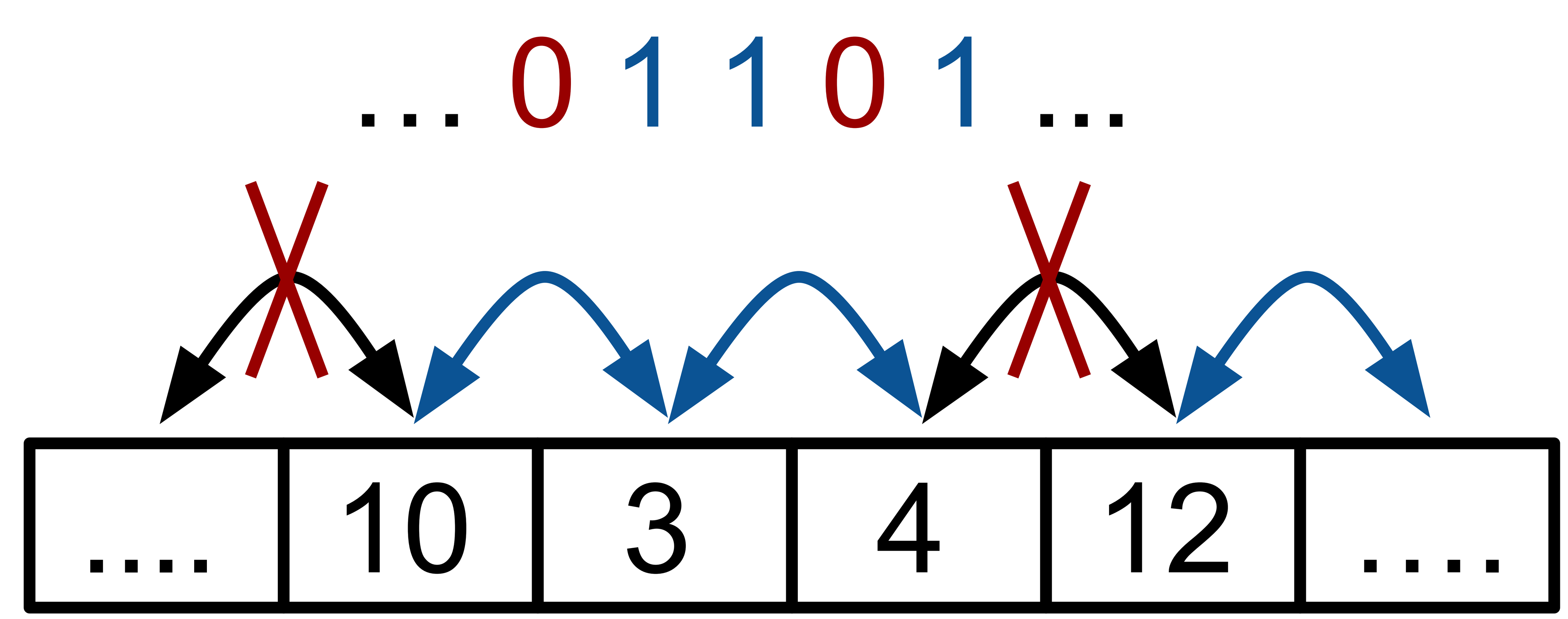
Methods:

Value Encoding: To store vectors, minnow constructs an oct-tree and encodes each vector as an offset from that vector's oct instead of as an offset from some global zero-point.

Numerous compression techniques are used to efficiently convert the structure of the tree into a byte stream which can be unambiguously decoded.

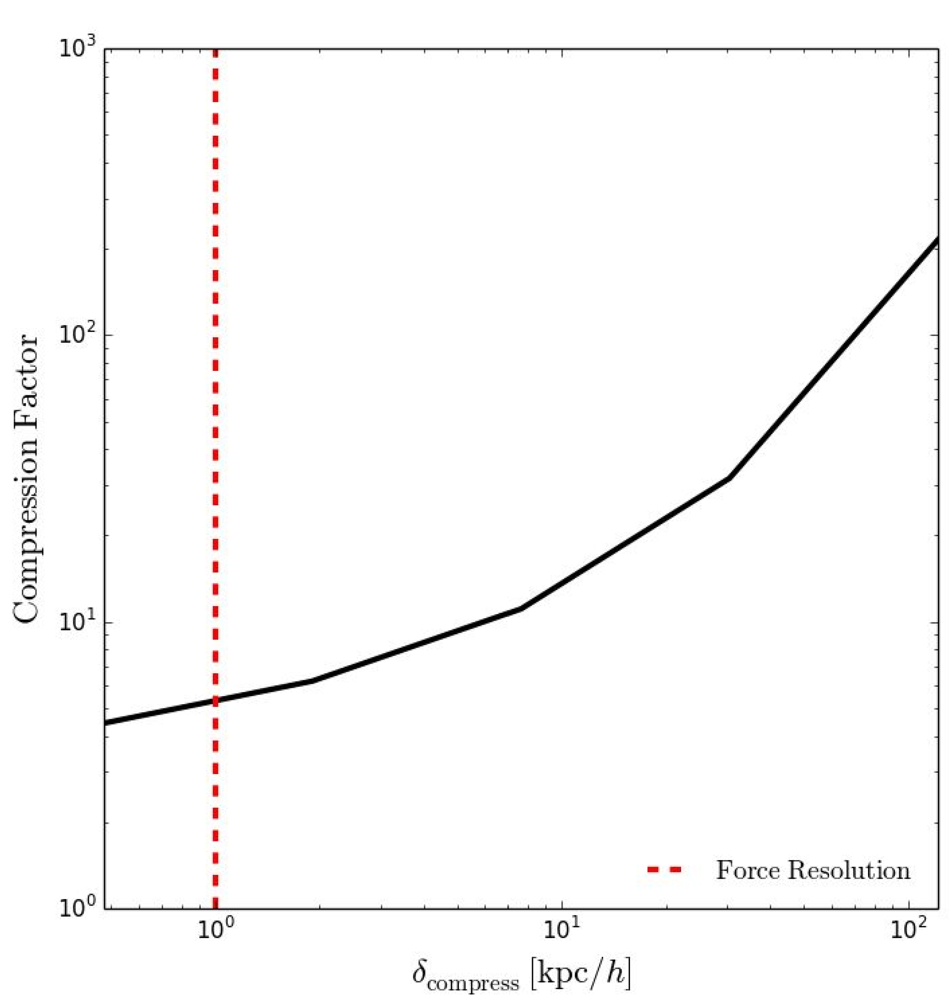


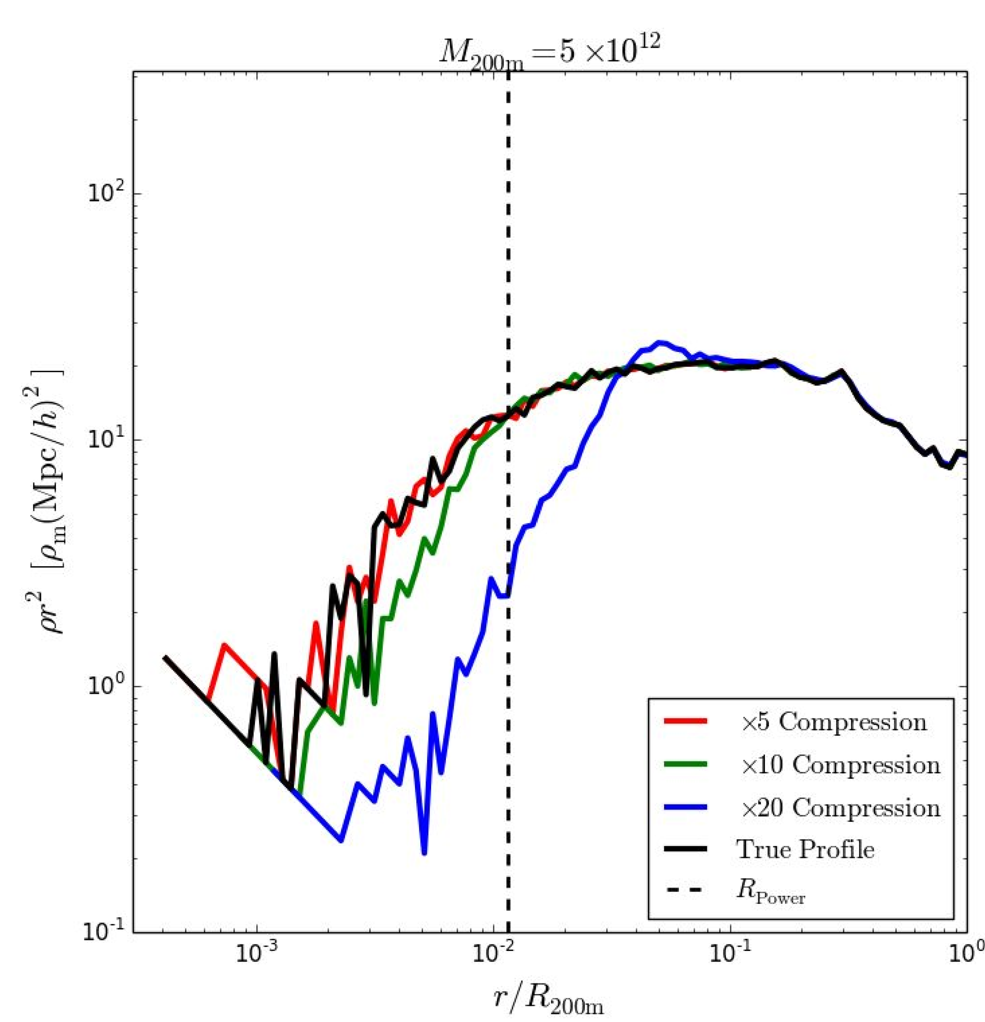
Order Encoding: As part of the oct-tree encoding process, the vector order is partially shuffled. To store the correct order, minnow creates a bit stream representing the swaps that would be made by an optimized, adaptive sorting algorithm when converting from one ordering to another.



Results:

Compression Factors: Comparison of the compression factors achieved when encoding the positions of a billion particle simulation to different accuracies.





Profile Accuracy: Comparison of a typical halo profile at different levels of compression. The true profile is only converged to the right of the vertical line.

NSF

THE KAVLI FOUNDATION