Nearest neighbor on manycore

Lawrence Cayton
Max Planck Institute, Tübingen
The point of this work

Want:

Fast nearest neighbor retrieval in (intrinsically low-dim) metric spaces....

Effective (and easy to implement) on modern many-core CPUs and GPUs.
Standard approaches to NN

Decompose space; hopefully will only have to look at a small part

Explore tree in conditional way.

recent results: [Beygelzimer et al., 2006, Ram et al., 2009]
Challenges for parallelism

1. Memory issues, practical and theoretical.
2. Trees explored in a data-dependent way.

Complex conditional computation seems difficult to distribute.
Random ball cover

$r$ random representatives

ball around representatives containing $s$ points
3 brute force searches; but only between subsets of the DB

e.g. all representatives to all DB points

1. Reduces work required for NN search,
2. but still parallelizes effectively (like brute force).
3. Simple to implement.
Code available for download.