Efficient Estimation of N-point Spatial Statistics

- *n-point correlation functions* give the probability of points occurring in a given configuration

- A general, powerful spatial statistic, capable of fully characterizing any distribution

- Previously used to understand:
  - Hierarchical structure formation
  - Gaussianity of the early universe
  - Models of galaxy mass bias
Computational Task

- Estimate n-point functions by counting $n$-tuples of points satisfying some distance constraints - $O(N^n)$ directly, per set of constraints
- Need many sets of constraints - repeat computation $M$ times
- Need to estimate variance - repeat the computation for $J$ subsamples
- Need large $n$ (at least 3) to accurately distinguish distributions

SDSS (millions of points)  Virgo Sim. (billions of points)

Overall complexity:  $O(J \cdot M \cdot N^n)$
Efficient Computation

- Build \( kd \)-trees on the data
- Compare \( n \) nodes, prune if distance bounds allow
- Share information among different matchers
  - overcome dependence on \( M \)
- Incorporate jackknife resampling directly
  - overcome dependence on \( J \)

\( kd \)-tree Level 2  \( kd \)-tree Level 4

New

\[ r_1 \]

\[ r_2 \]

prune if \( d > r_1 \)
Preliminary Results & Ongoing Work

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2 point cor.</td>
<td>4.96 s</td>
<td>352.8 s</td>
<td>2.0 x $10^7$ s</td>
</tr>
<tr>
<td>100 matchers</td>
<td>new</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 point cor.</td>
<td>13.58 s</td>
<td>891.6 s</td>
<td>1.1 x $10^{11}$ s</td>
</tr>
<tr>
<td>243 matchers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 point cor.</td>
<td>503.6 s</td>
<td>14530 s</td>
<td>2.3 x $10^{14}$ s</td>
</tr>
<tr>
<td>216 matchers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10$^6$ mock galaxies

- Heterogeneous Architectures: perform leaf-leaf computations very efficiently on GPU
- Massively Parallel tree code: scales to thousands of processors