

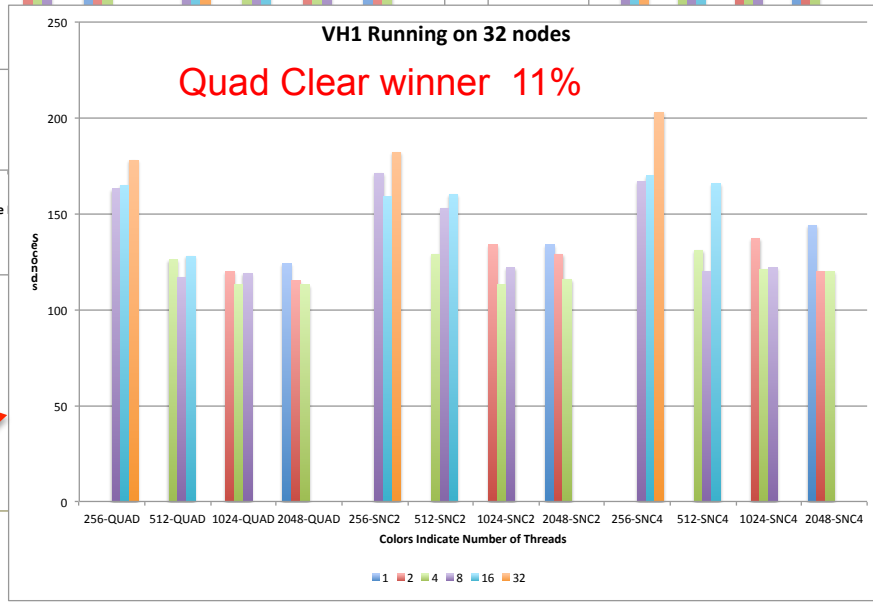
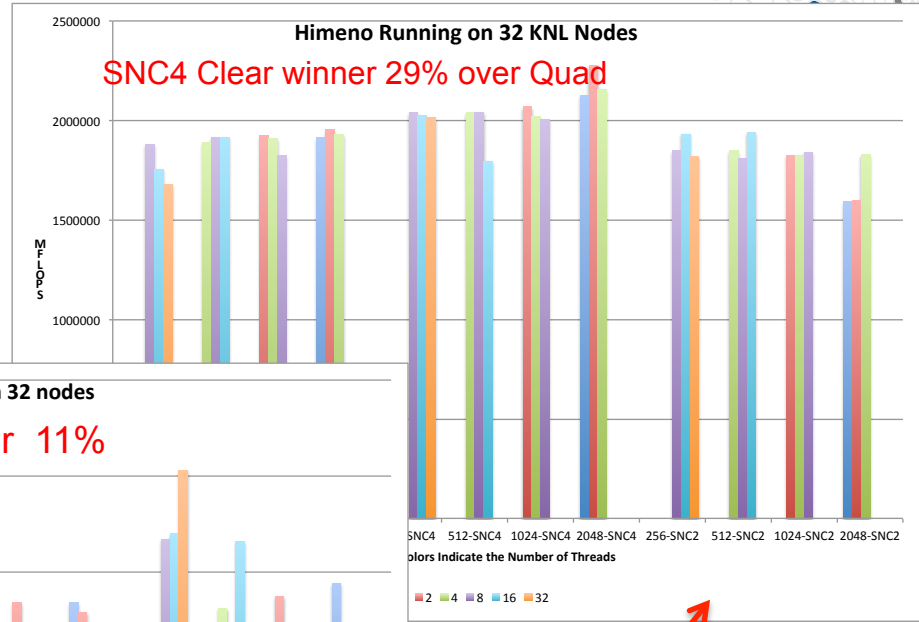
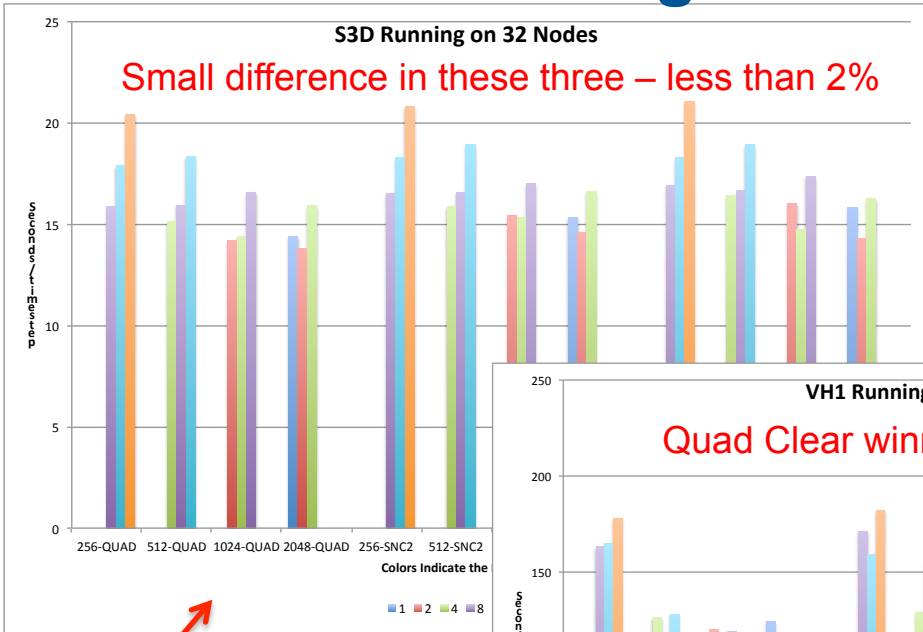
Looking at Clustering and MCDRAM options on large applications

John M Levesque
Director Cray's Supercomputing Center of
Excellence
CTO Office

MCDRAM Clustering Modes

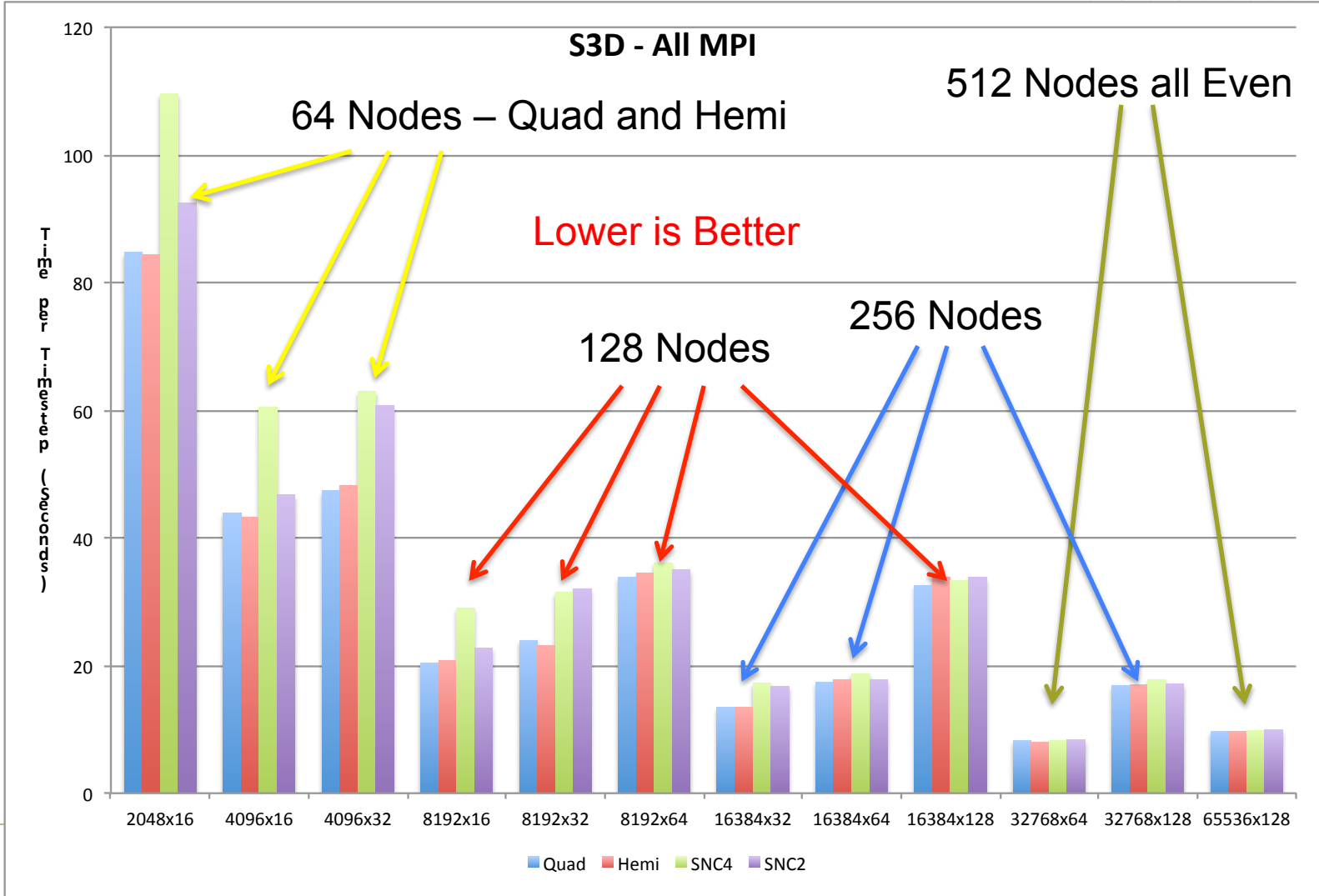
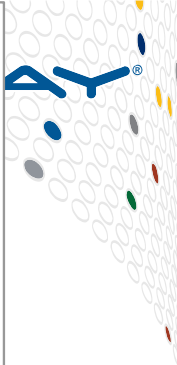
| | Cache Homing Agent | Memory | Bandwidth within Region | All MPI | OpenMP on all cores |
|------------------------------|--------------------|---------------|-------------------------|---------|---------------------|
| All to All Only 1 assured | Anywhere | Anywhere | All | Bad | Bad |
| Hemi | Anywhere | Within Region | All | OK | Bad |
| Quad 2&3 assured | Anywhere | Within Region | All | Best | Bad |
| SNC2 | Within Region | Within Region | 1/2 of Bandwidth | Poor | Good |
| SNC4 1,2&3 assured | Within Region | Within Region | 1/4 of Bandwidth | Bad | Best |

Different Clustering Modes on 32 Nodes

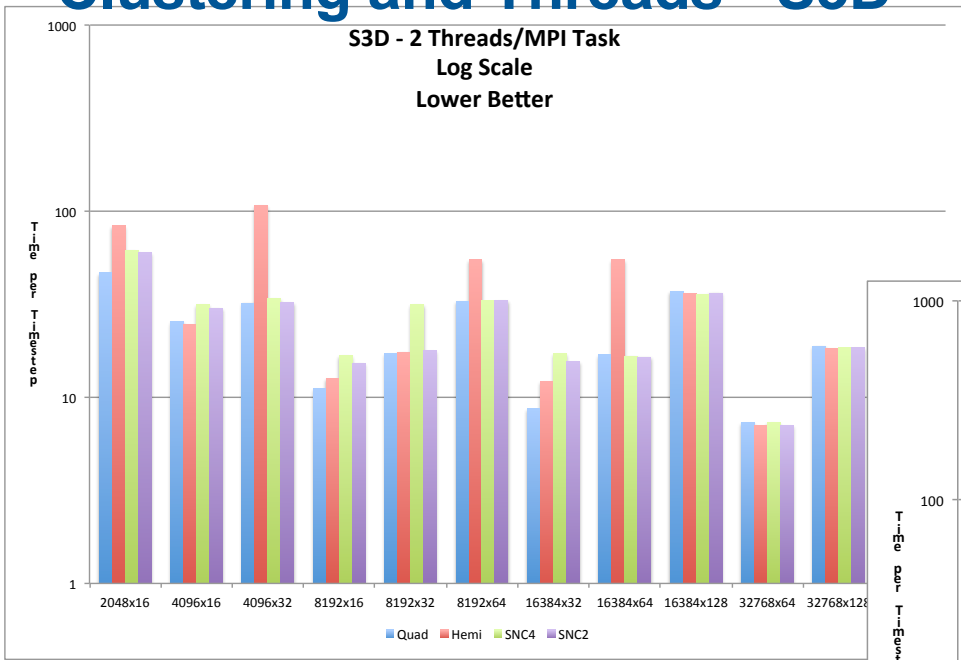


Lower Better

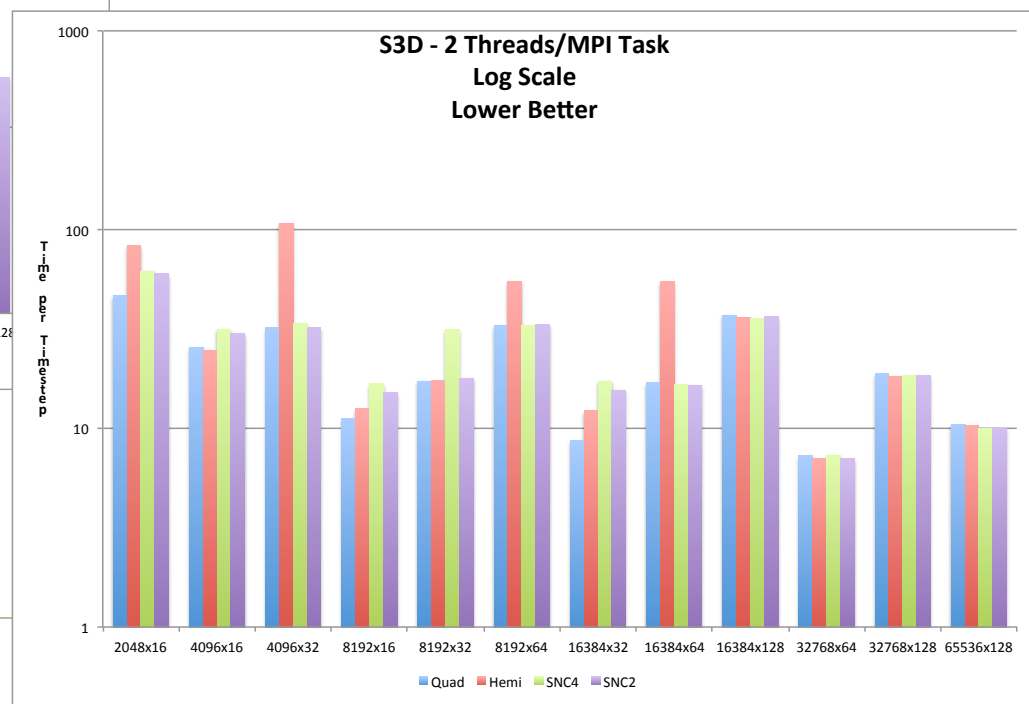
Higher Better



Clustering and Threads -S3D – Cache Friendly



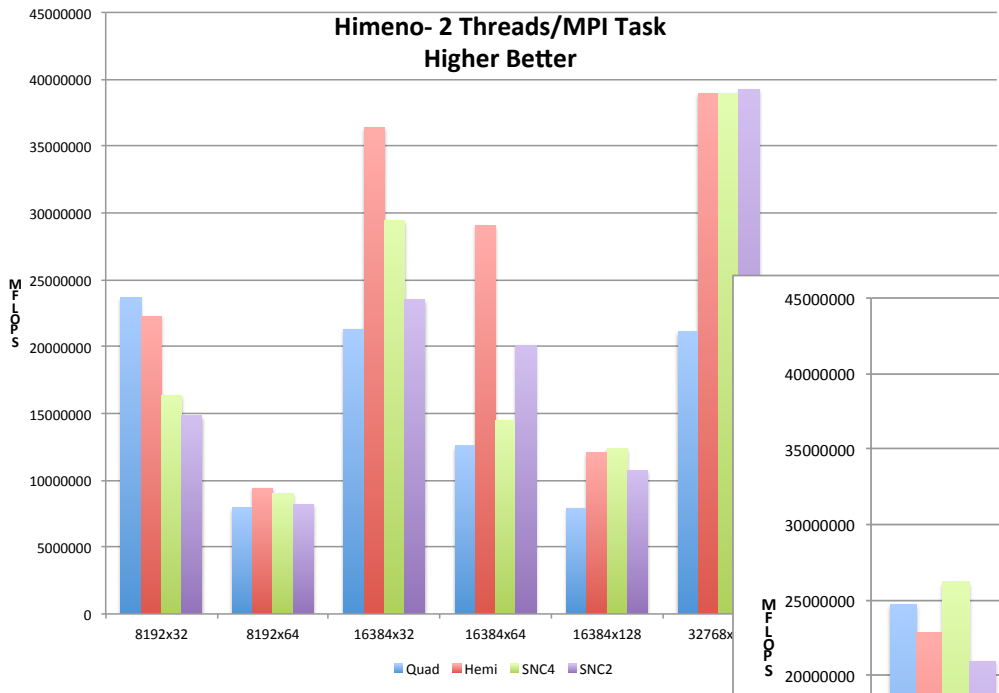
Lower is Better



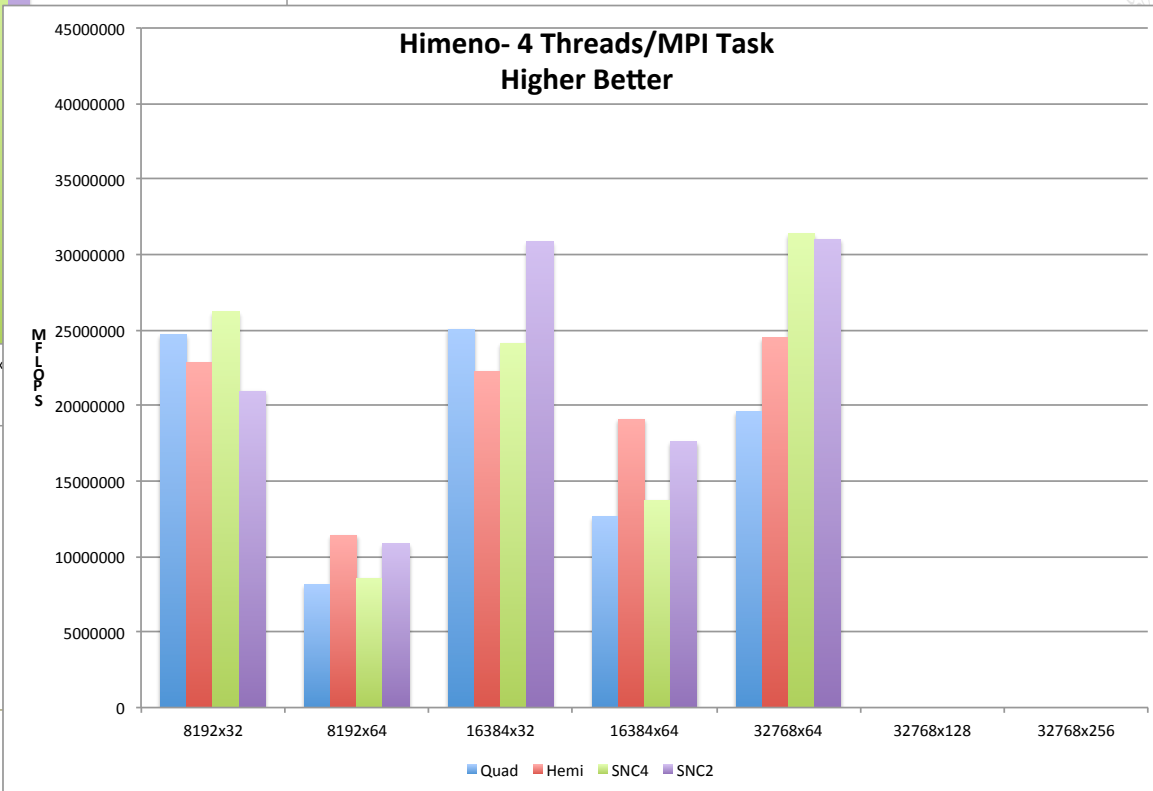


Himeno – Memory Bandwidth Limited

Himeno- 2 Threads/MPI Task Higher Better

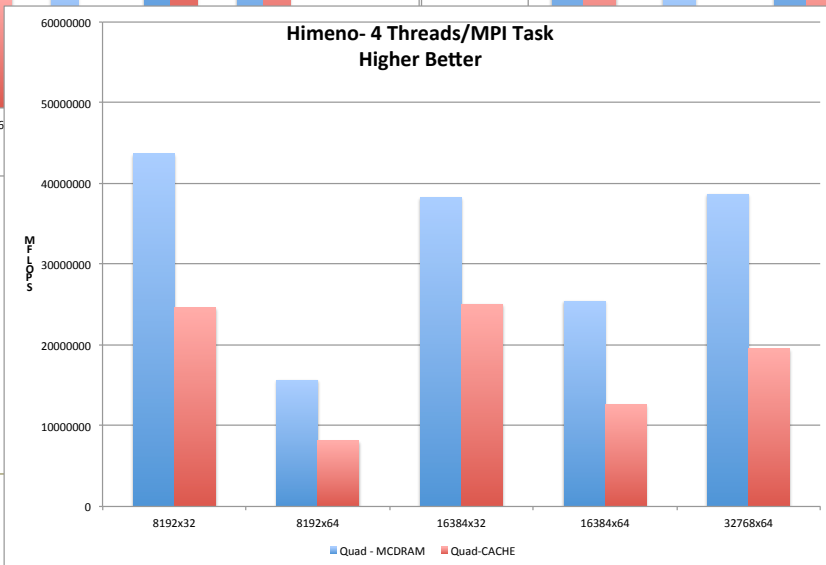
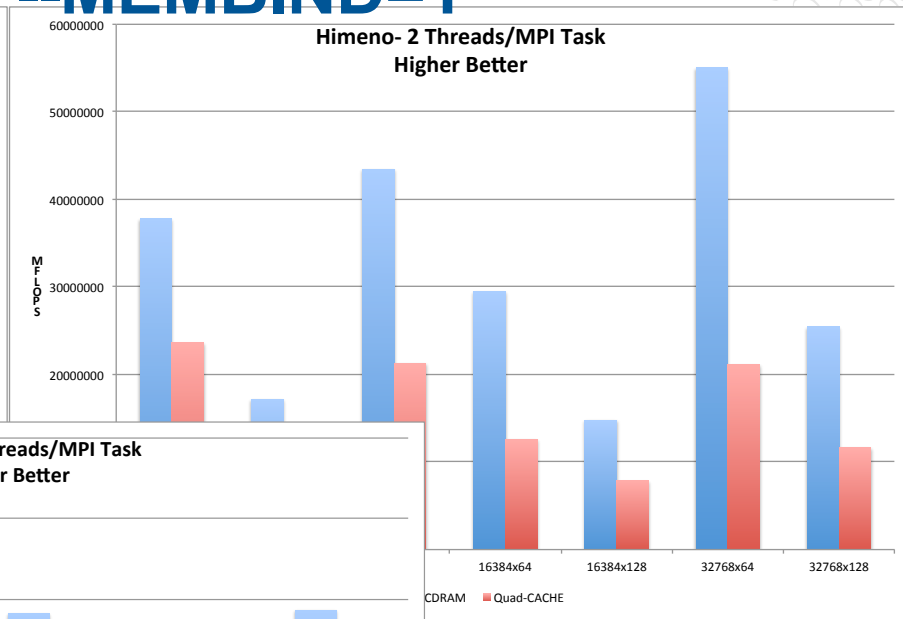
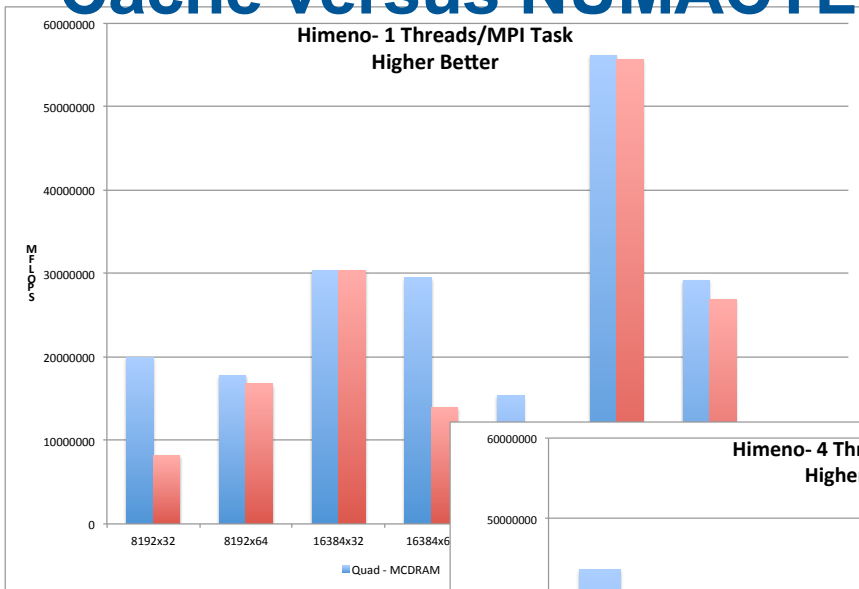


Himeno- 4 Threads/MPI Task Higher Better

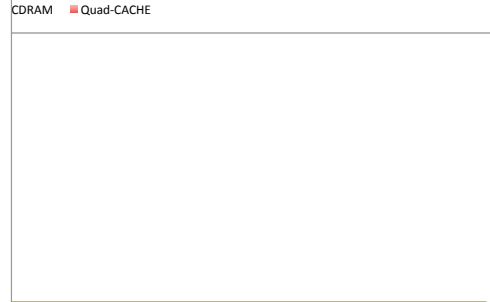


Higher is Better

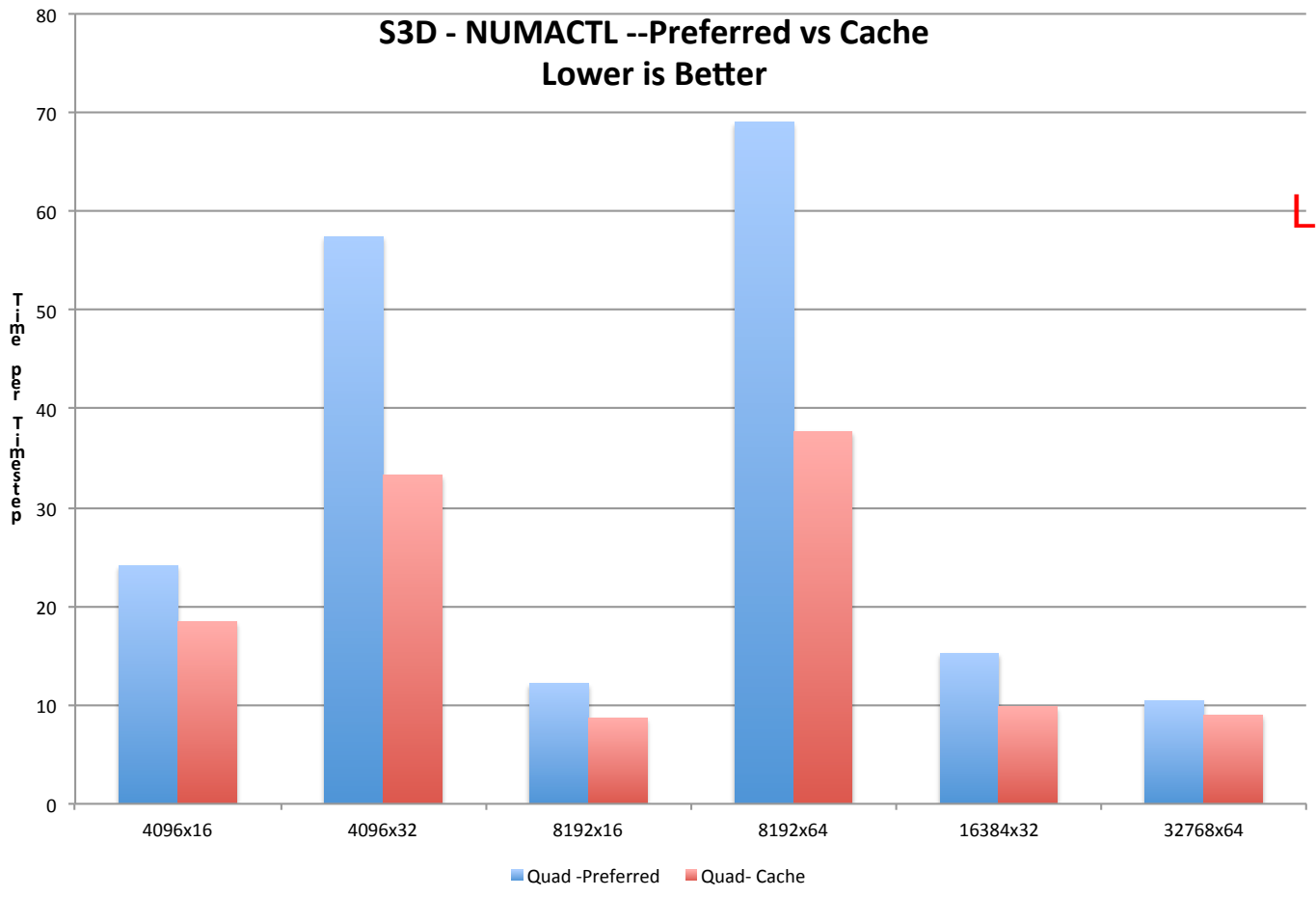
Cache versus NUMACTL --MEMBIND=1



Higher is Better



Cache versus NUMCTL --Preferred =1



Lower is Better

Conclusions

● Clustering

- Unfortunately choice is application dependent
- Re-provisioning is expensive
 - Sites should standardize on Quad-cache

● MCDRAM Memory Modes

- Unfortunately choice is application dependent
- Re-provisioning is expensive
 - Sites should standardize on Quad-cache