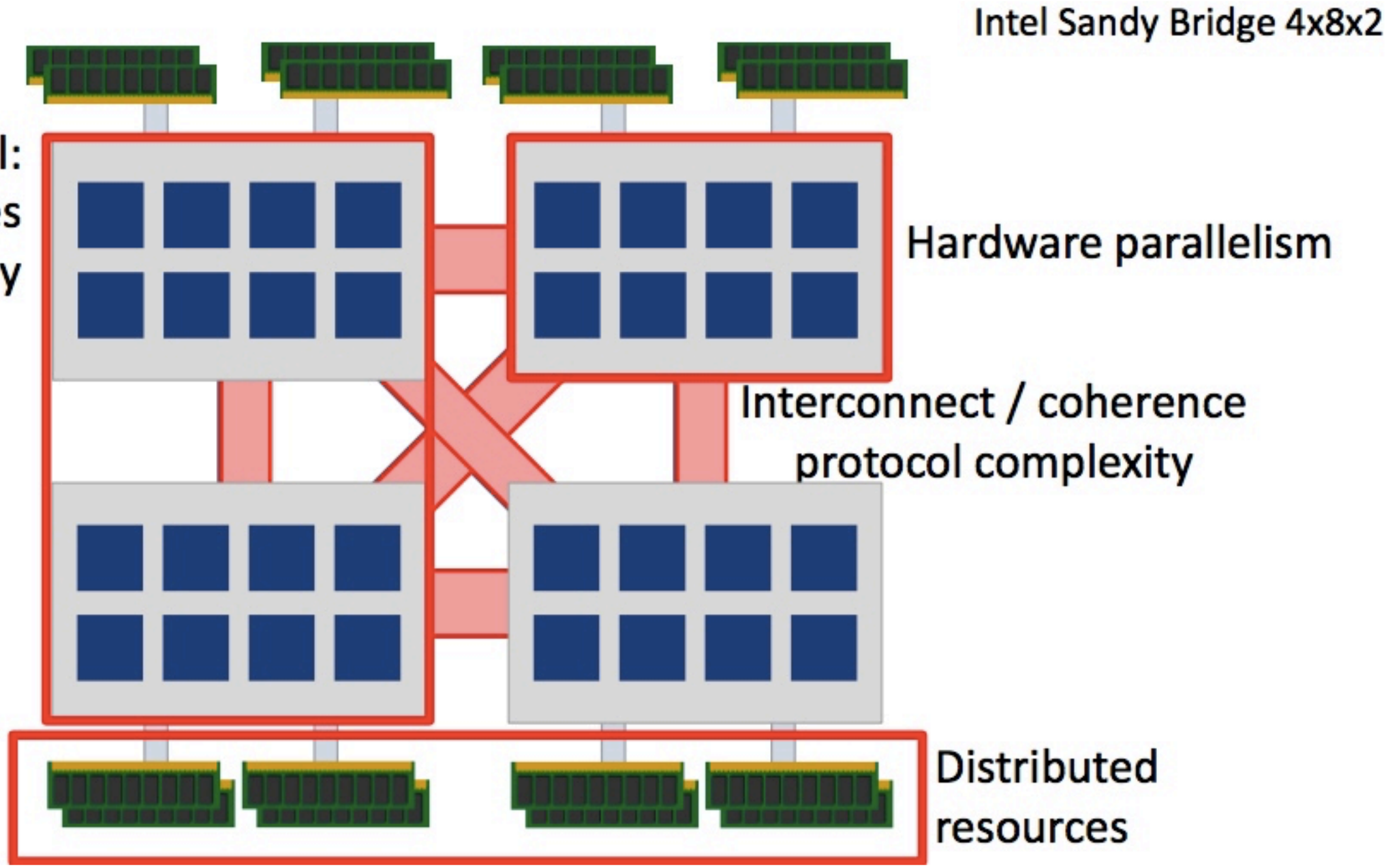
Taking the most out of Asymmetric NUMA Systems

11th EuroSys Doctoral Workshop (EuroDW'17)

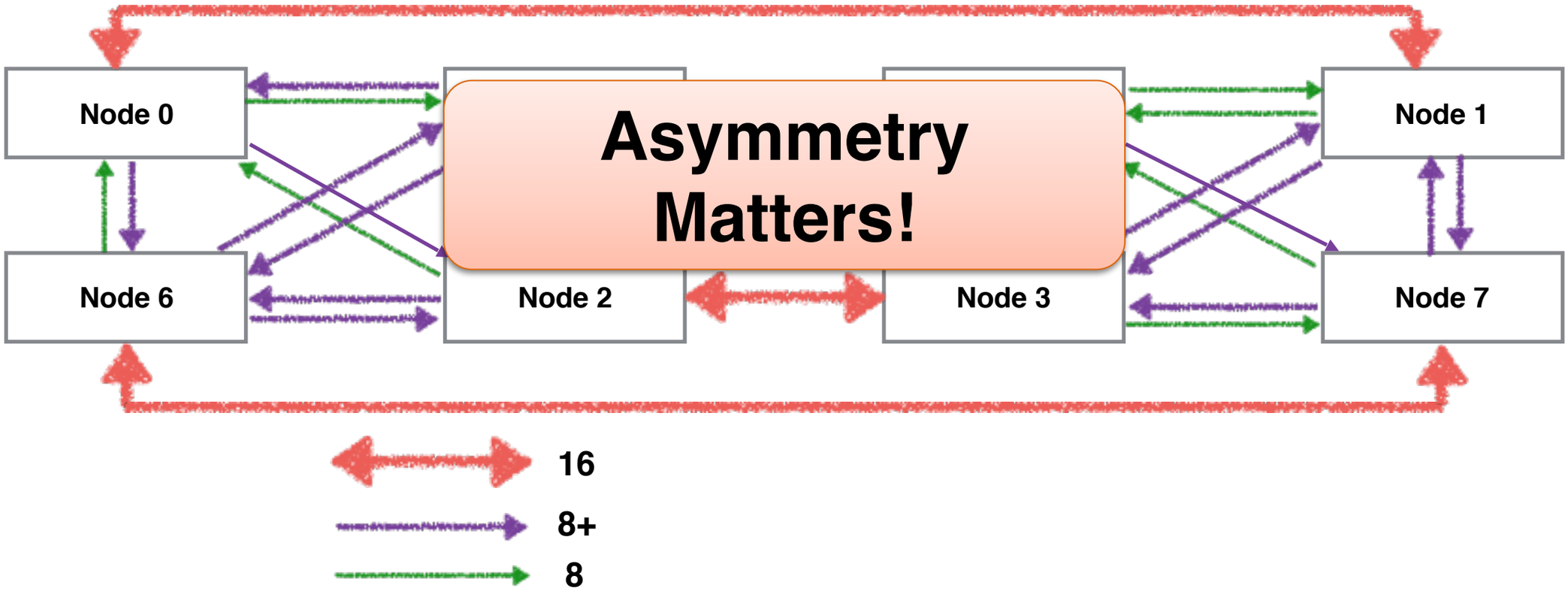
David Daharewa Gureya, Joao Barreto, Paolo Romano, Rodrigo Rodrigues {IST, U. de Lisboa & INESC-ID}, Vivien Quema {Grenoble INP}, Pramod Bhatotia {University of Edinburgh}

Modern architectures are complex



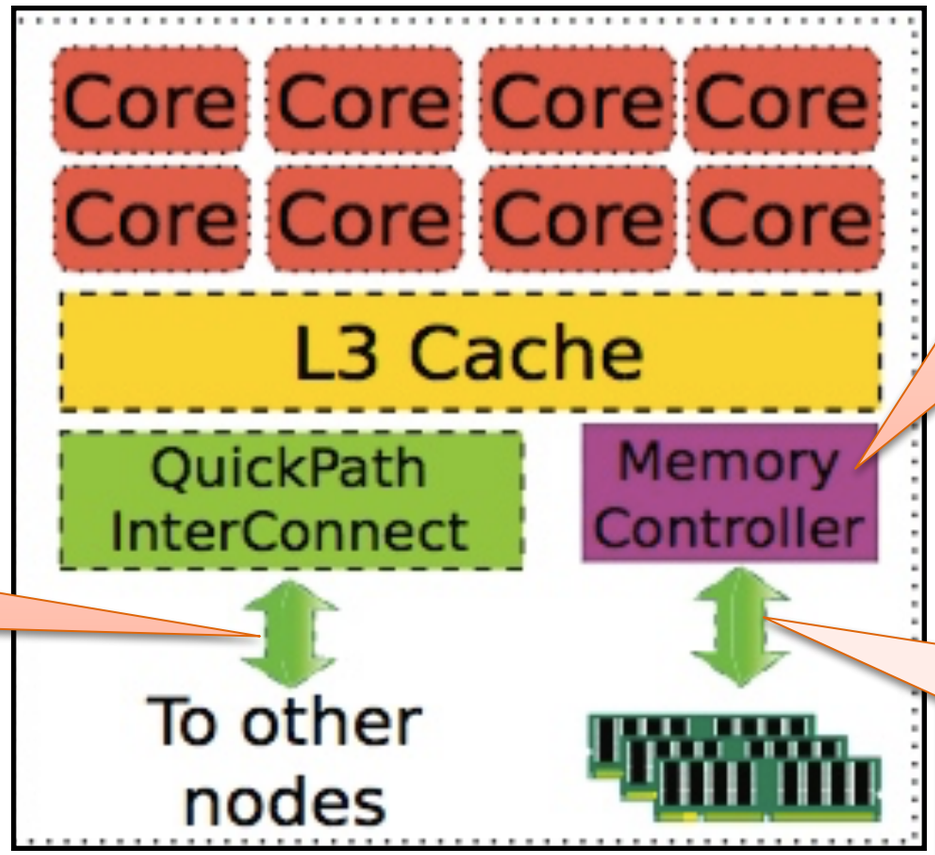
Machines are Asymmetrically NUMA

AMD Opteron 8x6x1



- Interconnect links exhibit different bandwidths
- Direct and Indirect links

Shared resource contention can lead to performance degradation



Load Imbalance of Memory Controllers

Congestion on Interconnect

Congestion on the Memory Controllers

Example: Streamcluster running on 2 nodes

- Streamcluster (data-intensive application) executing with 12 threads on 2 nodes

✓ Where to place threads?

✓ Where to allocate pages?

Example: Streamcluster running on 2 nodes

Best placement

Thread Placement	Data Placement	Execution Time (s)
0,1	0 1	188 182

Nodes	% perf. relative to the best configuration
5, 6	**-91%**

Worst placement

0,7	0 7 0,7	**219** 328 195

Goal: place threads and data optimally

- **Performance of parallel application depends on; memory allocation, thread placement and data structures used (distributed/replicated)**
- **suboptimal allocation = bad performance**

Challenges

- **Actually, thread/data placement is hard because**:

  ✓ It depends on accurate online measurements of communication patterns

  ✓ It is combinatorially difficult

  ✓ It needs to adapt to changes in workloads

  ✓ All the above becomes harder when multiple applications exist

Current Optimizations

- Focus exclusively on Thread Placement [*NuCore HPCA'16]* or Data Placement [*Shoal USENIX ATC'15, Carrefour ASPLOS'13]*

- Rely on heuristic simplifications of hardware topology and memory behavior of the application [*AsymSched USENIX ATC'15]*

- Neglect high impact micro-bursts

Our approach

- Unify thread and data placement
- Model all factors that affect performance on an asymmetric NUMA System
  - ✓ Hardware topology
  - ✓ Memory behavior of the application
  - ✓ High impact micro-burst events
- Dynamically adjusting data and thread placement as workload changes

THANK YOU!

Email: *dgureya@gsd.inesc-id.pt*