

# Comments on Future of Clusters for LQCD

Don Holmgren

LQCD SciDAC-2 Software Workshop

Nov 7-8, 2008

# FY08/FY09 Cluster

- New cluster at FNAL: JPsi
  - FY08 piece (5.8 TFlops) online Jan 1 (friendly user mode starts Thanksgiving)
  - FY09 piece will increase to 8.6 TFlops
    - Online by June 1, probably much earlier (Mar 1)
    - Waiting on FY09 funds
  - Based on 2.1 GHz quad-core Opterons
    - Similar to JLab 7n, but faster memory bus
    - FY08: 586 nodes
    - FY09: 280+ additional nodes

# Intel Nehalem

- New architecture, shipping in volume March 2009
  - Performance is still under NDA
  - NUMA, same as current Opteron
    - At least 2.5x the memory bandwidth of Opteron
    - Pricing uncertain
  - LQCD performance on single node and Infiniband cluster was more than 2X over JPsi
    - Outruns DDR Infiniband – have to shift to QDR

# Nehalem cont'd

- Probable 2010 cluster:
  - Dual-socket, quad-core
  - QDR IB
- Possible perturbations (doubtful for 2010):
  - Dual-socket, hex-core
  - Quad-socket, quad- or hex-core
- 8+ cores/socket in 2011 forward

# GPUs

- PCI-E gen2 probably critical to feed GPU's for our sort of application
  - Only Intel hosts support this now
  - Waiting for NVidia to release support for inter-GPU communications (and possibly inter-node, inter-GPU communications)
- Hybrid clusters currently difficult to build
  - PCI-E gen2 x16 ports not typically available on cluster-style 1U servers
- FNAL, JLab have budgeted for prototypes in FY09