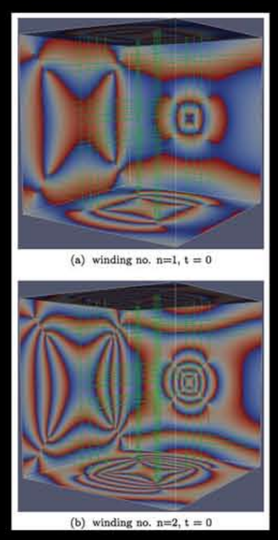
Connecting Communities Through HPC

SC11

2011 Chair
Scott Lathrop
Seattle, WA



(a) winding no. n=1, t = 0

(b) winding no. n=2, t = 0

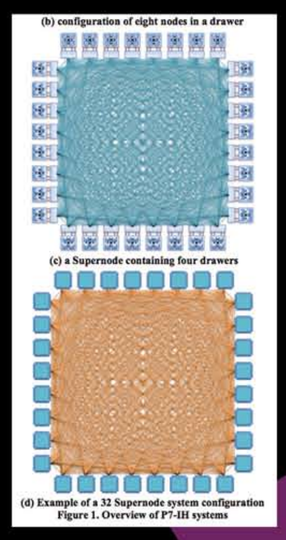


Figure 1. Overview of P7-IH systems

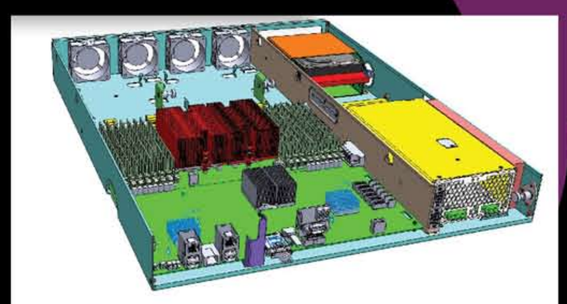


Figure 5: Motherboard and chassis layout for efficient airflow. Two CPUs and Eighteen DIMMs are placed side by side, yielding the best thermal result based on airflow coming from front to rear. They are the primary power consumers and heat generators, taking up over 80% of the total system power, so spreading them apart ensures that they do not heat up each other. All the I/O ports, including the debug card (shown attached here, but normally unneeded), are placed on the front to facilitate easy access for rack service. The PCIe x16 slot placement (front left) can accommodate a full-height PCIe card. The PSU power mates directly with the motherboard, eliminating PSU power cable routing (only remaining cables to the fans and HDD are flush to the sides). Mounting holes' placement offers the best support for motherboard. Extra space for up to five more HDDs is visible behind the motherboard and PSU.

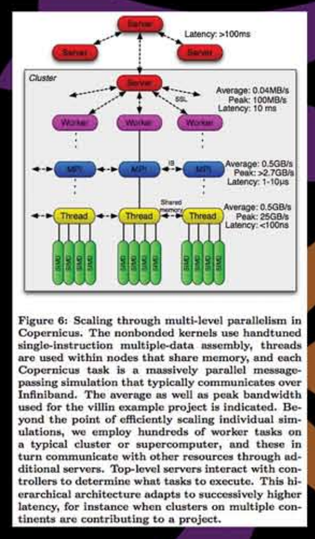


Figure 6: Scaling through multi-level parallelism in Copernicus. The nonbonded kernels use handtuned single-instruction multiple-data assembly, threads are used within nodes that share memory, and each Copernicus task is a massively parallel message-passing simulation that typically communicates over Infiniband. The average as well as peak bandwidth used for the villin example project is indicated. Beyond the point of efficiently scaling individual simulations, we employ hundreds of worker tasks on a typical cluster or supercomputer, and these in turn communicate with other resources through additional servers. Top-level servers interact with controllers to determine what tasks to execute. This hierarchical architecture adapts to successively higher latency, for instance when clusters on multiple continents are contributing to a project.

2011

Notable Systems first mentioned this year in the proceedings:

- Cray XT5
- Jaguar
- T2K Open Supercomputer (Tokyo)
- Roadrunner
- NCSA Lincoln GPU cluster
- Amazon EC2
- Ranger

Application/Problem Studies:

- Communication optimization in asymmetric interconnects
- Optical memory access networks
- Heart simulation
- High resolution weather prediction
- Molecular Dynamics Discrete Particle Simulation
- Effects of System Noise on Application

Noteworthy Architecture Topics:

- All-to-All Communications
- Evaluating networks
- Multi-domain dynamic power and clock frequency management for chip MPUs
- GPU acceleration of memory-intensive application
- Adding new levels/devices to memory hierarchies
- GPGPUs and expanded programming models for them
- Formal verification of MPI programs

Other Topics:

- Scaleable checkpointing

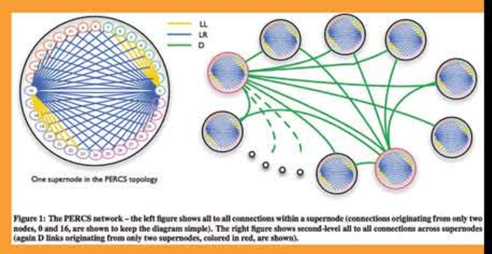


Figure 1: The PERCS network – the left figure shows all to all connections within a supernode (connections originating from only two nodes, 0 and 16, are shown to keep the diagram simple). The right figure shows second-level all to all connections across supernodes (again D links originating from only two supernodes, colored in red, are shown).

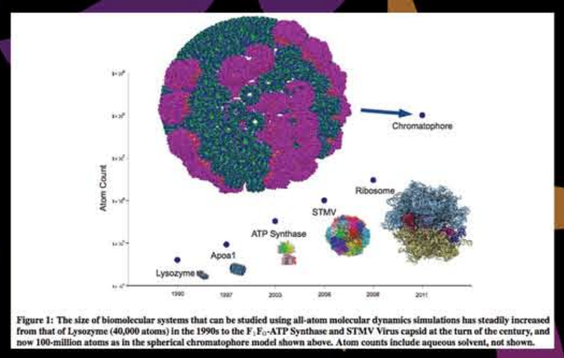


Figure 1: The size of biomolecular systems that can be studied using all-atom molecular dynamics simulations has steadily increased from that of Lysozyme (40,000 atoms) in the 1990s to the F_1F_0-ATP Synthase and STMV Virus capsid at the turn of the century, and now 100-million atoms as in the spherical chromatophore model shown above. Atom counts include aqueous solvent, not shown.

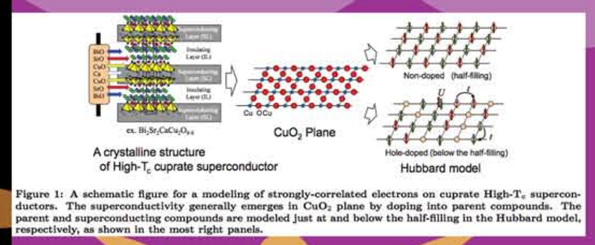


Figure 1: A schematic figure for a modeling of strongly-correlated electrons on cuprate High-T_c superconductors. The superconductivity generally emerges in CuO_2 plane by doping into parent compounds. The parent and superconducting compounds are modeled just at and below the half-filling in the Hubbard model, respectively, as shown in the most right panels.

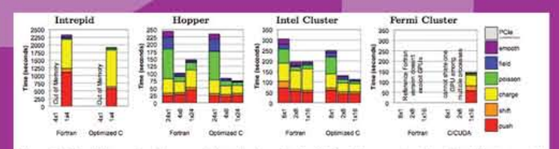


Figure 3: Breakdown of runtime spent in each phase for the B20 problem as a function of architecture, optimization, and threading. Note, minor axis is "processes×threads", or equivalently, "spartdom×OMP_NUM_THREADS". Additionally, observe each is plotted on a different timescale. PCIe time has been removed from each kernel's time and tabulated in a separate bar.