

TRENDS AND DIRECTIONS IN HPC

2014



PCPC Direct is
the nation's leading provider
of multi-brand HPC solutions.





INFINITE COMPUTE FOR FREE
...everything else is a compromise.



HPC IS ALWAYS A BALANCING ACT



HPC is always a balancing act between:

- Sockets/node
- Cores/Socket
- Memory/node
- I/O fabric between nodes
- Storage performance and capacity
- Total nodes
- Desire>\$

Sockets/node

- Current designs will stay close to what we have today. One, two and four processor nodes.
- Soon, we will see more designs like HP Moonshot, and Intel's Rack Scale Compute Architecture.



HP Moonshot - 45 servers in 4u

Intel Rack Scale Architecture Innovation

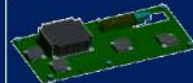


Orchestration

Open Network Platform



Intel® Silicon Photonics



Silicon: Intel® Atom™ & Xeon



CPU / Mem Modules

CORNING 



ANNOUNCING
TODAY

✓ New MXC connector

Up to 1.6 Terabits per second bandwidth¹

✓ New "ClearCurve" fiber

Enables >3x cable length (300m) at 25Gbps²

Higher Bandwidth, Modularity & Density

*Other brands and names are the property of their respective owners.

¹Measured per fiber bandwidth on an Agilent Bit Error Rate Tester (BERT) that included a N4960A-CJ1 controller, N4951A-H32 pattern generator and N4952A-E32 error detector. MXC connector used had 32 fibers for an actual data rate of .8 tera-bits. Mechanical models and CAD simulations show that the MXC can accommodate up to 64 fibers for a theoretical total bandwidth of 1.6 Tera-bits per second.

²ClearCurve fiber operating at 300 meters was tested using 300 meters of new ClearCurve fiber connected to an Agilent Bit Error Rate Tester (BERT) that included a N4960A-CJ1 controller, N4951A-H32 pattern generator and N4952A-E32 error detector.



Cores/Socket

- **ARM designs don't seem to fit seismic workloads**
- **Intel 2600v3 (Haswell) – the new Core i3 will have a 13w version, but focus is on the higher core counts**
- **Xeon PHI – Knights Landing will disrupt Nvidia CUDA, but how much?**
- **ARM – lot of buzz at SC13, but limited work being done**
- **IBM Watson – What if it takes some geology classes?**

Cores/Socket

- Intel will continue to drive cores/socket up
- Haswell has 18 cores available at launch, roadmap up to 20 core
- Nvidia Cuda will continue to be dominant
- Xeon Phi with Knight's Landing potentially more disruptive. Release expected 2H '15

Memory/node

- **DDR4 faster, lower power but still a price premium.**
- **8 or 16 DIMM slots on the low end; up to 24 or more available.**
- **IBM 3950 - 8 sockets; over 6TB RAM in a standard package**
- **Cray and SGI still pursuing proprietary NUMA hardware approaches**
- **ScaleMP's software approach moves faster, outperforms hardware approach for many workloads**
- **ScaleMP released free version, 2 cpus+8 nodes of memory total 128+ memory slots in an affordable footprint**
- **And one 16GB DIMMs is now less than two 8 GB DIMMs**

I/O Fabric – InfiniBand deathmatch!

- **QDR (Intel True Scale)**
 - Now available on some system boards at no additional cost
 - Switch and cabling still expensive, switches are coming down in price faster
 - Best latency for the price for MPI or RDMA workloads
- **FDR14 still pricey, but coming down**
 - Best throughput, but HPC is usually latency, not throughput
 - Often not worth the money, but customers chase “Ego-width”

I/O Fabric - Ethernet (cont.)

- 1 GB still around for system management, and many workloads, usually found on the motherboard
- 10 GB wider adoption on servers, many Haswell systems will have 10 GB LAN on Motherboard (LOM)
- 40 GB Top of Rack (ToR) to core switches and 40 GB ToR; and 40 GB adapters make “poor man’s InfiniBand”
- 100 GB available today for the core, and *maybe* late 2015 for ToR uplink?
- As 40 and 100 GB Ethernet becomes widespread, it will eventually kill InfiniBand, but that will take several years.

Storage

- **Storage is not usually a bottleneck that customers spend money to address, they would rather buy more nodes.**
- **DDN and NetApp are the biggest HPC storage providers**
- **Flash is on the same price/capacity path as DIMMs, and will eventually replace all spinning disk. Sooner than most people think.**
- **Scalable file systems are probably more important**

Total Nodes - Exascale

- **Being worked on today, best guess is about 2020. Probably Japan or China (even money bet)**
- **With Exascale, we will never get a complete answer, development paradigms need to change to be satisfied with a “good enough” answer.**
- **At Exascale, something will always be broken**

Total Nodes – Departmental

- **Fastest growing space, but not by a lot (3% vs 7%)**
- **Applying HPC techniques and applications to new problems and industries**
- **How to manage it? Don't necessarily want to hire a full time sys admin**

Total Nodes – Cloud

- **Cloud not seeing much growth in commercial space, mainly public sector**
- **It's ok for occasional “raw compute” and “embarrassingly parallel” jobs**
- **Sticker shock**
- **Some are looking at a hybrid model**
- **Some decent tools to make all of this work; more to come**

Market Forces

- **HP – Splitting into ink and less profitable stuff**
- **Lenovo rapidly absorbing IBM System x – Very good products at great prices?**
- **Dell went private – Doesn't seem to be “buying the business” currently**
- **SuperMicro – business as usual**
- **OpenCompute – New manufacturers, but they don't understand HPC.**

Thank you!

- **Addison Snell – Principal, Intersect 360**
- **Egan Ford – IBM Distinguished Engineer**
- **Scott Denham – Cray Computer**
- **Benzi Galili – COO, ScaleMP**
- **Kim Hawes – Intel**
- **Dan Cummings – IBM HPC Team**
- **Too many others to count**
- **And, thank you for your time today!**