

The Cray X1E – a Unique Capability Computer

James L. Schwarzmeier

Cray Inc.

jads@cray.com

715-726-4756

Two Types of HPC Applications

- **Highly scalable apps** ← XT3 space
 - problem size scales with P (*weak scaling*)
 - goal: advance science by increasing resolution or computational region for ~ *fixed runtime*
 - needs **highly scalable app, network, software**
- **Runtime-sensitive apps** ← X1E space
 - problem size fixed but add P (*strong scaling*)
 - goal: advance science by *reducing runtime* so new physics or algorithms can be evaluated quickly – week/month-long runtimes a severe impediment
 - with limited MPI parallelism in app, also attack *data level parallelism* → need **powerful (vector) processors, memory system, network**

What Makes the Cray X1E a Powerful Processor?

- 18 GFLOPS peak (MSP) + 30% eff on finite element CFD (from AHPCCRC). How? Because of **balance**, throughout CPU & memory system
 - high peak → need high L2 and memory bandwidth (~13 GB/s/MSP STREAMS TRIAD for 8 MSPs)
 - high memory bandwidth – *more than just pin bw*
 - → many outstanding memory references (2048/MSP)
 - → **compiler reduces bandwidth need** by interchange, fuse & unroll loops, vectorize *outer* loops → many registers (8192 register elements/MSP) → CPU needs many in-flight operations (1000's) → 'chain' dependent instruction streams (vectorize *across* dependence chains)
 - L2 control → *non-allocating vector references when no temporal locality*
- Need well-vectorized code (SSP mode can help)

What Gives the Cray X1E a High Capability Network?

- **All-to-all** → high bi-section bandwidth/proc
 - 8-cabinet, 1024 MSP system has payload **bi-bw 0.4 GB/s/MSP**
- **Collectives** → low latency network
 - **0.85 μsec** nearest neighbor latency with CAF/UPC
 - global AMO instructions
- **General fine-grain, dword-level communication via CAF/UPC enabled by:**
 - → **global addressability** to whole machine with load/store instructions → extend address range → avoid remote TLB misses → avoid cache coherence traffic across network
 - CAF/UPC can **overlap local and remote memory references and computation**
 - 504 MSP X1 runs HPCC Random Access @ **7 GUPS** w UPC
 - fully adaptive (**every Δt**) finite element CFD (AHPARC) w UPC