Interesting Times Ahead

Bob Lucas
Operational Director, USC – Lockheed Martin Quantum Computing Center
University of Southern California

Interesting Times Ahead

Bob Lucas
Operational Director, USC – Lockheed Martin Quantum Computing Center
University of Southern California

Institute for Defense Analyses/Center for Computing Sciences
Livermore Software Technology Corporation

ANSYS, Leidos, and Micron

Interesting Times Ahead

Bob Lucas
Operational Director, USC – Lockheed Martin Quantum Computing Center
University of Southern California

Institute for Defense Analyses/Center for Computing Sciences
Livermore Software Technology Corporation

ANSYS, Leidos, and Micron

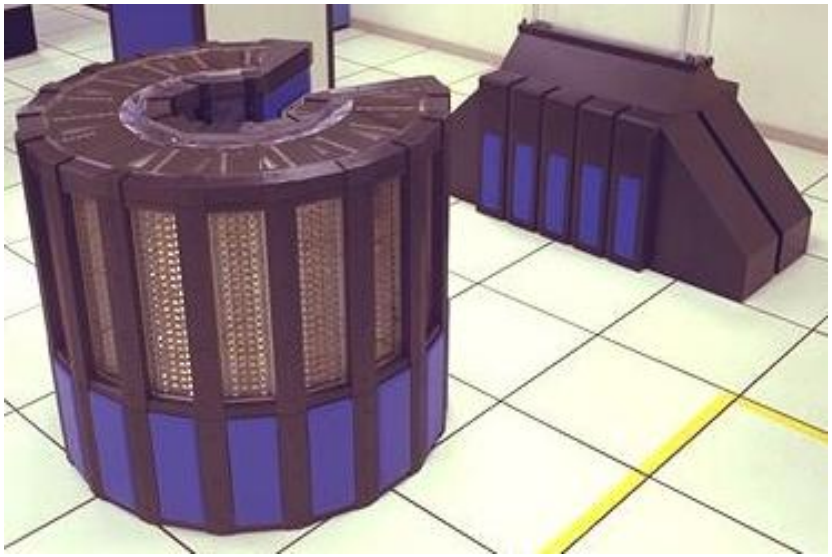
Your name here ☺

Overview

- **déjà vu, all over again?**
  - I'm reminded of my youth
- **Software Perspective**
  - It'll be harder this time around
- **Co-design**
  - Let's get the systems we need

Supercomputing in the late 1980s

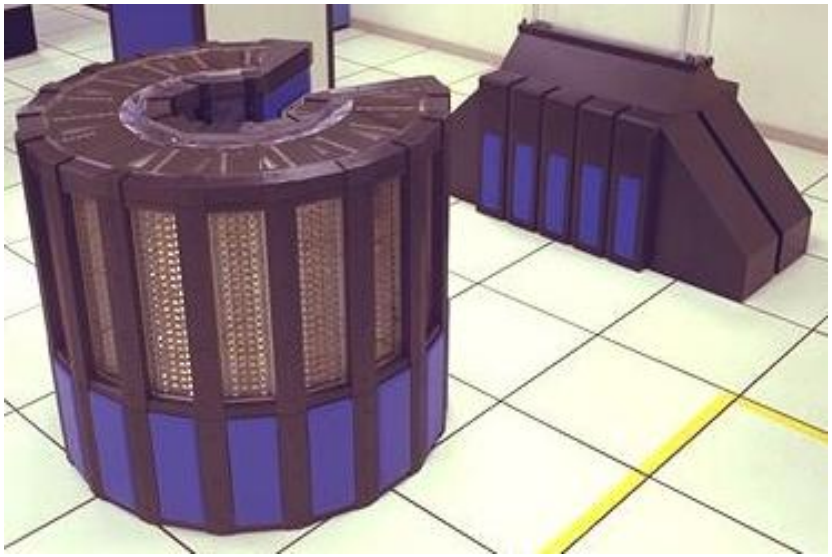
- **Cray-2 was my first supercomputer**
  - Supercomputing Research Center (now CCS)
- **Shared memory, vector mainframe**
  - O($10M)
  - Four 250 MHz ECL CPUs
  - Three warm bodies



NASA Cray-2 from Wikipedia

Supercomputing in the late 1980s

- **Cray-2 was my first supercomputer**
  - Supercomputing Research Center (now CCS)
- **Shared memory, vector mainframe**
  - O($10M)
  - Four 250 MHz ECL CPUs
  - Three warm bodies
  - People were cheap



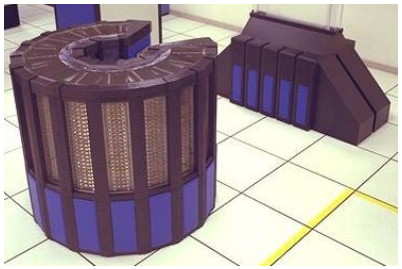
NASA Cray-2 from Wikipedia

Disruptive Technology

- **Field-Effect Transistor**
  - Patent filed in 1925
- **Metal-Oxide-Semiconductor**
  - Invented in 1959
- **Complementary MOSFETs**
  - Latch-up solved in mid-1980s
  - Lots of DOD $$$ (e.g., SEMATEC)
- **Personal Computers**
  - High volume
  - Low cost, O($1K)

IBM PC from Wikipedia

Driving down the cost of HPC



Mainframes

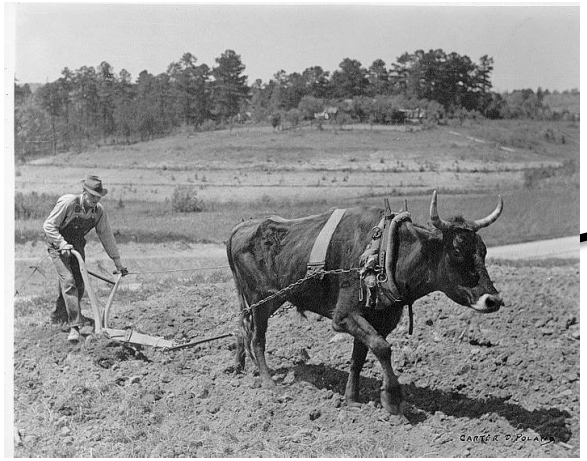
Scalable COTS

Cosmic Cube

Log Performance

Years

Seymour's perspective



Sturdy Ox

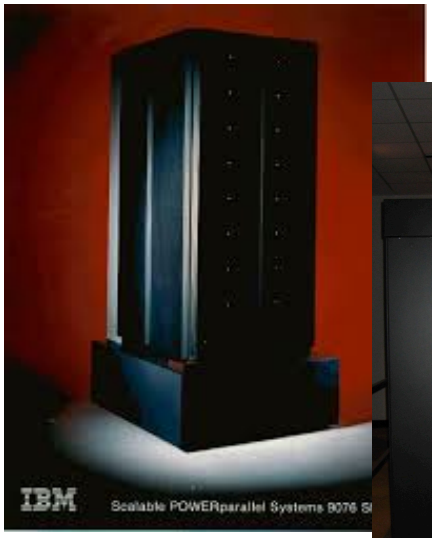
Bunch of Chickens

Log Performance

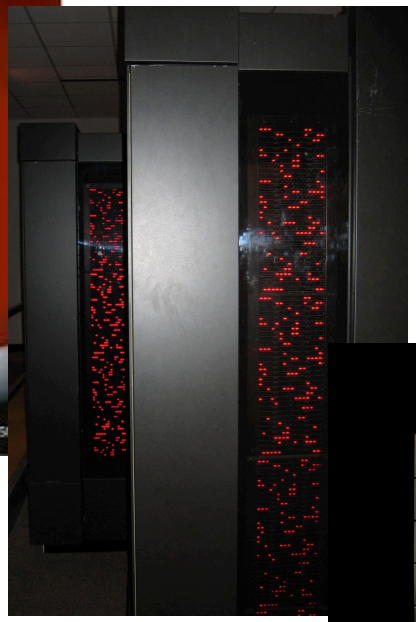
Years

Decade of Innovation

Intel Touchstone Delta

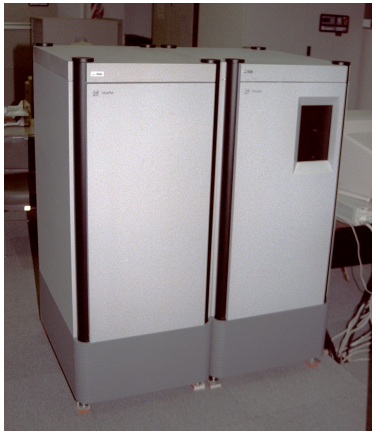


IBM SP1



TMC CM5

Convex SPP



Maspar

Cray T3D

Today's Linux Server Clusters

- **Cheap Hardware**
  - Commodity volumes
- **Free software**
  - If you can install it yourself

Beowulf at Caltech

Today's Linux Server Clusters

- **Cheap Hardware**
  - Commodity volumes
- **Free software**
  - If you can install it yourself
- **Often inefficient**
  - Memory wall
  - Message passing

Beowulf at Caltech

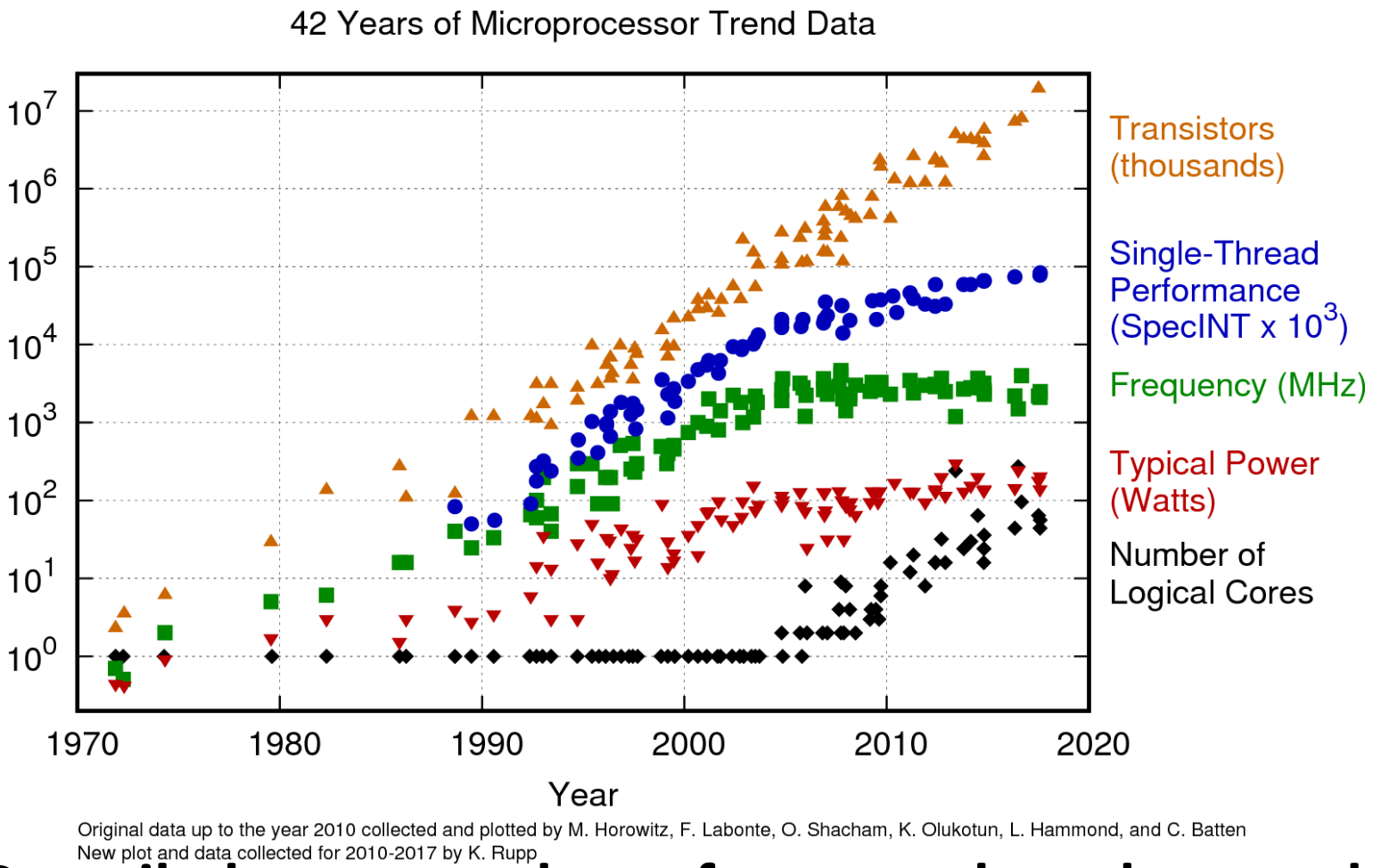
Today's Linux Server Clusters

- **Cheap Hardware**
  - Commodity volumes
- **Free software**
  - If you can install it yourself
- **Often inefficient**
  - Memory wall
  - Message passing
- **Just Buy More** …
  - Amazon buys it by the acre
  - Cloud service model
  - People are expensive

Beowulf at Caltech

Northern VA datacenter

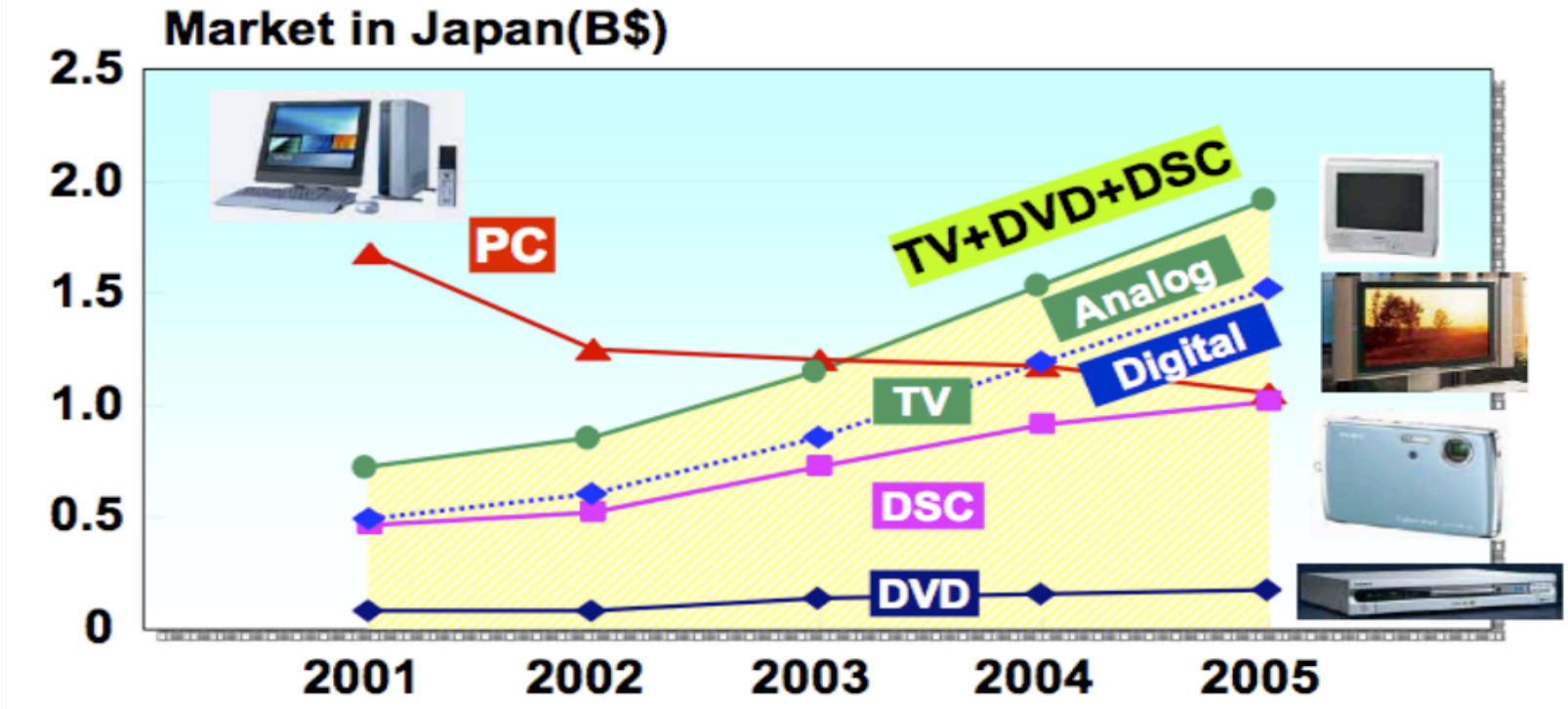
Another technology transition is happening



- **Compiled source code performance has plateaued**
  - Libraries still benefit from wider SIMD ALUs

Déjà vu?

- **John Shalf saw Tsugio Makimoto's talk at ISC '06**
  - PCs haven't been the low-cost, high-growth market for a decade!
  - Are they today's mainframes, ready to be undercut by cheaper H/W?



John's Vision of SoC for HPC (circa 2008)

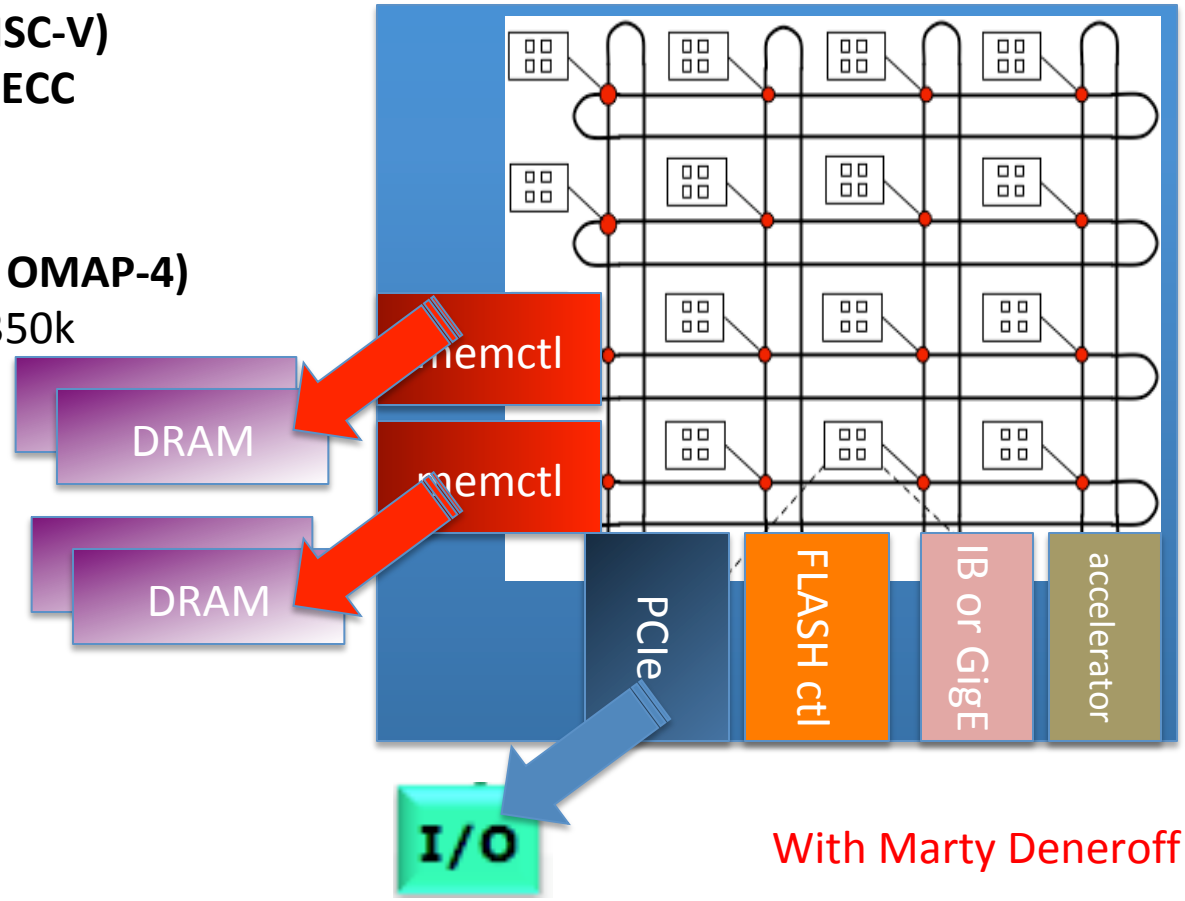
Processor Core (ARM, Tensilica, RISC-V)
With extra "options" like DP FPU, ECC
IP license cost $0-$500k

NoC Fabric: (Arteris, Denali, other OMAP-4)
IP License cost: $200k-$350k

DDR3 1600 memory controller
(Denali / Cadence, SiCreations)
+ Phy and Programmable PLL
IP License: $250-$350k

PCIe Gen3 Root complex
IP License: $250k

Integrated FLASH Controller
IP License: $150k



With Marty Deneroff

10GigE or IB DDR 4x Channel
IP License: $150k-$250k

Accelerator?

Innovative accelerators for AI

- **Low-precision arithmetic for ML**
  - Volta & TPU
- **Neuromorphic for vision**
  - True North
- **Annealers for optimization**
  - Fujitsu & D-Wave
- **Startups**
  - Cerebras, EMU & SambaNova

Overview

- **déjà vu, all over again?**
  - I'm reminded of my youth
- **Software Perspective**
  - It'll be harder this time around
- **Co-design**
  - Let's get the systems we need

Value of software

- **Hardware fabrication largely automated**
  - O($1B) of NRE can be amortized over a big production run
- **Software is more of an art form, or craft**
  - Productivity is O(1) SLOC per hour
  - LS-DYNA represents over 1M hours of labor
- **Software is often more valuable than hardware**
  - GM used to have O(10K) IBM POWER processors
  - They spent more money on software licenses
- **Trust and acceptance by user's is priceless**

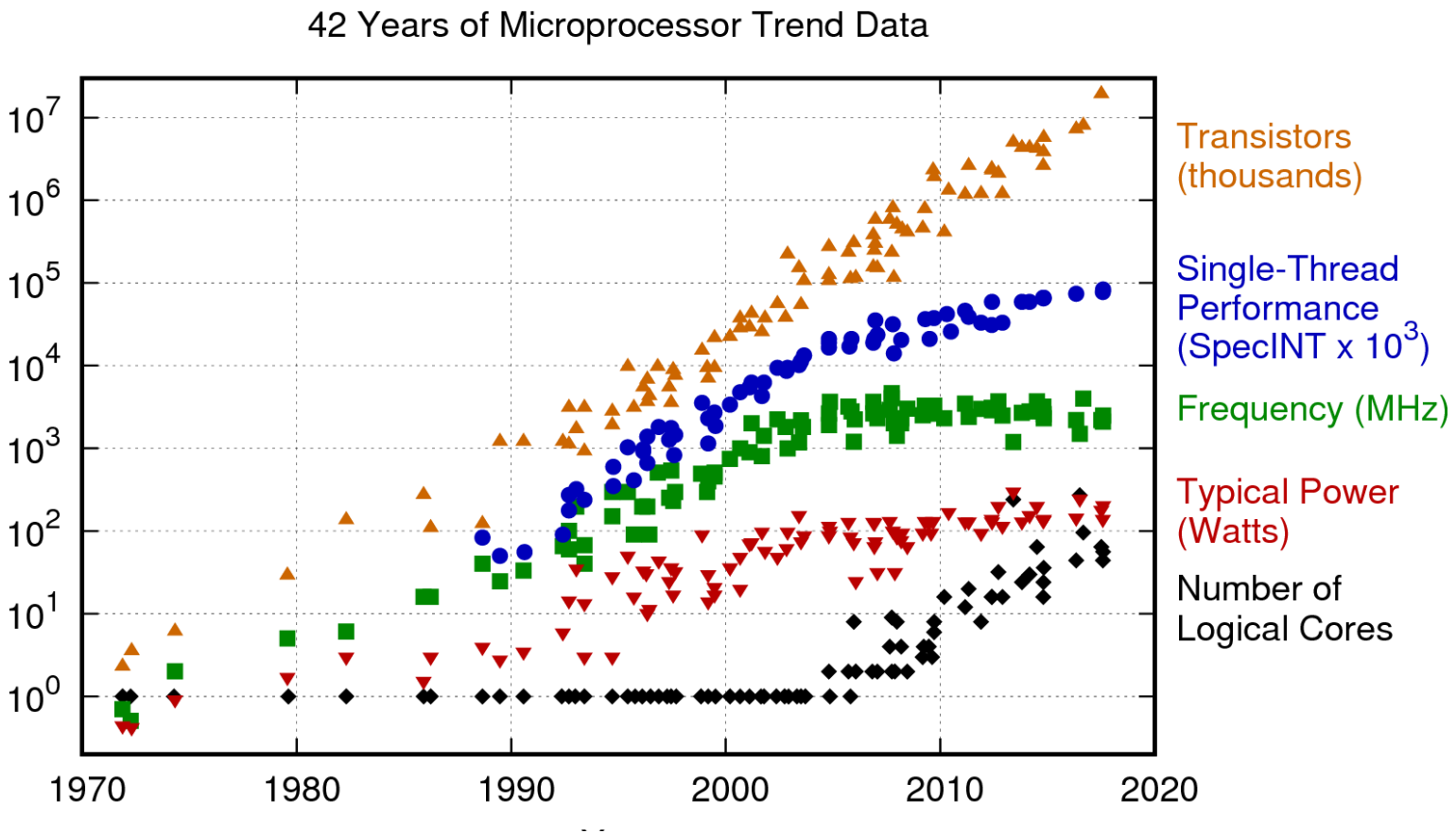
Its hard to displace existing software

- **Math, science, and engineering predate computing**
  - We've had seven decades to build codes for them
  - Modifying existing codes is often the easiest path forward
    - And it preserves existing investment
  - Initially successful ASCI burn codes predated ASCI
- **Imagine displacing Windows, Google, or Facebook**
  - Hundreds of millions of SLOCs
  - You'd need to replicate their data too
- **Easier to complete in a new marketplace**
  - Apple outflanked its competitors with iPod and iPhone
  - Today's innovation in AI

Disruption in application software

- **Nevertheless, sometimes new software is required**
  - New research problems or markets (e.g., ML today)
  - New hardware or software technology to exploit
    - LLNL's brand-new DYNA3D was rewritten for the Cray 1 in 1979
    - Makoto Asai created Geant 4 to use new software technology, C++
- **Successful software then evolves**
  - LLNL's DYNA3D was adopted by commercial companies
    - Like NASTRAN, SPICE, and many other public or academic codes
  - LS-DYNA is now O(10M) lines of source code
    - Primarily F77, but increasingly F95 and C
  - LSTC started working on message passing in 1993
    - Before MPI released

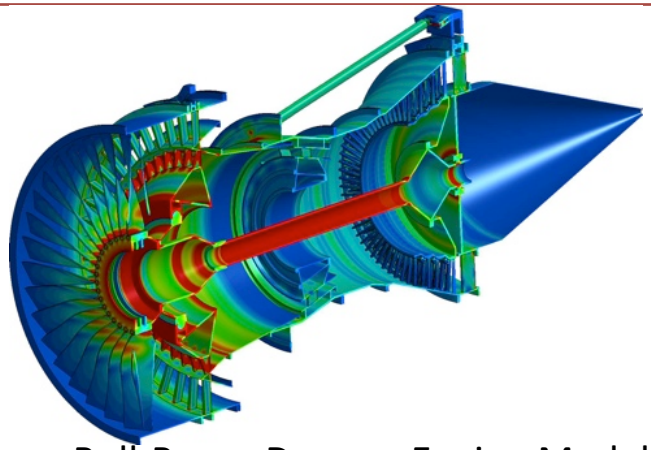
Exponential growth of demand



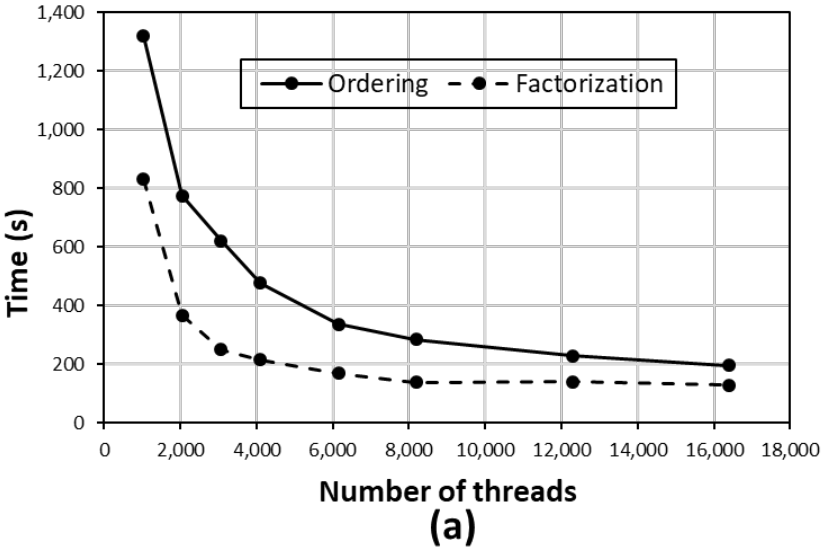
"… the current code is limited to 4096 processes so I cannot run the job up to the 96k cores I wanted to."

Scaling LS-DYNA

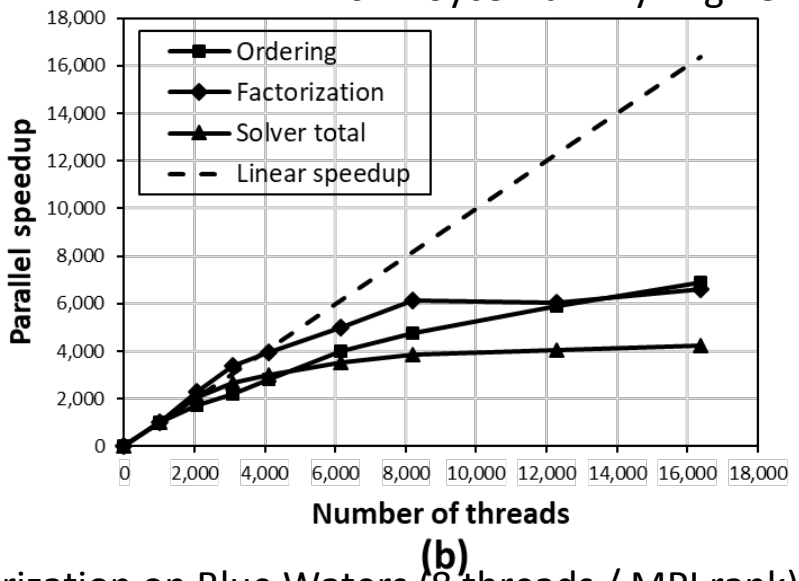
- **Users' requirements are unbounded**
  - Rolls-Royce wants virtual certification
  - Models are O(100M) elements, and growing
- **Trying to scale LS-DYNA accordingly**
  - Working with Cray, NCSA, and Rolls-Royce



Roll-Royce Dummy Engine Model



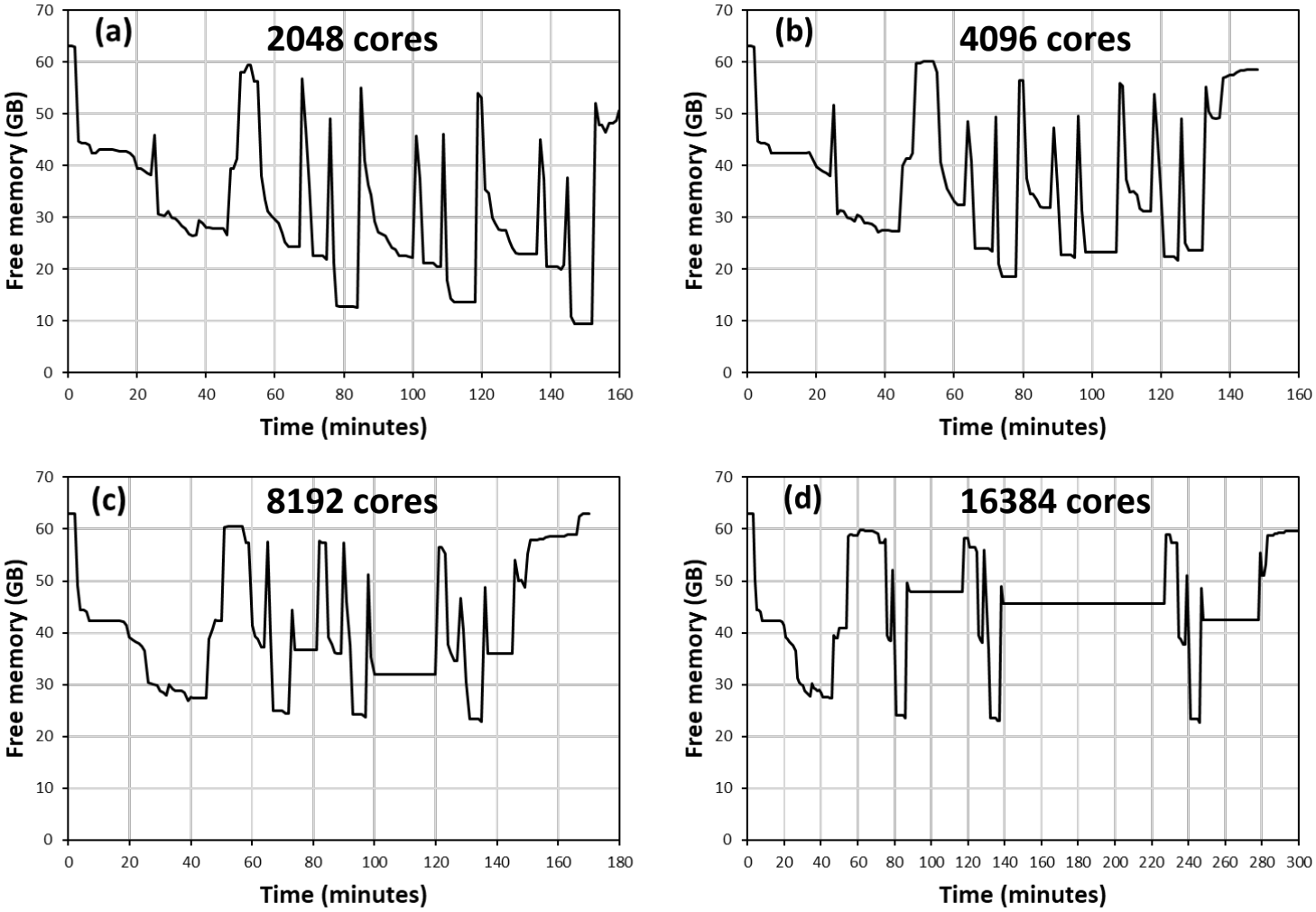
(a)



(b)

LS-DYNA sparse matrix reordering and factorization on Blue Waters (8 threads / MPI rank)
Figures courtesy of Erman Guleryuz (NCSA)

2017 LS-DYNA behavior vs cores

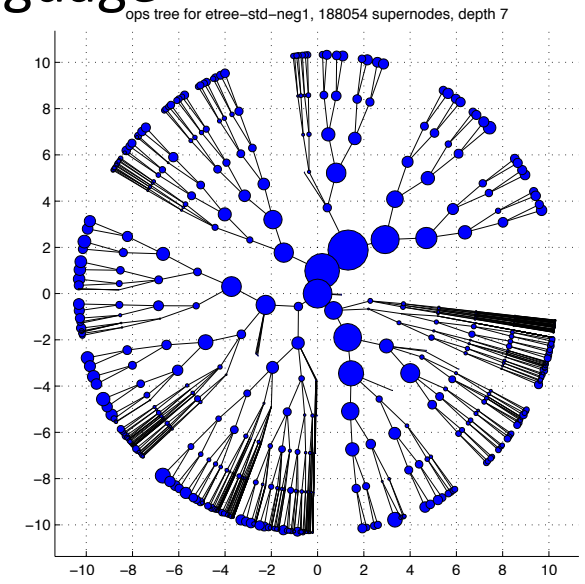


Available memory on MPI rank 0 while running 105M DOF engine in LS-DYNA on Blue Waters
Figures courtesy of Erman Guleryuz (NCSA)

Bounds on change

mpirun −np 2048 mppdyna i=control.k ncpu=8

- **Change needs to come from "below"**
  - Libraries written in C can be linked with Fortran
    - E.g., Metis, MPI and MUMPS
  - OpenMP is a graceful extension of the language
    - Brings back happy memories of autotasking
    - UPC too
- **Fix what's broken**
  - Multifrontal elimination tree is a DAG
  - Traverse branches with ICL's PaRSEC?
    - #3 on my LSTC to do list



Heterogeneous nodes

- **Floating point accelerators (APU, GPU)**
- **I'm 0:1 at integrating NVIDIA GPUs into LS-DYNA**
  - There're lots of DGEMM calls in a multifrontal code
  - First to receive the CUBLAS. Perhaps too early?
- **ANSYS succeeded**
  - Reverse engineered our experiment (twice!)
  - Their users' models are different than LSTC's (solids)
  - They've gone beyond me, exploiting GPU memory B/W too
- **Time for another look**
  - CUDA Fortran this time around
  - NVIDIA's helping

Overview

- **déjà vu, all over again?**
  - I'm reminded of my youth
- **Software Perspective**
  - It'll be harder this time around
- **Co-design**
  - Let's get the systems we need

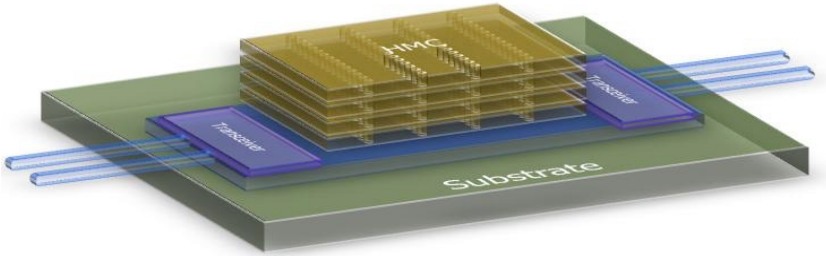
Managing complexity

- **The ground is shifting underneath us**
- **Software approaches to shield developers**
  - Domain Specific Languages
  - Kokkos
  - LANL's Ristra
- **Why not an equivalent hardware effort?**
  - Other half of co-design
  - Lower the bar for the software people

Unique role for HPC community

- **Peter Kogge's EXACUBE processor-in-memory**
  - IBM had 80386 IP
  - IBM had DRAM fabs
    - Several of them
- **Today, only Samsung makes both logic and memory**
- **Specialization constrains imagination**
- **Logic already exists in the memory**
  - What should Micron add?
  - How would it impact the host?

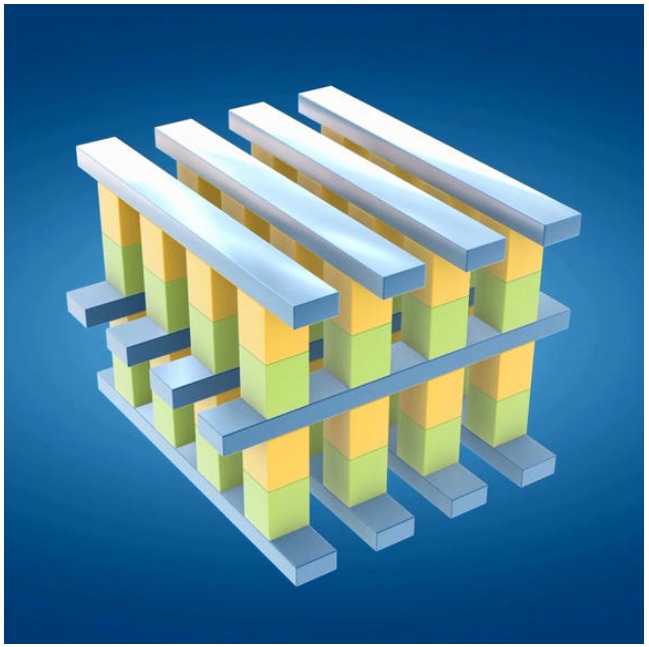
Hybrid Memory Cube



Self Test, Self Repair, Scrubbing, Refresh, Autonomous Functions

Steer the integration of new technology

- **Intel's Apache Pass**
  - 3D Xpoint in a DDR4 form factor
- **Multiple configurations offered**
  - High bandwidth file system
  - High bandwidth swap space
  - Direct user access
- **I want "door #3"**
  - Keep sparse matrix factors in 3D Xpoint
    - Eliminate file I/O abstraction (and overhead) for out-of-core
  - The rest of LS-DYNA can live in DDR4
    - Or better, HBM

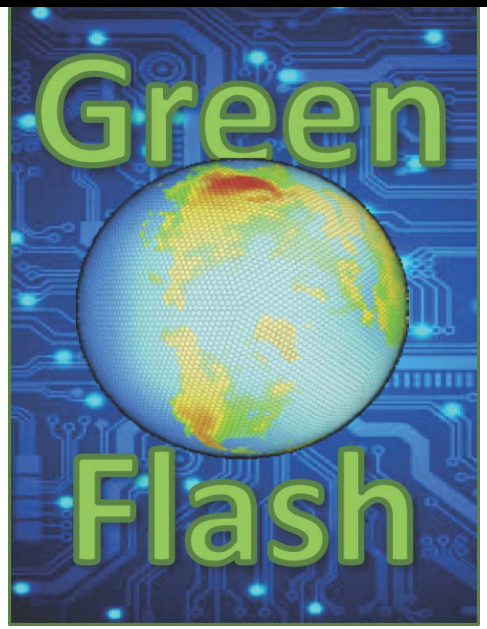


3D Xpoint illustration

SoC ecosystem provides a big knob

- **Anton**
  - Two orders-of-magnitude more capability

- **Green Flash**
  - Two orders-of-magnitude less energy



Bespoke systems

- **We have the talent to exploit SoC technology**
  - Spread thinly over the HPC user community
  - LBL's David Donofrio designed chips for Intel and Apple
  - CCS's Bill Carlson hacked GNU C to create the initial UPC
  - USC's Jeff Draper got a sole source award from DARPA
- **System vendors certainly do**
  - Still HPE's business model?
    - It was before Greg Astfalk retired
  - Sunway can too
    - TaihuLight

Bespoke systems

- **We have the talent to exploit SoC technology**
  - Spread thinly over the HPC user community
  - LBL's David Donofrio designed chips for Intel and Apple
  - CCS's Bill Carlson hacked GNU C to create the initial UPC
  - USC's Jeff Draper got a sole source award from DARPA
- **System vendors certainly do**
  - Still HPE's business model?
    - It was before Greg Astfalk retired
  - Sunway can too
    - TaihuLight
  - Low-volume business model
    - Sell engineering, not chips

B2 NRE ~2X production cost

Address things that limit performance

- **Replacing Pentiums with "free" RISC-Vs won't be enough**
  - Only free until you fab them
- **Distributed address space**
  - I want my E-registers back
    - UPC-like global memory abstraction
    - Surely the patents have expired by now
- **Virtual memory hierarchy**
  - Non-unit stride, gather/scatter, indirect addressing, etc.
    - Utah's Impulse project
  - Si Hammond's store with floating point accumulate
    - Perform "one-touch" functions in the memory
- **Synchronization**
  - I have lots of 8-byte MPI_ALLREDUCEs for error status
  - BlueGene/L had a combining network

Make SoC ecosystem HPC friendly

- **Create IP specifically for HPC**
  - Our needs will differ from commodity markets
- **Fill software gaps too.**
  - LANL reportedly finds ARM's Fortran environment lacking
- **Invest in ECAD R&D**
  - Reduce the cost of engineering SoC systems
  - Andres Olofsson is working this
- **Enduring advantage**
  - GOTS IP, not available to others

Summary

- **There's still "gas in the tank" for CMOS**
  - Rich Linderman's phrase
  - Specialization offers a path forward, beyond Moore's Law
- **Evolution**
  - HPC software is often more valuable that the hardware
  - The pace of change must allow for adaptation
- **Collaborate**
  - Science and engineering have much in common
  - One destructive engine test is O($100M)
    - Roughly the same cost as Anton