Improving HPC Software

Pete Beckman & Jack Dongarra

Outline

- Current State: HPC Software
- Background: Activities in Europe and Japan
- The Changing Architecture
- The IESP Workshops
- Roadmap and Outcomes

The Open Source Community Provides Most of the World's HPC Software

Jaguar	Total	XT5
Peak Performance	1,645	1,382
AMD Opteron Cores	**181,504**	**150,176**
System Memory (TB)	**362**	**300**
Disk Bandwidth (GB/s)	**284**	**240**
Disk Space (TB)	**10,750**	**10,000**
Interconnect Bandwidth (TB/s)	**532**	**374**

U.S. Department of Energy
Office of Science

National Energy Research Scientific Computing Center (NERSC)

- Located at Lawrence Berkeley National Lab
  - Cray XT-4 Franklin: 102 Tflop/s, 9,660 nodes, 19,320 cores
  - IBM Power 5 Bassi: 6.7 Tflop/s, 888 cores
  - Linux Opteron Cluster Jacquard: 3.1 Tflop/s, 712 cores
- Franklin quad-core currently being upgraded 350 Tflop/s, 38,640 cores
- NERSC-6 Project
  - RFP issued in September 2008
  - Proposals are being reviewed

Franklin

Argonne's IBM Blue Gene/P – 556 TFs

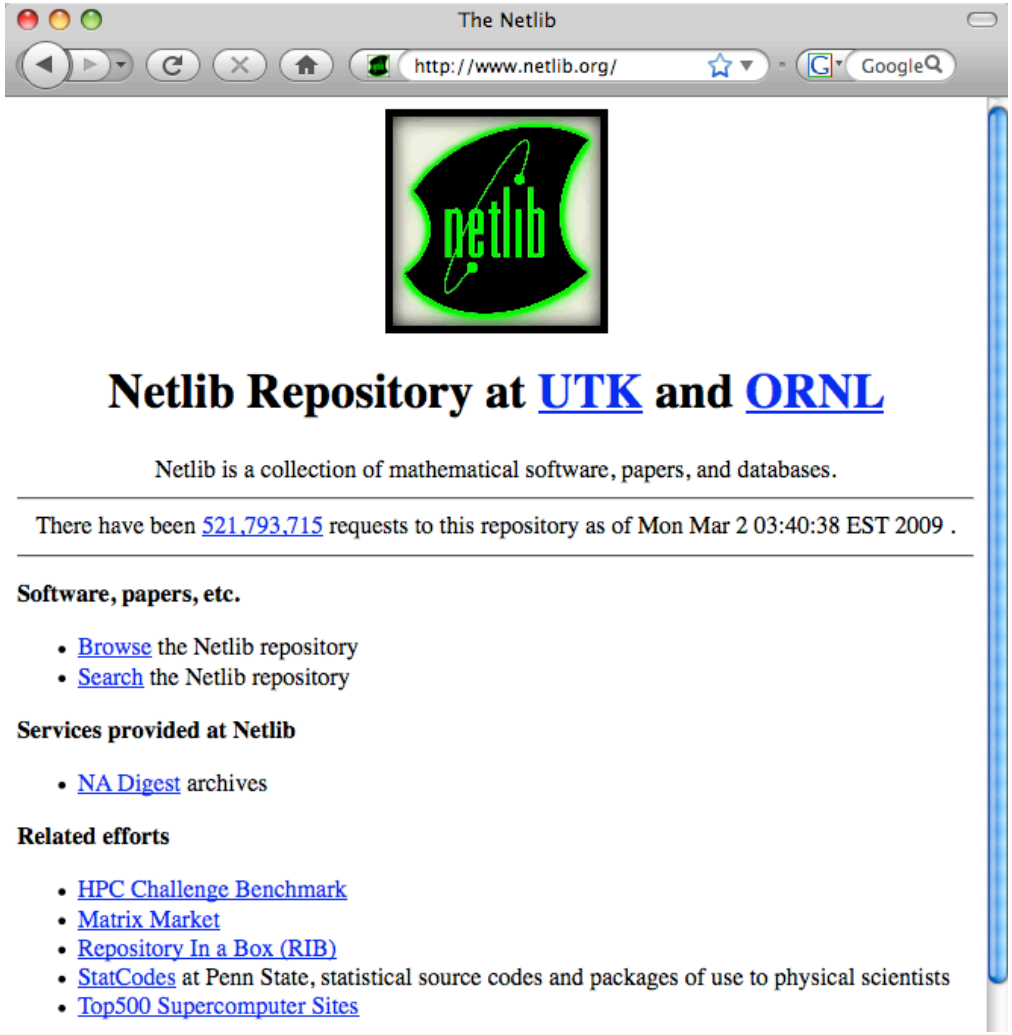
The Community is Diverse and Robust

- Over the last 10 years, the galvanization of the Open Source movement has dramatically improved HPC software

A very small sample:

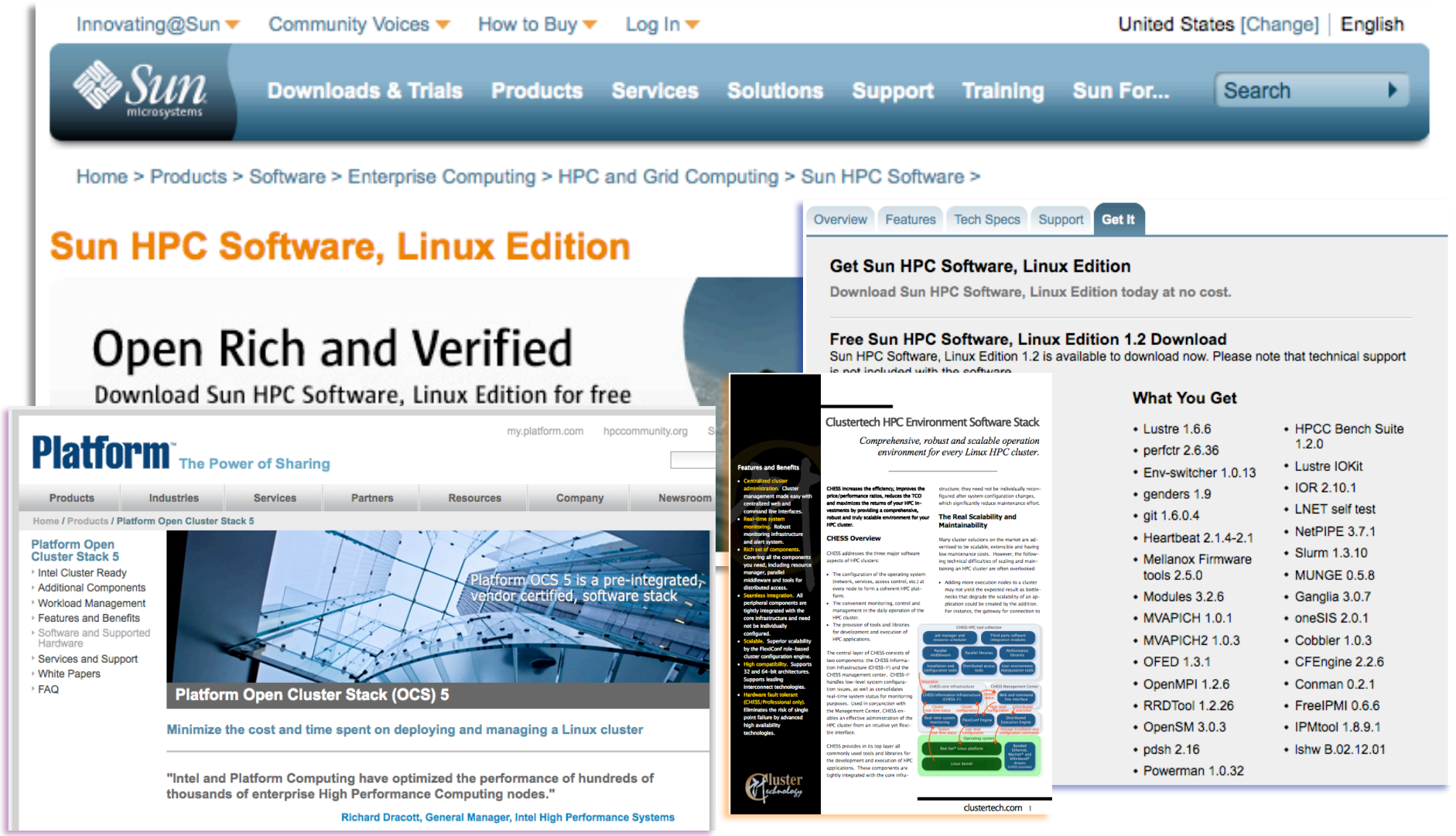
- Linux Operating System, libc
- Python, Perl
- PAPI, TAU, Kojak
- UPC
- MPICH, OpenMPI
- ScaLAPACK
- VisIt
- GASNet, ARMCI/GA
- PVFS
- CFEngine, bconfig
- Ganglia
- SLURM, Cobalt
- Dyninst
- Torque/Moab, OpenPBS
- Charm++
- pNetCDF, HDF5
- GridFTP
- FFTW

A Long History of Collaboration



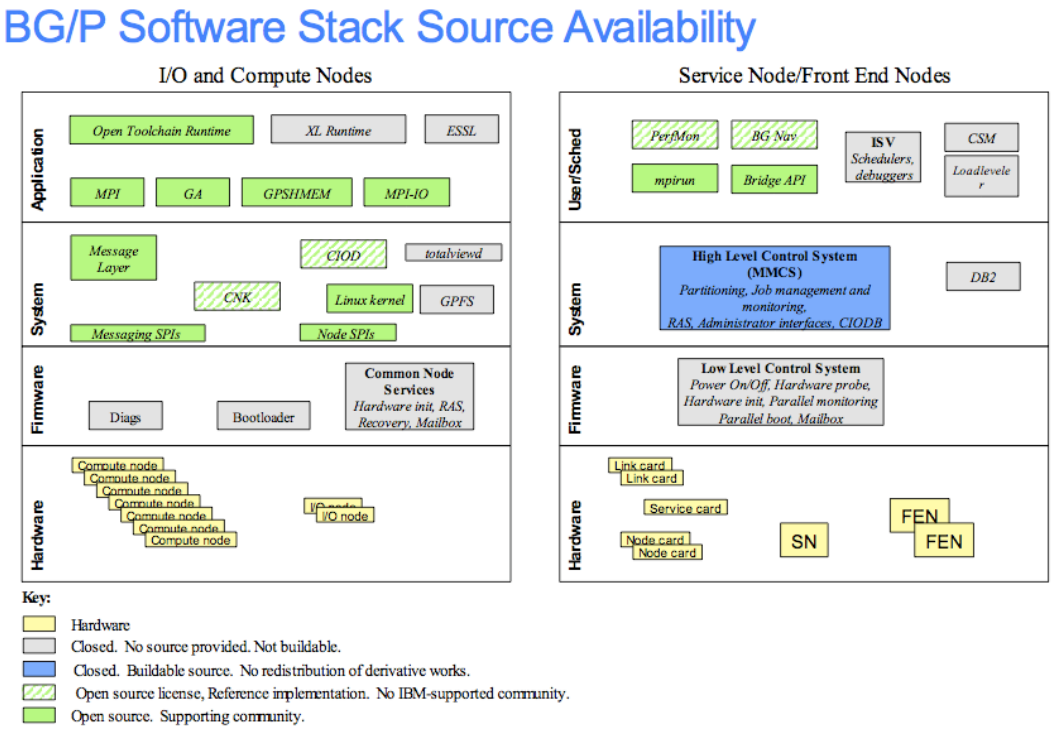
The Result….

Open Source HPC Software Stacks for Small Linux Clusters are Everywhere



Got Scale?

- For some markets, a closed source business model continues to work well
  - Single-node optimized math libraries & compilers
  - Debuggers for small clusters
  - Some queuing systems, parallel file systems, HSMs
  - Small cluster applications: Fluent, CFD++, etc



Why Seek to Improve This?

- The largest scale systems are becoming more complex, with designs supported by large consortium
  - The software community has responded slowly
- Significant architectural changes arriving
  - Software must dramatically change
- Our ad hoc community coordinates poorly, both with other software components and with the vendors
  - Computational science could achieve more with improved development and coordination

Extreme-Scale Platform Design:

Industrial revolution and globalization has arrived

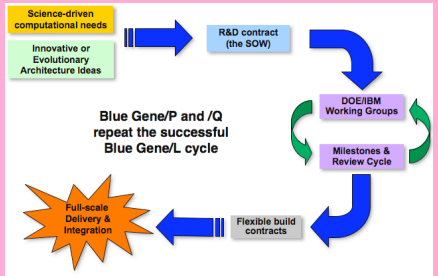
Yesterday **Today** **Tomorrow**

Seymour & team designs and hand builds set of computers

Dozen HPC companies flourish: incompatible OS & components

Commodity components and Open Source move effort to integration

Design-Build partnerships for extreme machines (e.g. LLNL/ANL/IBM)



Japan 10PF: Hitachi, NEC, Riken, Universities

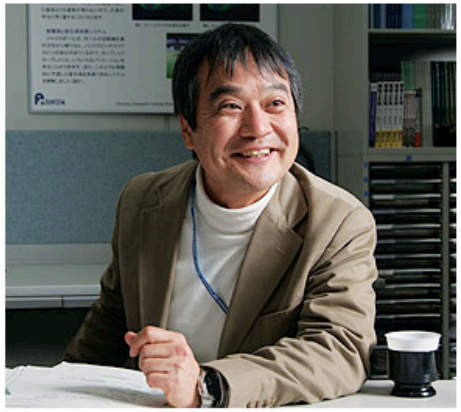
Globally Distributed teams, Diverse technology providers, Open Source Software

Mar 2012… the Japanese "> 10 PF" SC to be Online

25~30MW Power
~1mil ft2 floorspace
$1 bil construction

Kobe, Japan
"the site"

Interview with Ryutaro Himeno, Dr. Eng.
Development Group Director at the Next-Generation Supercomputer R&D Center



HIMENO Ryutaro

A 10-petaflop* supercomputer will be built in 2012. The supercomputer will boost a computing speed that is 50 times faster than the world's current fastest computer. RIKEN is responsible for the development, construction, and operation of the supercomputer, which is a huge government project with a total budget of 110 billion yen including facilities for the system and the research grid project. We interviewed Dr. Ryutaro Himeno, Development Group Director at the newly-established Next-Generation Supercomputer R&D Center to ask about why a speed of 10 petaflops, and what they are hoping to achieve with a 10-petaflop supercomputer.

Home
About PRACE
Activities
Use cases
Documents
Press corner
HPC Training
Contact us

PRACE newsletter

Your e-mail address

HTML
Text
Subscribe

Welcome to PRACE

The Partnership for Advanced Computing in Europe prepares the creation of a persistent pan-European HPC service, consisting of several tier-0 centres providing European researchers with access to capability computers and forming the top level of the European HPC ecosystem. PRACE is a project funded in part by the EU's 7th Framework Programme.

Supercomputers are indispensable tools for solving the most challenging and complex scientific and technological problems through simulations. To remain internationally competitive, European scientists and engineers must be provided with leadership-class supercomputer systems. PRACE, the Partnership for Advanced Computing in Europe will create a persistent pan-European high performance computing (HPC) service and infrastructure. This infrastructure will be managed as a single European entity. European scientists and technologists will be provided world-class leadership supercomputers with capabilities equal to or better than those available in the USA and Japan. The service will comprise three to five superior HPC centers strengthened by regional and national supercomputing centers working in tight

News

» HPC Infrastructures for Petascale Applications - DEISA PRACE Symposium 2009 2009-02-25
» PRACE hosts highly successful Winter School 2009-02-25
» PRACE held All Hands Meeting in Jülich, Germany, February 12-13, 2009 2009-02-16
» Serbia joins the PRACE initiative 2009-02-12
» PRACE Part of Zero-In Magazine - Call for Papers Open 2009-02-03

more...

Events

» OGF25 / EGEE User Forum, 2-6 March, Catania, Italy
» 24th Forum ORAP, 26 March, Lille, France

Traditional Sources of Performance Improvement are Flat-Lining (2004)

- New Constraints
  - 15 years of *exponential* clock rate growth has ended
- Moore's Law reinterpreted:
  - How do we use all of those transistors to keep performance increasing at historical rates?
  - Industry Response: parallelism doubles every 18 months *instead* of clock frequency!

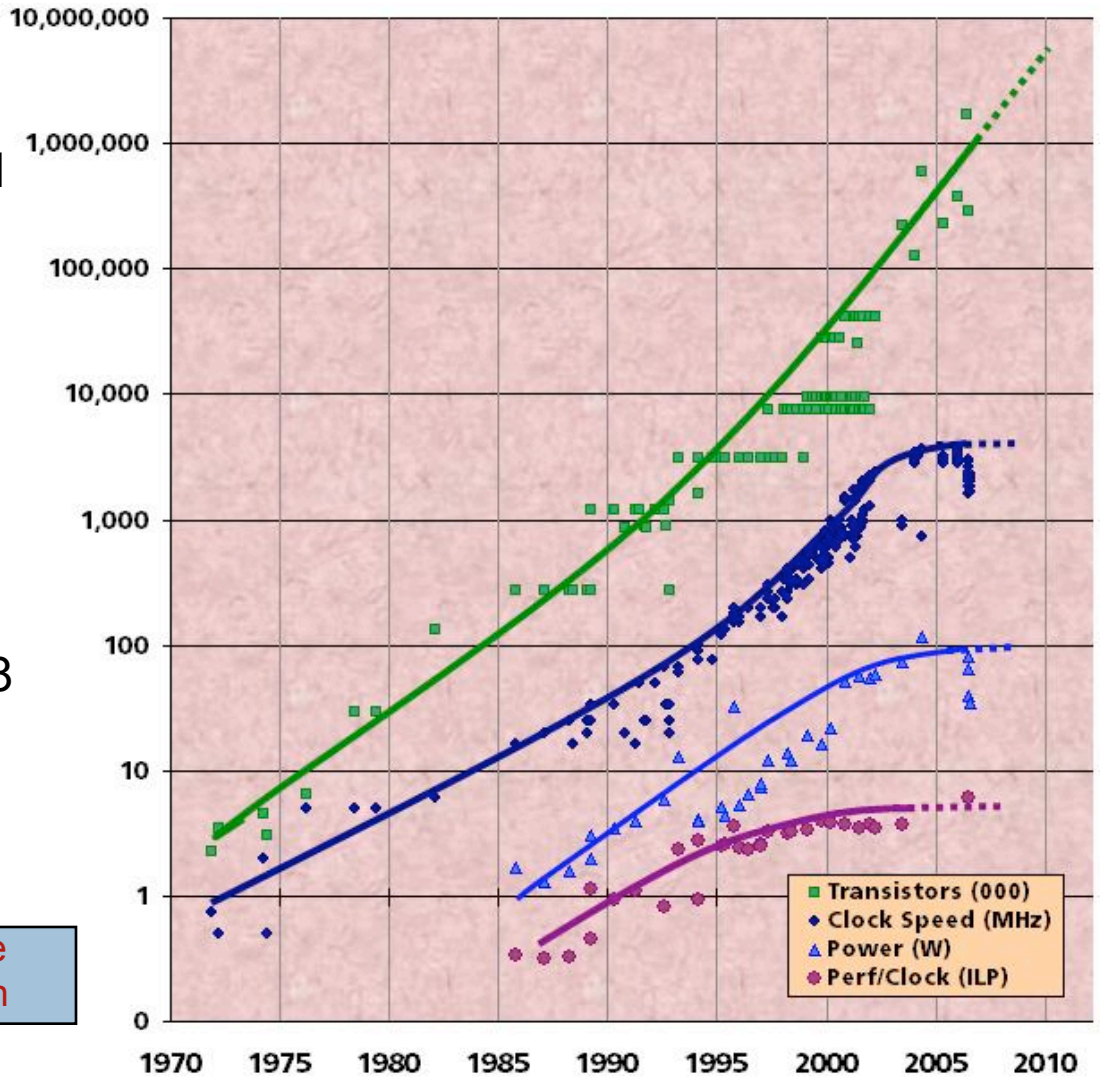
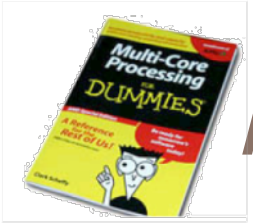
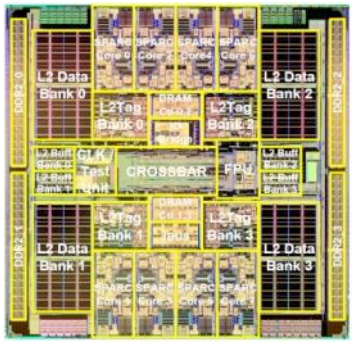


Figure courtesy of Kunle Olukotun, Lance Hammond, Herb Sutter, and Burton Smith



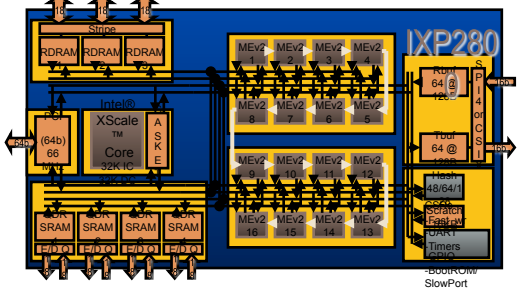
Multicore comes in a wide variety

- Multiple parallel general-purpose processors (GPPs)
- Multiple application-specific processors (ASPs)

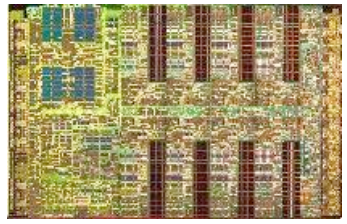


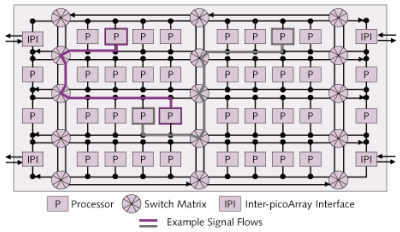
Sun Niagara
8 GPP cores (32 threads)

Intel Network Processor
1 GPP Core
16 ASPs (128 threads)

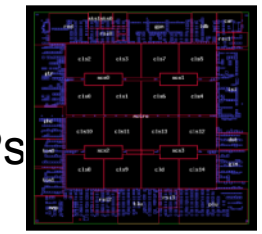


IBM Cell
1 GPP (2 threads)
8 ASPs

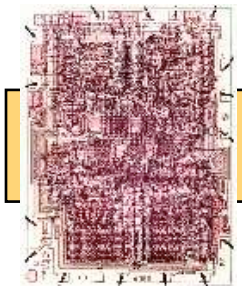




Picochip DSP
1 GPP core
248 ASPs



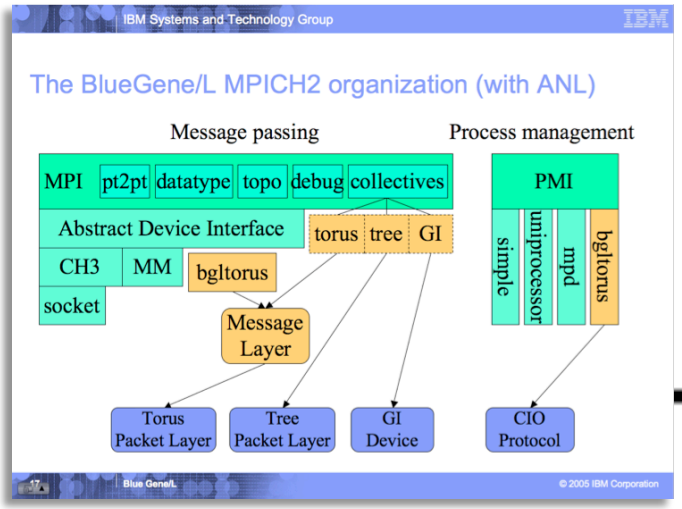
Cisco CRS-1
188 Tensilica GPPs

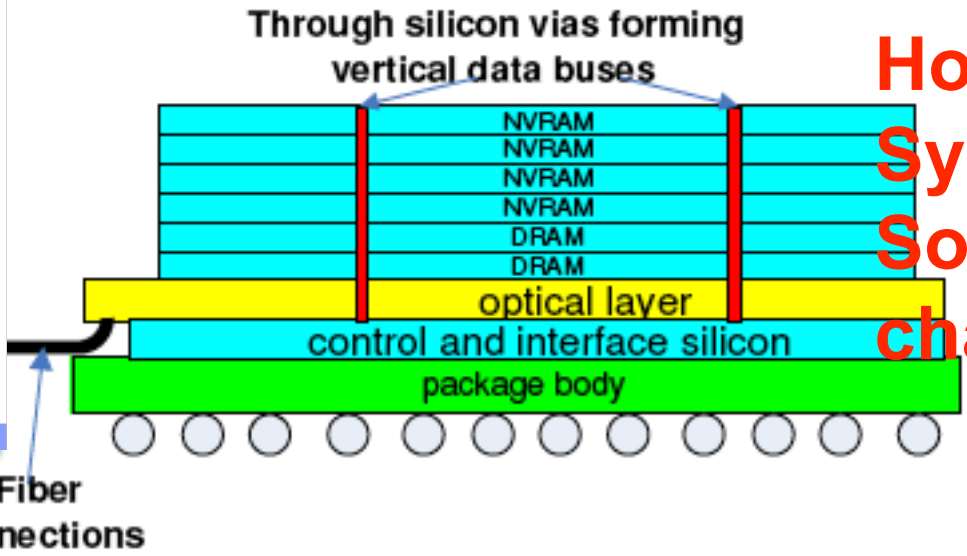


Intel 4004 (1971):
4-bit processor,
2312 transistors,
~100 KIPS,
10 micron PMOS,
11 mm^2 chip

"The Processor is the new Transistor" [Rowen]

3D Packaging: Changing Paradigms





How will System Software change?

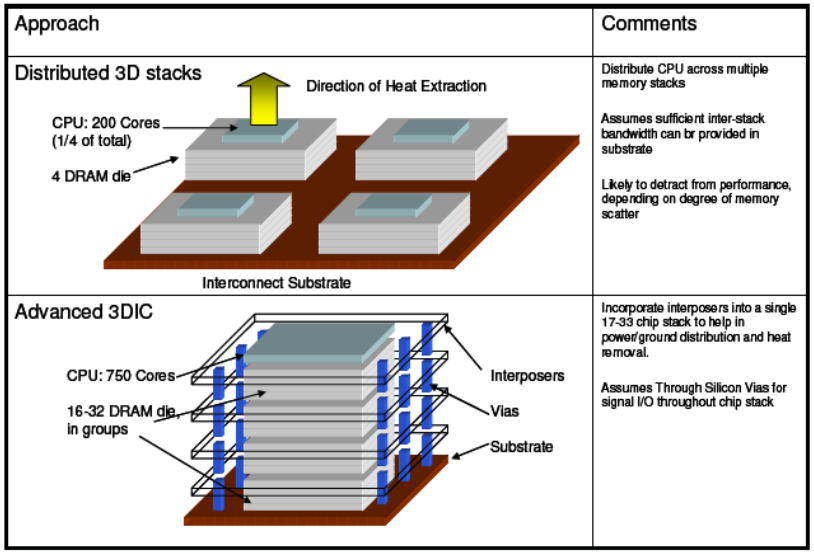


Figure 7.5: Potential directions for 3D packaging (A).

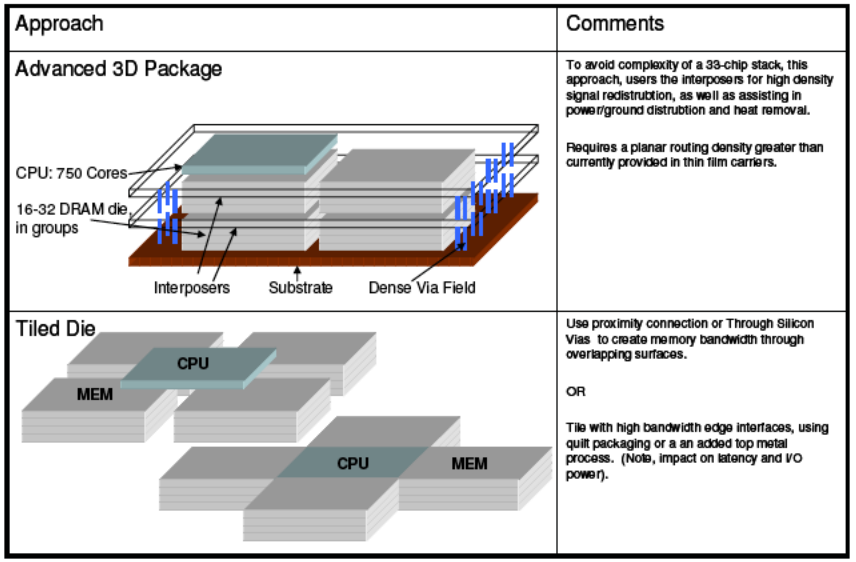


Figure 7.6: Potential directions for 3D packaging (B).

Where We Are Today:

We are not prepared for the changes coming

- Hardware features are uncoordinated with software development
  - (power mgmt, multicore tools, math libraries, advanced memory models, etc)
- Only basic acceptance test software is delivered with platform
  - UPC, HPCToolkit, Optimized libraries, PAPI, can be *YEARS* late
- Vendors often "snapshot" key Open Source components and then deliver a stale code branch
  - Counterexample: A model that works – MPICH for BG/P
- Community codes unprepared for sea change in architectures
- Coordination via SOW/contract is poor and only involves 2 parties
- No global evaluation of key missing components

The IESP Workshops:

- Goal: Improve the world's simulation and modeling capability by improving the coordination and development of the HPC software environment.
  - Build a plan for how the international community can join together to improve software available for high-end systems over the next 2 to 10 years.
- The DOE, NSF, and EU have committed their support for the workshops.
- The first workshop will be Santa Fe, April 7-8.
  - White papers encouraged

International Community Effort

- We believe this needs to be international collaboration for various reasons including:
  - The scale of investment
  - The need for international input on requirements
  - Europeans, Asians, and others are working on their own software that should be part of a larger vision for HPC.
- The process must be totally open

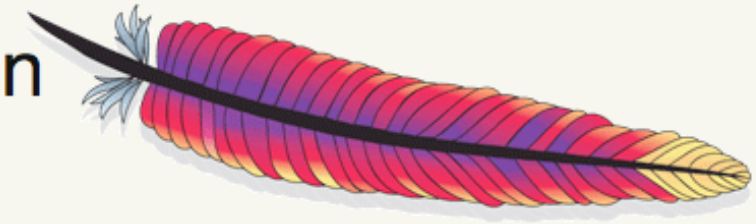
Executive Committee:

Co-Chair: Jack Dongarra, Univ, of Tennesse / ORNL, US
Co-Chair: Pete Beckman, Argonne National Laboratory, US
Franck Cappello, INRIA, FR
Thomas Lippert, Jülich Supercomputing Centre, DE
Satoshi Matsuoka, Tokyo Institute of Technology, JP
Paul Messina, Argonne National Laboratory, US

An Example Development Community

The Apache Software Foundation

Meritocracy in Action.



The Apache Software Foundation provides support for the Apache community of open-source software projects. The **Apache projects** are characterized by a collaborative, consensus based development process, an open and pragmatic software license, and a desire to create high quality software that leads the way in its field.

We consider ourselves not simply a group of projects sharing a server, but rather a *community of developers and users*.

This page will give you everything you always wanted to know about the foundation but were afraid to ask. The difference between membership and committership, who decides what, how elections take place, how is our infrastructure setup, what is the board, what is a PMC, what's the philosophy behind the incubator, why is the foundation moving away from project containment. Come and see behind the scenes of the ASF.

What is the Apache Software Foundation?

The Apache Software Foundation (ASF) is a 501(c)3 non-profit organization incorporated in the United States of America and was formed primarily to:

- provide a foundation for open, collaborative software development projects by supplying hardware, communication, and business infrastructure
- create an independent legal entity to which companies and individuals can donate resources and be assured that those resources will be used for the public benefit
- provide a means for individual volunteers to be sheltered from legal suits directed at the

Apache Foundation

- Create a foundation for open, collaborative software development projects by supplying hardware, communication, and business infrastructure
- Incubator projects can become Apache projects
- 800 “committers”
- The ASF Infrastructure is mostly composed of the following services:
  - the web serving environment (web sites and wikis)
  - the code repositories
  - the mail management environment
  - the issue/ bug tracking
  - the distribution mirroring system

A Plan Could Include:

- Work with vendors to create the HPC equivalent to the ITRS (Int'l Tech Roadmap for Semiconductors)
  - Get community working on software before machine becomes available
- Community proposed unified roadmap for exascale software
- Identify missing components for future architectures and a plan to address them
- Develop models for working more closely with vendors
  - (support, acceptance tests, target features)
- Identify key application areas to drive development
- Community software development models
- Funding and organizational models (Apache, etc)

Achievable Outcomes

- Improve the capability of computational science
- Build and strengthen international collaborations and leadership; deliver more capable, productive HPC systems
- Build and improve R&D program developing new programming models and tools addressing extreme scale
- Open source HPC development guided by roadmap with better coordination and fewer missing components
- Joint programs in education and training for the next generation of computational scientists.
- Vendor engagement and coordination for more capable software supporting exascale science

Workshops and Report

- 3 workshops over the next year
  - 1: Santa Fe, April 7-8
  - 2: France, week of July 20th
  - 3: Japan in the early Fall
- Broad engagement by the community
- Initial reports in summer 2009
- Final report for first year at SC09
- Planning for *IMMEDIATE* payoff
  - Could begin ramping up next year
- Stay tuned…

www.exascale.org

Home Meetings Documents Community Help

Main Page

Page Discussion View source History

Workshop Information

- Workshop Location
- Workshop Agenda (draft)
- Executive Committee
- Organizing Committee
- Background Material

The mission of the **International Exascale Software Project (IESP)** is to lay the foundation for exascale computing by mobilizing the global open source software community to combine and coordinate their collective efforts far more efficiently and effectively than ever before. The IESP will hold a series of three workshops to organize and structure this community wide effort. The first, invitation-only workshop will occur on April 7th and 8th in Sante Fe, New Mexico, US, with people arriving in time for a reception on April 6th. Attendees will include members from industry, academia, and government, with expertise in a range of critical areas.

Goals for the first meeting include the following:

- Assess the short-term, medium-term and long-term needs of applications for peta/exascale systems
- Explore how laboratories, universities, and vendors can work together on coordinated HPC software
- Understand existing R&D plans addressing new programming models and tools addressing extreme scale, multicore, heterogeneity and performance
- Start development of a roadmap for software on extreme-scale systems

Attendance at the workshop is by invitation only. Additional details on registration will be coming soon.