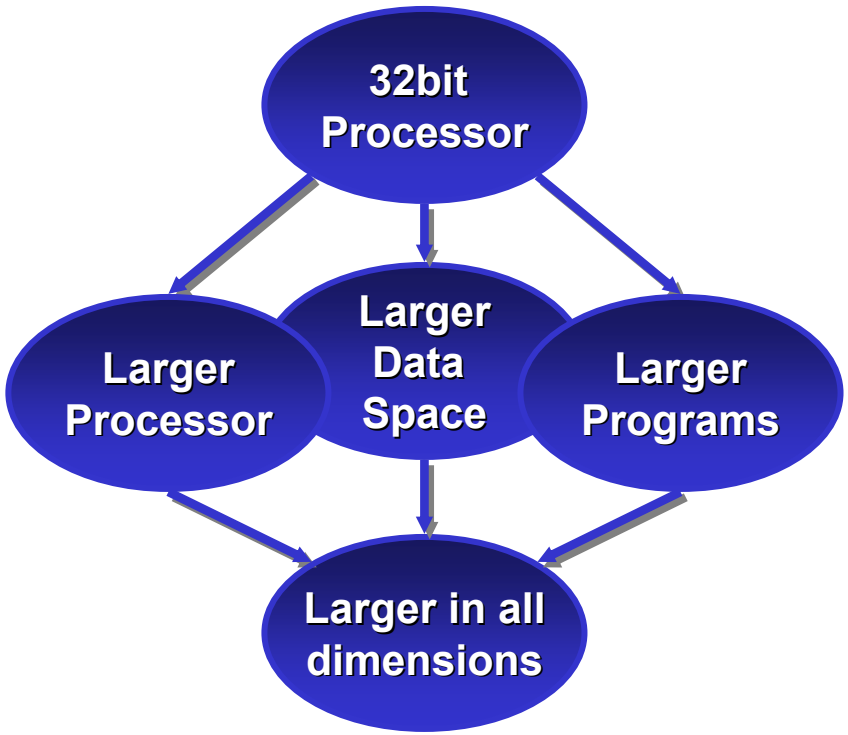
Itanium System Performance

Jan 17, 2003
Bob.Kuhn@intel.com

Outline – Intel Software Tools Improving Performance in 3 Dimensions



- What are the tools
- The data –

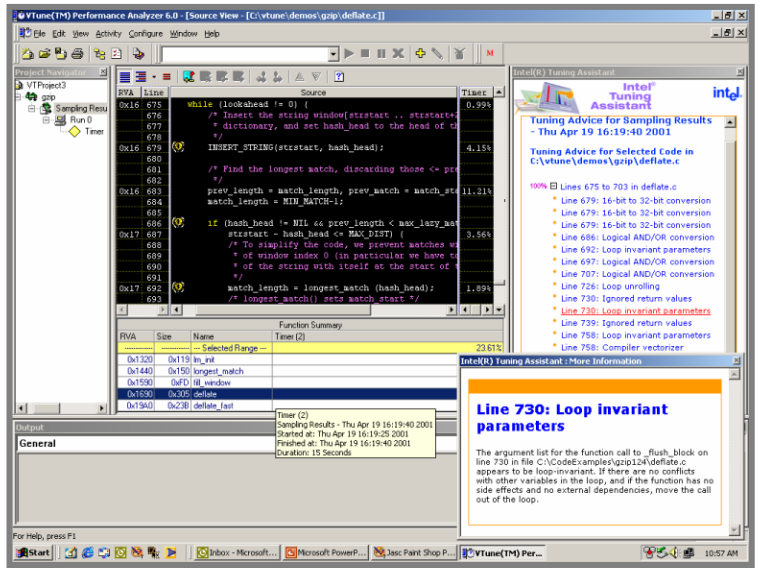
1. Larger processor
   - Itanium performance
2. Larger data space
   - SMPs
   - Interconnects
   - DVSM
3. Larger programs
   - Portals and DRMs
   - Multiphysics, Multichem

Software Tools from Intel

- VTune performance analysis
- Compilers
- Math libraries
- Thread Checker

Intel® VTune™ Performance Analyzer 6.0

- **Allows you to save time in the development cycle by quickly identifying "hot spots" for review**
- **Identifies performance bottlenecks in Source Code using three modes**
  - **1.) Sampling – events and time based**
  - **2.) Call Graph – presents program flow**
  - **3.) Counter Monitor – monitors process against the CPU**
- **Supports Intel Processors**
- **New Features:**
  - **Linux remote data collectors (32- and 64-bit)**
  - **Support for .Net and C#***
  - **Enhanced Intel Tuning Assistant**
    - **Provides system and source level tuning insights based on OS and CPU performance counters**



New Feature: Intel® Tuning Assistant

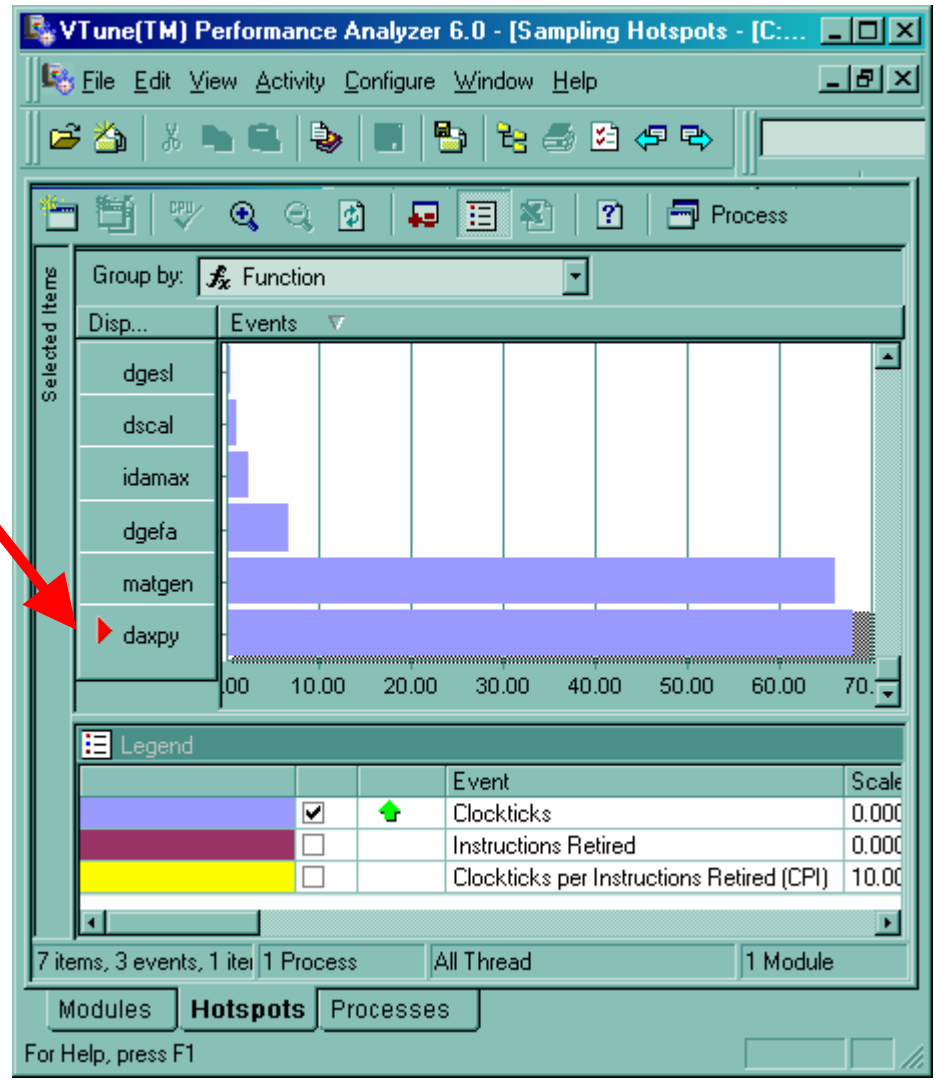
www.intel.com/software/products

Hotspot Drill Down

Most Samples within daxpy

Use in combo with gprof for better understanding

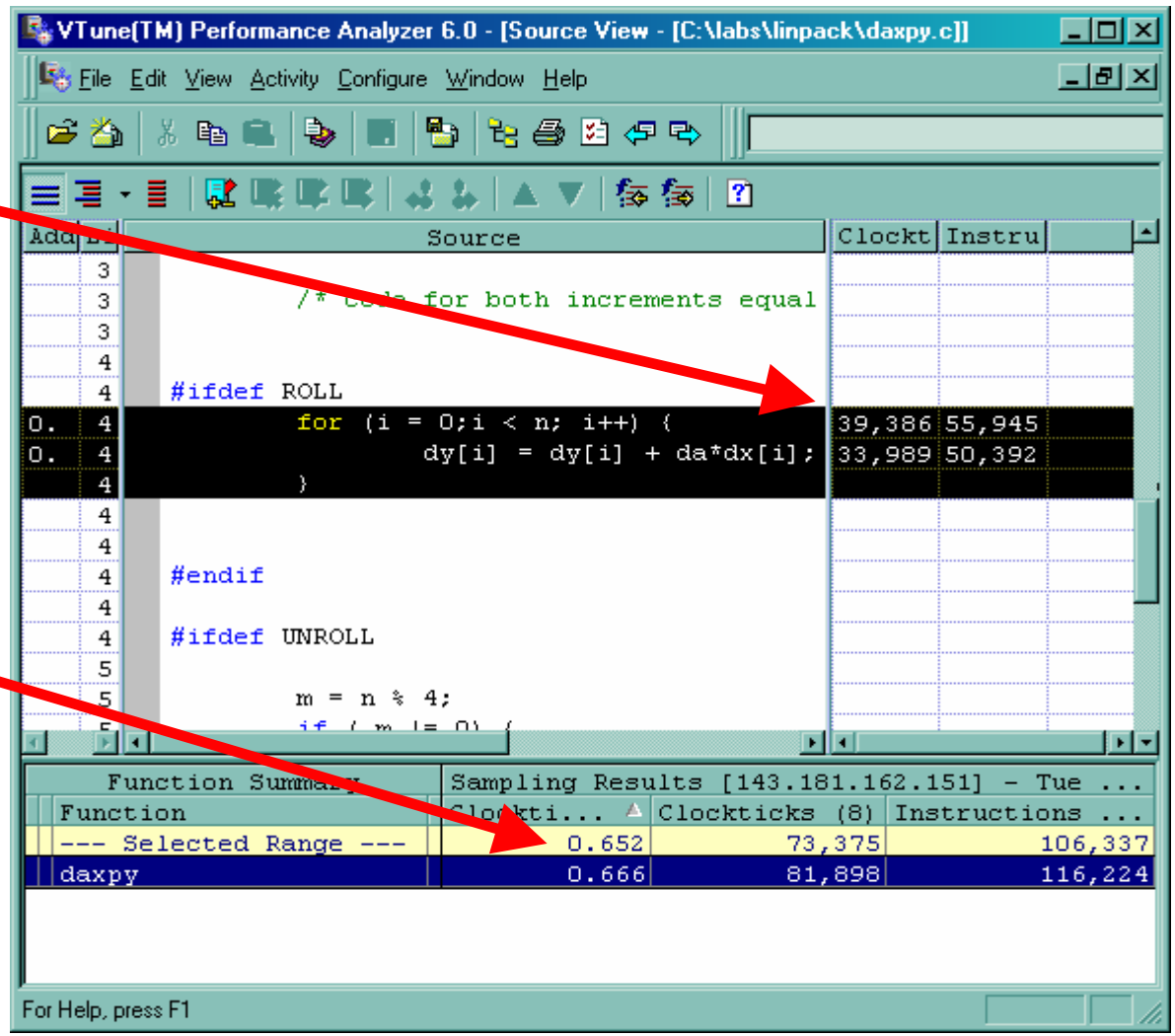
Call Graph Coming Soon!



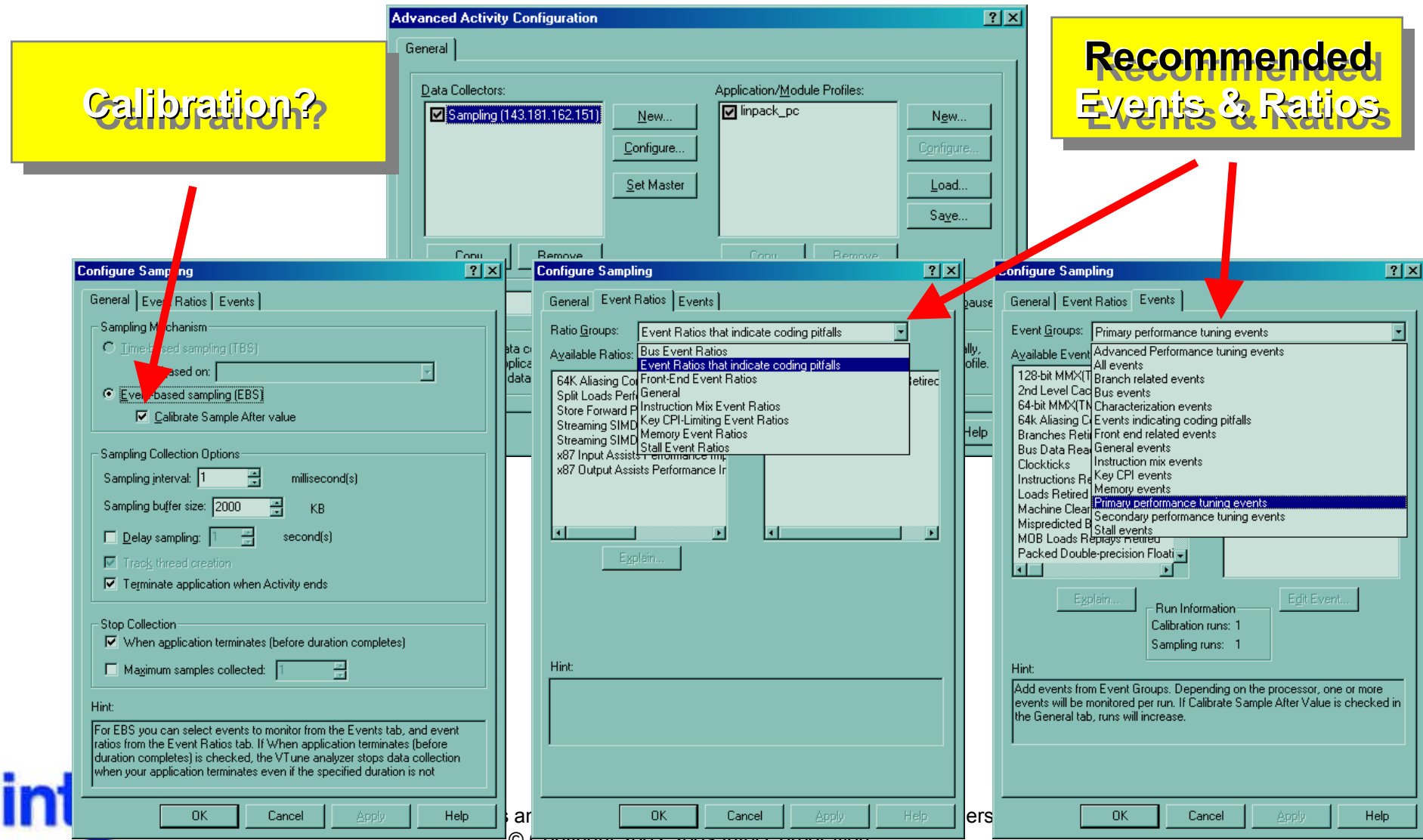
Source Level View

HotSpot

Efficiency?



Advanced Sampling



7.0 Compiler Features

- Itanium 2 Performance Features
- Fortran Bridge Product Features: addresses the market requirements for subset of Visual Fortran features
    - Intel Fortran VS.Net IDE integration*1
    - Fortran Source Code Colorizer*1
    - CVF Command line options (Subset)
    - CVF Non-core libraries compatibility
    - Array Visualizer for Win32 C/C++ and Fortran
    - VF reporter for Windows

1 These are Windows features that depend on Visual Studio.NET

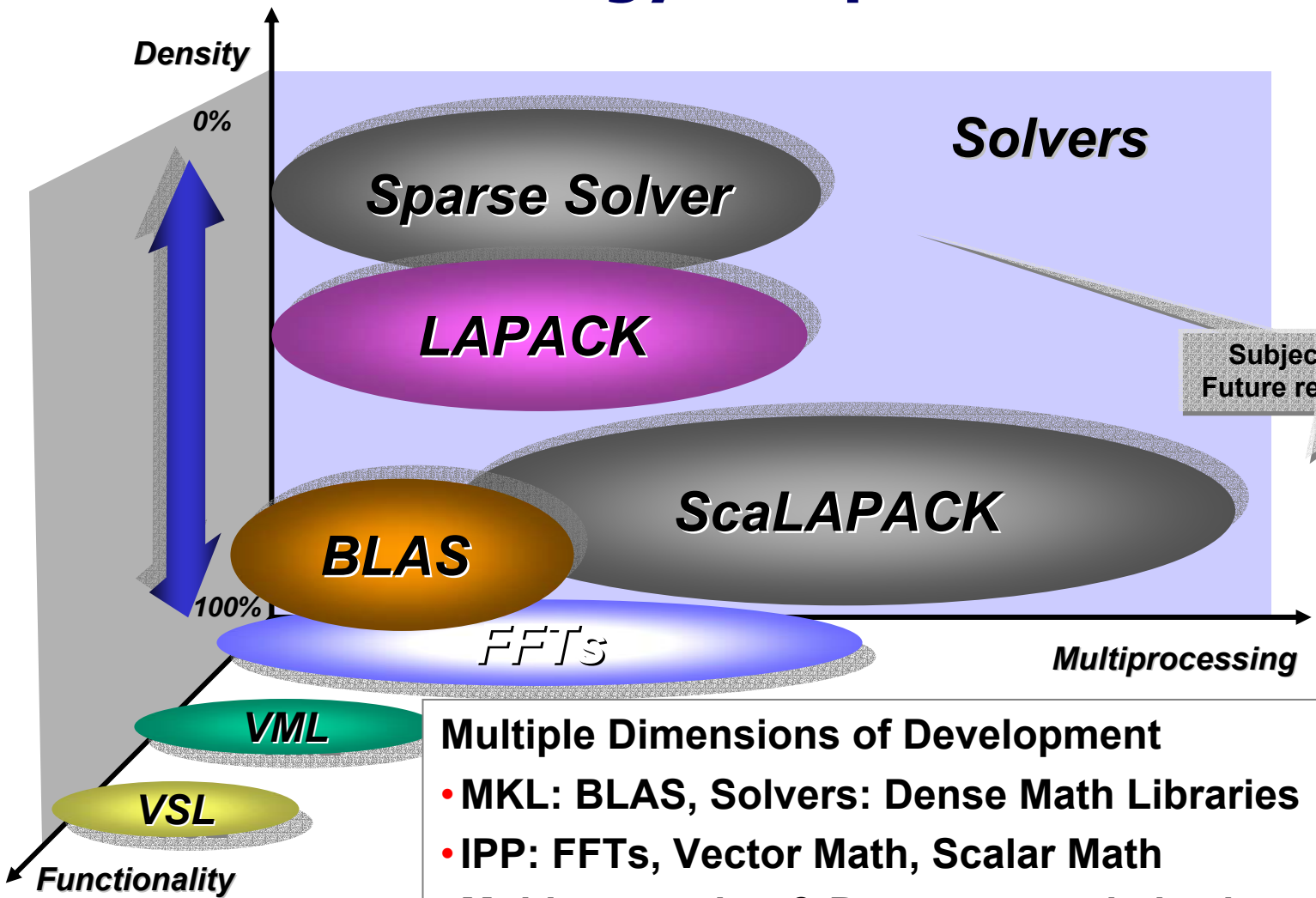
7.0 Compiler Features (cont)

- Support for Threading Tool Kit
  - New Format Assure (Thread Checker) and Guideview (Thread Profiler)
  - OpenMP
    - C/C++ 2.0 (Workqueue), Fortran 2.0 except for Workshare
- GCC Compatibilty Support
  - Build Linux Kernel, Distributions and Applications
  - GCC 3.x C: Source, Library and Object Compatibility
  - GCC 3.x C++: Source Compatibility, Conformance to C++ ABI
- MS Visual Studio Support
  - Intel C/C++ Compilers VS6 and VS.Net IDE integration
  - VC++.Net Language and Object Compatibility
    - No Support for Wizards with ATL headers
  - No Managed Code, Managed Data and CIL
- C99 Compatibility (subset)

7.0 Product Features

- LSB Compliance
- EDB Debugger on Windows
- Ladebug Debugger on Linux
  - Command Line Debugger on Linux32 and Linux64
  - GDB command line compatibility subset - same as LDB
- IA-32: Supported Environments
  - Win2K-32, Win98, WinNT 4.0, Windows XP, Windows.Net Server
  - Linux32: Kernel 2.4 with glibc 2.2.2 and glibc 2.2.4
- IPF: Supported Environments
  - Cross Compilers: Win2K-32, WinNT 4.0, Windows XP, Windows.Net Server
  - Native Compilers: Win2K-64, Windows.Net Server
  - Linux64: Kernel 2.4 with glibc 2.2.3 and glibc 2.2.4

MKL Technology Components

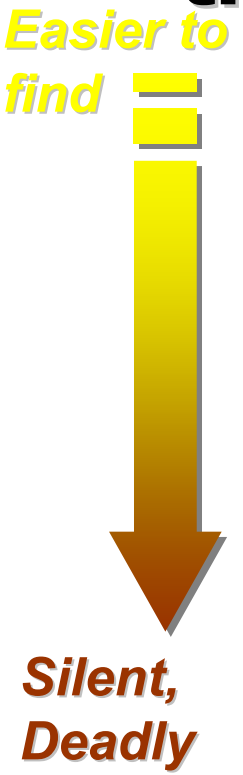


Multiple Dimensions of Development

- **MKL: BLAS, Solvers: Dense Math Libraries**
- **IPP: FFTs, Vector Math, Scalar Math**
- **Multiprocessing & Processor optimizations**

Intel® Thread Checker V1.0 Features

- **Automates detection of most threading errors**



- Deadlocks (detection and prediction)
- Thread stalls (potential deadlocks), waits
- Improperly synchronized I/O
- Threaded calls to non-reentrant routines
- Memory access issues
- Race conditions
- Potential and realized rata races / dependencies
- Invalid threading-library calls, arguments, returns

Thread Checker Usage

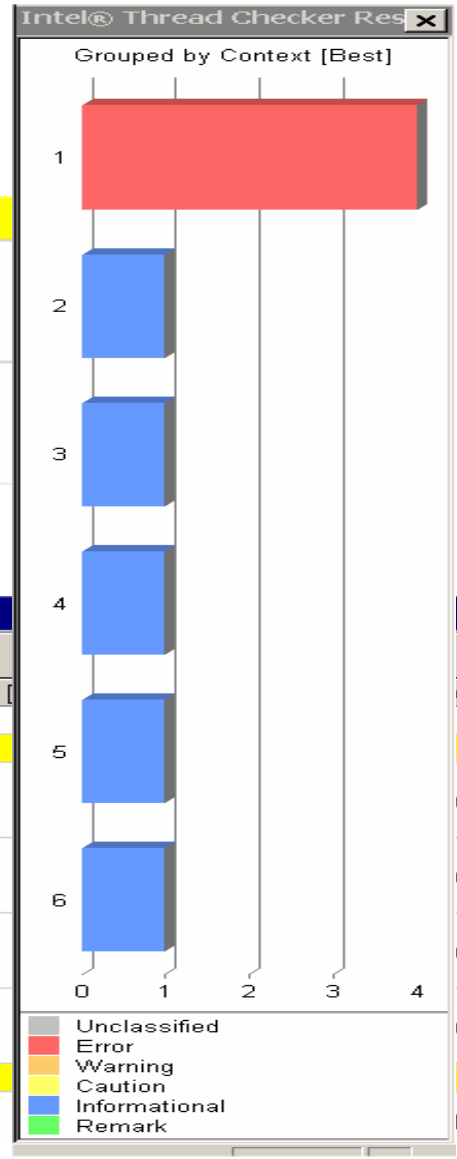
- **Intel® Thread Checker improves productivity**
  - **Isolates application thread errors in hours, that could take weeks to isolate in typical debug sessions**
  - **Use in development mode – isolate problems as soon as they appear**
  - **Use in testing/validation phase – insure new code does not break existing threading code**

Thread Checker Error List

0	●	Memory read of pArg at "Pi.cpp" : 13 conflicts with a prior memory write of i at "Pi.cpp" : 51 (flow dependence)	16	main	i
1	●	Memory write of dSum at "Pi.cpp" : 24 conflicts with a prior memory read dSum at "Pi.cpp" : 24 (anti dependenc...	1240	PiFunc	dSum

Intel® Thread Checker Results - Thu Sep 19 18:04:37 2002: Error List

	Context [Best]	ID	Seve...	Description	Counts	1st Access [
Total						
Group 1 : "Pi.cpp" : 1						
	"Pi.cpp" : 12	0	●	conflicts with a prior memory write of i at "Pi.cpp" : 51 (flow dependence)	16	main
	"Pi.cpp" : 12	1	●	Memory write of dSum at "Pi.cpp" : 24 conflicts with a prior memory read dSum at "Pi.cpp" : 24 (anti dependenc...	1240	PiFunc
	"Pi.cpp" : 12	2	●	conflicts with a prior memory write of dSum at "Pi.cpp" : 24 (flow dependen...	1240	PiFunc
	"Pi.cpp" : 12	3	●	Memory write of dSum at "Pi.cpp" : 24 conflicts with a prior memory write of dSum at "Pi.cpp" : 24 (output depend...	1240	PiFunc
Group 2 : Whole Program 1						
	Whole Program 1	4	●	Thread Info at "Pi.cpp" : 55 - includes stack allocated of 0 and use of 1044	1	main



Location in Source Code

0	●	Memory read of pArg at "Pi.cpp" : 13 conflicts with a prior memory write of i at "Pi.cpp" : 51 (flow dependence)	16	main	i	"Pi.cpp" : 51
1	●	Memory write of dSum at "Pi.cpp" : 24 conflicts with a prior memory read dSum at "Pi.cpp" : 24 (anti dependenc...	1240	PiFunc	dSum	"Pi.cpp" : 24

Intel® Thread Checker Results - Thu Sep 19 18:04:37 2002 (ID=0)2nd Access

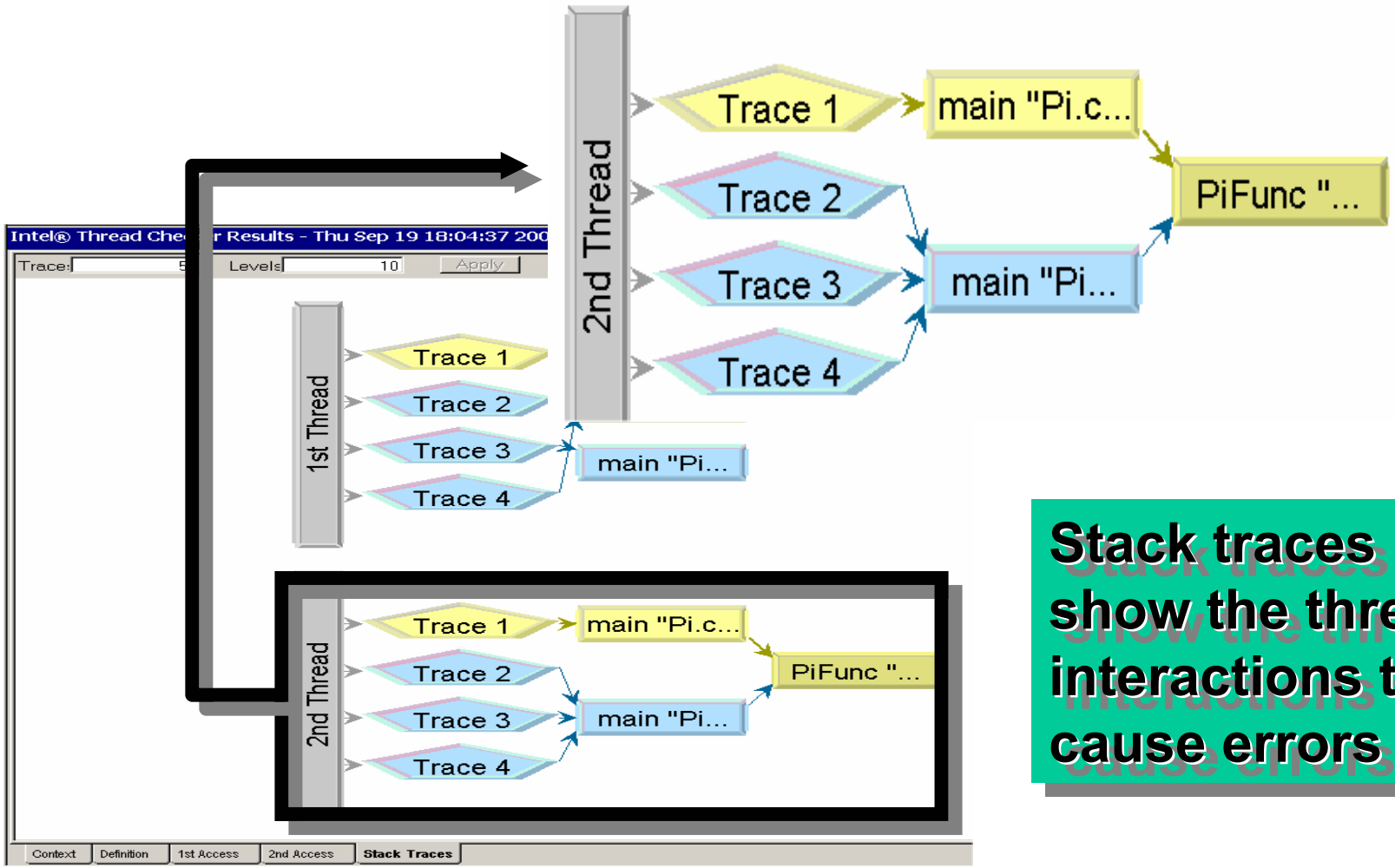
Stack Trace: PiFunc "Pi.cpp" : 13

Line		Source
10		
11		DWORD WINAPI PiFunc(LPVOID pArg)
12		{
13	●	int myThreadNum = (int)(*(int*)pArg);
14		int start;
15		double dx;

Context | Definition | 1st Access | **2nd Access** | Stack Traces

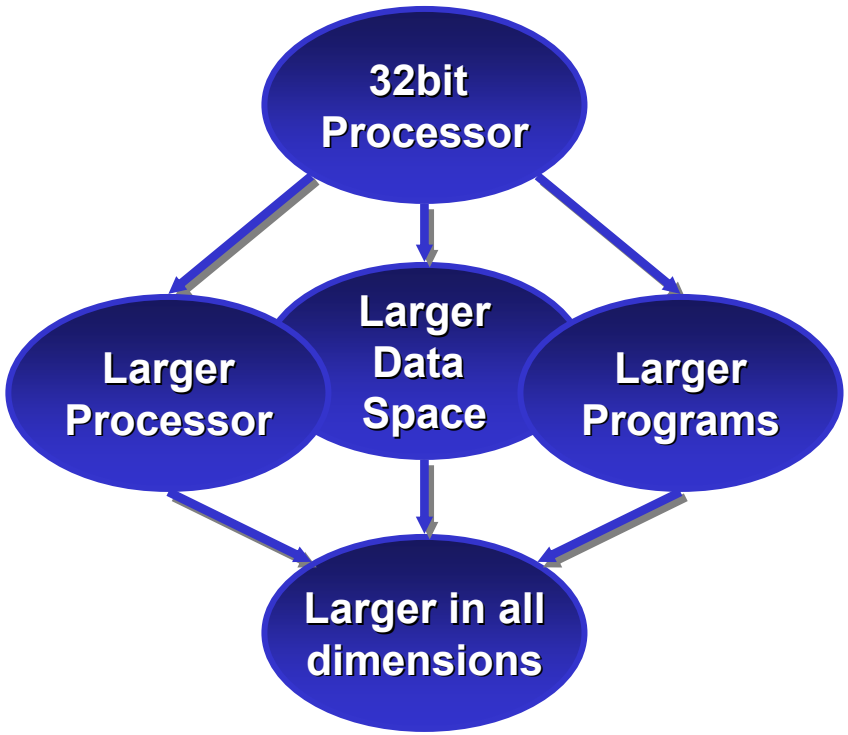
Each entry in the error list links to the location(s) in source code where error occurs

Stack Trace



Stack traces show the thread interactions that cause errors

Outline – Intel Software Tools Improving Performance in 3 Dimensions



- What are the tools.
- The data –

1. Larger processor
   - Itanium performance
2. Larger data space
   - SMPs
   - Interconnects
   - DVSM
3. Larger programs
   - Portals and DRMs
   - Multiphysics, Multichem

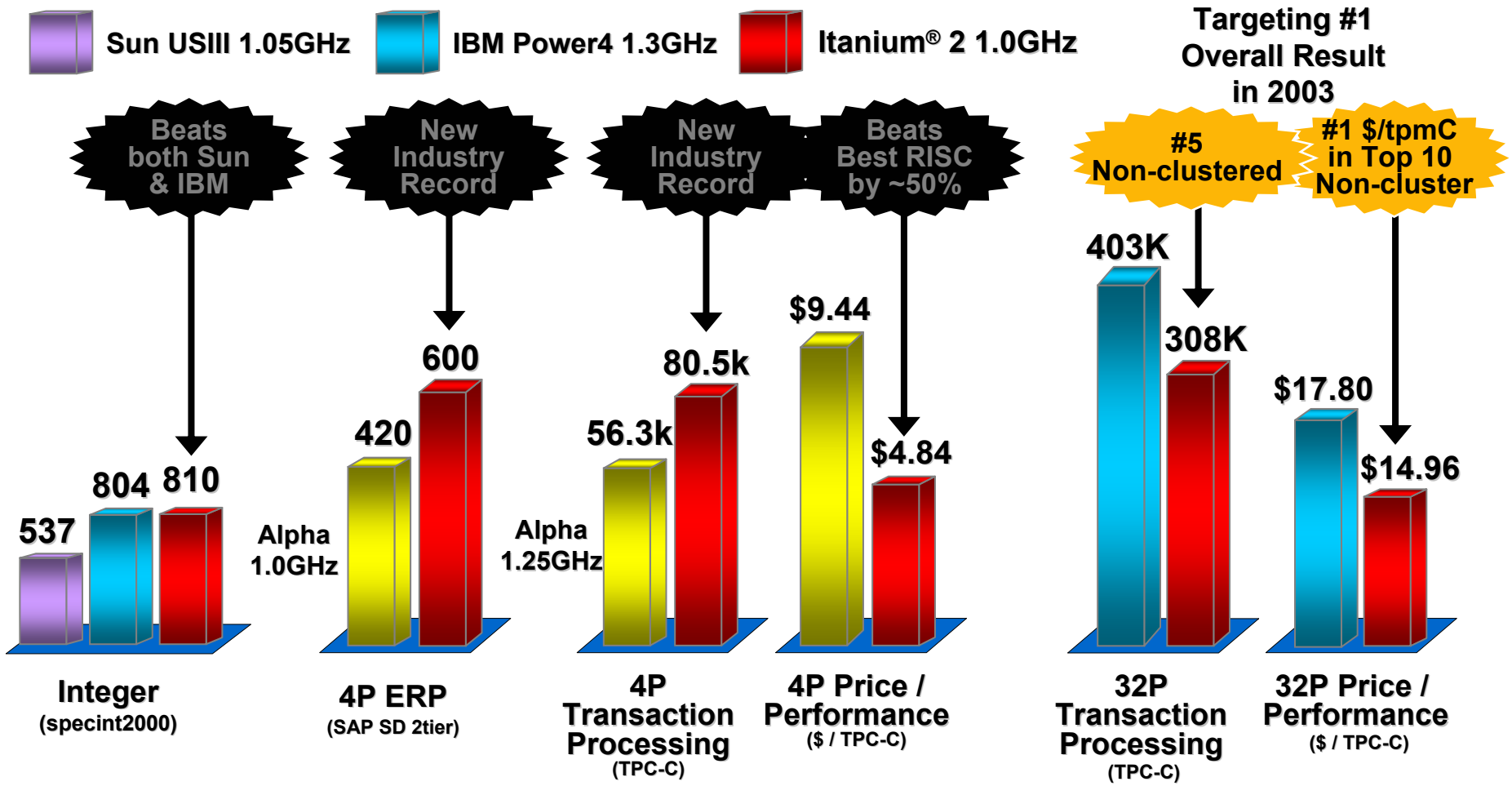
Performance Data

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel® products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference http://www.intel.com/procs/perf/limits.htm or call (U.S.) 1-800-628-8686 or 1-916-356-3104.

Benchmark	Performance Number
SPECint*_base2000	810
SPECfp*_base2000	1356
Stream TRIAD	3700 MB/s
Linpack-1000 (single processor)	3534 MFLOPS
Linpack-10K (4-way system)	13940 MFLOPS
Linpack-HPC (32-way system)	101770 MFLOPS
SPECweb99*_SSL	1520 simultaneous connections
32-way server TPC transactions	308,620 tpmC at $14.96/tpmC
4-way server TPC-C transactions	78,454 tpmC at $5.12/tpmC
2-way server TPC-C transactions	40,621 tpmC at $5.72/tpmC
SAP 2-tier SD 4-way server	470 SD users

Itanium® 2 Processor Performance & Value

Record Setting Performance for the Enterprise



Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference www.intel.com/procs/perf/limits.htm or call (U.S.) 1-800-628-8686 or 1-916-356-3104

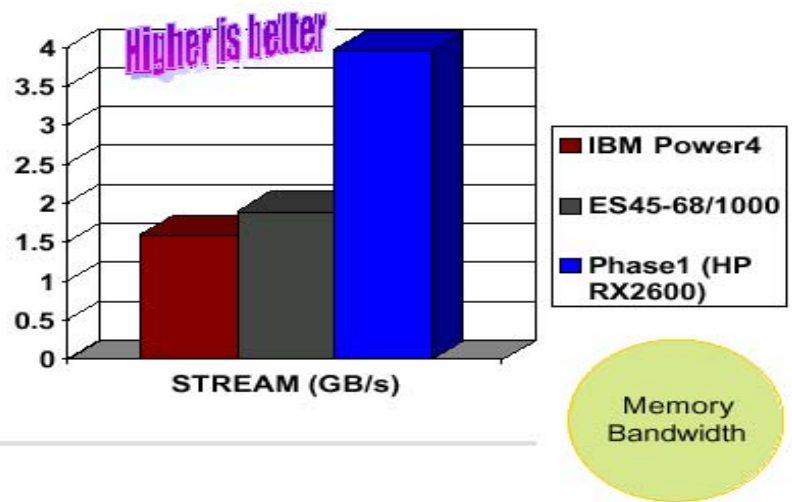
Stream Triad, PNNL P=1

EMSL

Molecular Science Computing Facility

Memory bandwidth

- Sustaining 3.9GB/sec on a STREAM Triad benchmark with a single CPU.
- This is 2.5X faster than RISC processors.



Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

intel

PNNL: Gaussian98 – Itanium 2, P=1

EMSL

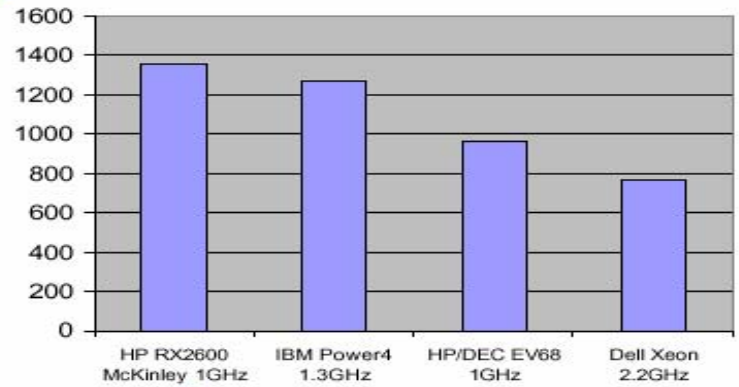
Molecular Science Computing Facility

Other single CPU benchmarks

SpecFP vs. PNNL Phase1

Higher is better

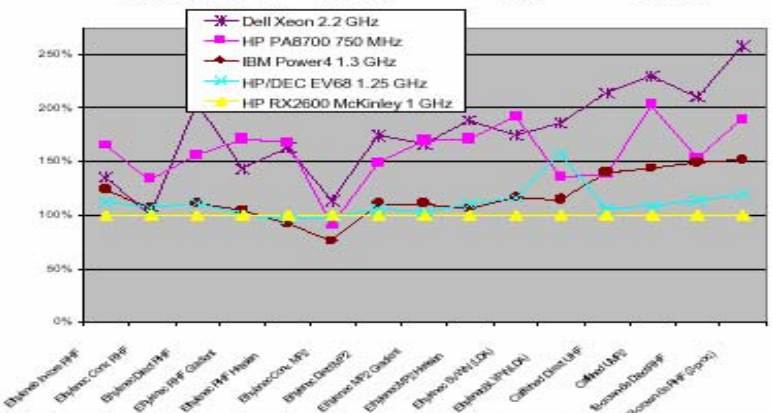
- 100% HP Itanium2 1GHz
- 93% IBM Power4 1.3 GHz
- 71% HP/DEC EV68 1 GHz
- 56% Intel Xeon 2.2GHz
- 43% HP PA8700 750 MHz



Average Gaussian98 time to solution vs. PNNL Phase1

Lower is better

- 100% HP Itanium2 1GHz
- 113% HP/DEC EV68 1.25 GHz
- 144% IBM Power4 1.3 GHz
- 160% HP PA8700 750 MHz
- 220% Intel Xeon 2.2GHz



NWChem is not the only code that is faster on this system than others

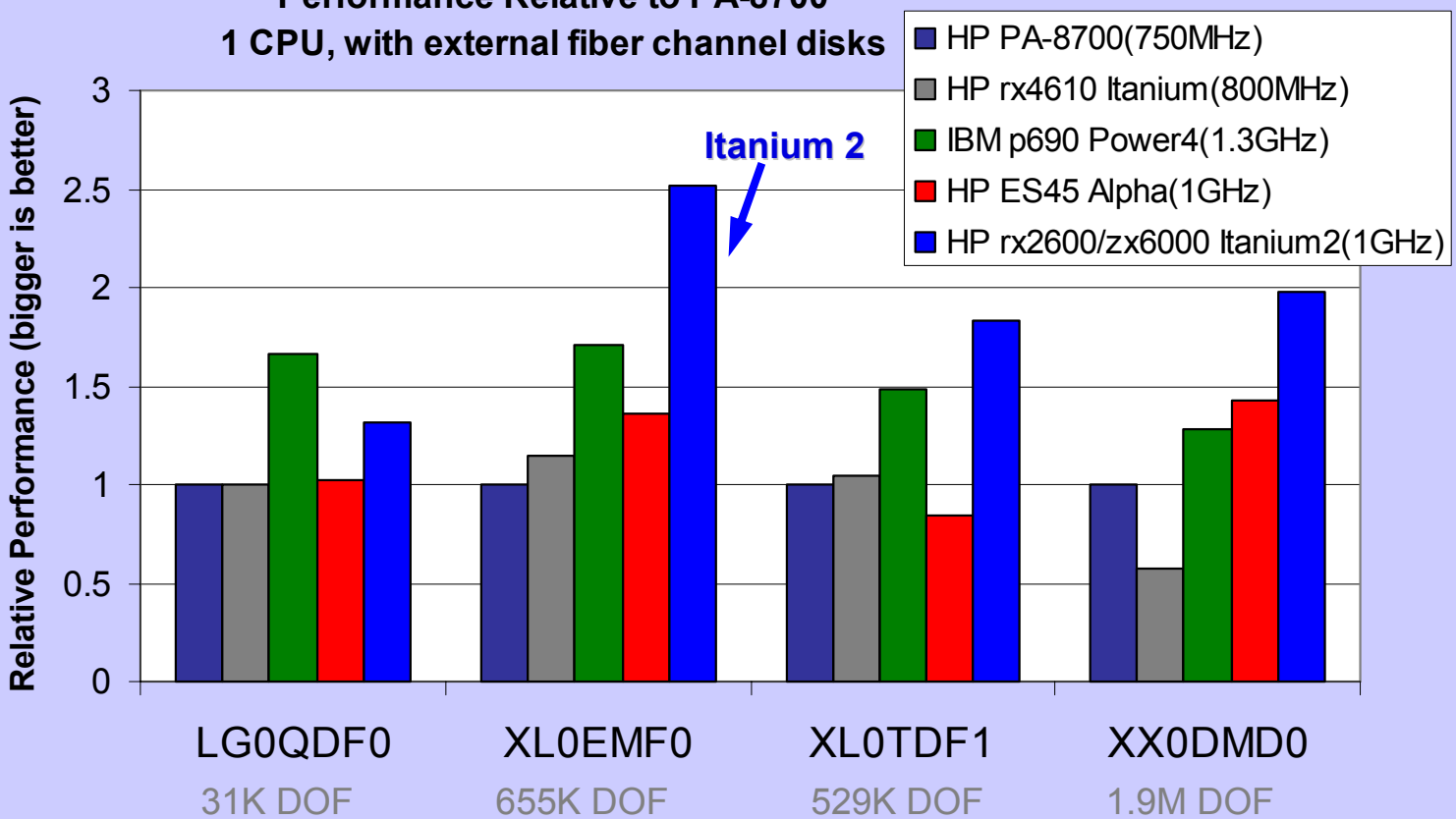
Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

MSC.NASTRAN

MSC SOFTWARE
SIMULATING REALITY

MSC NASTRAN -- STANDARD BENCHMARKS

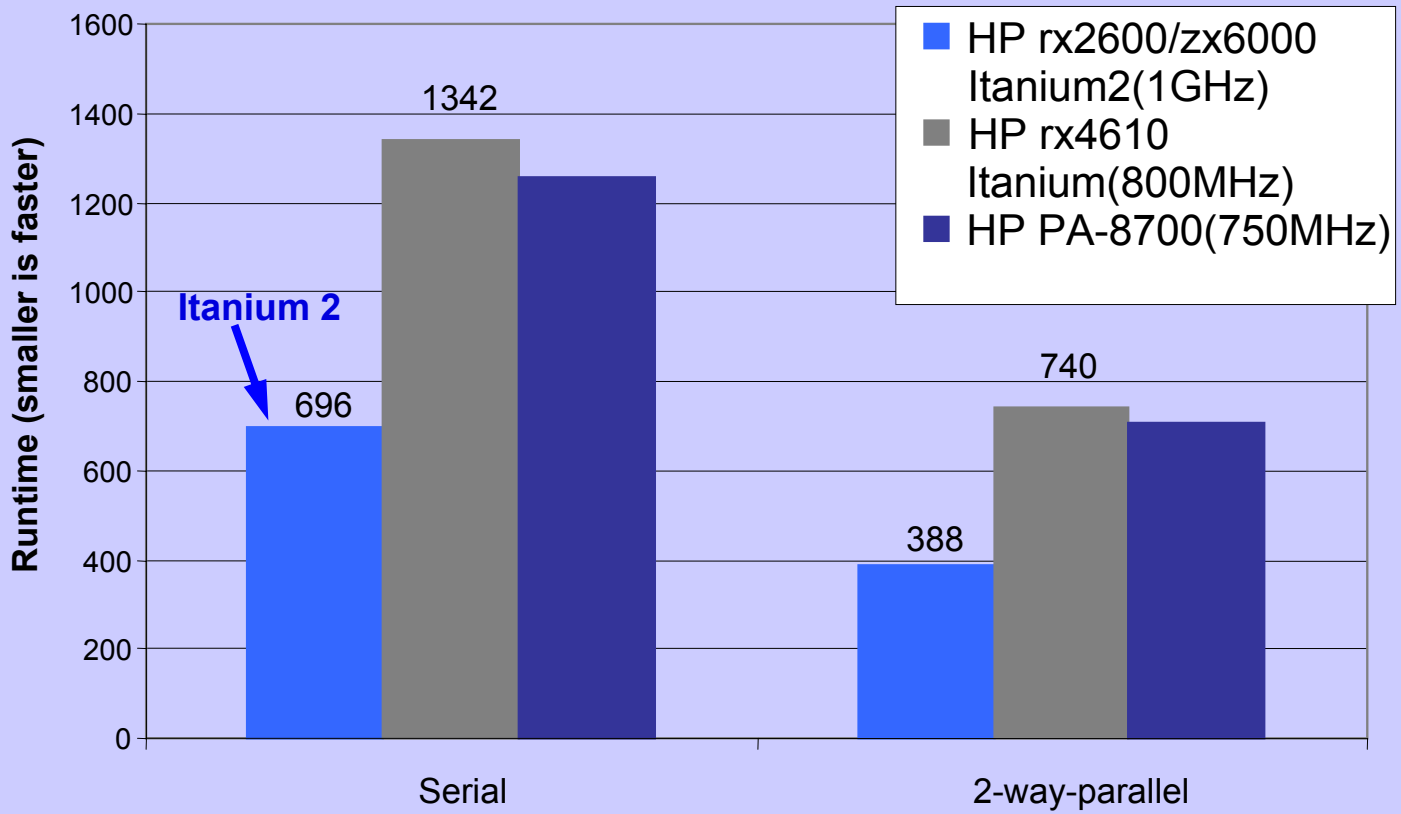
Performance Relative to PA-8700
1 CPU, with external fiber channel disks



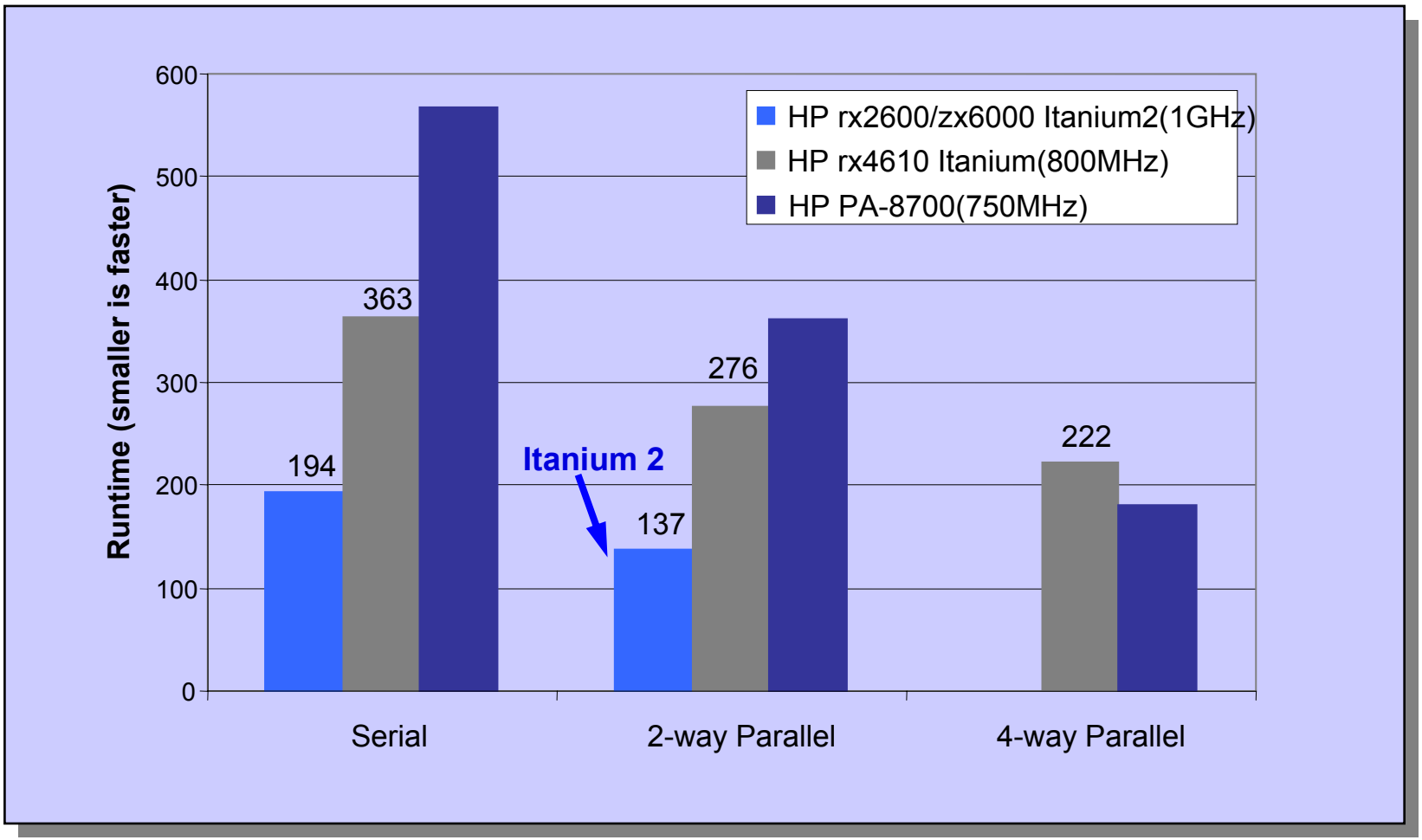
Reference: v2001 data posted at *www.mscsoftware.com/support/prod_support/nastran/performance/index.cfm*
HP-UX/Itanium2 data run with pre-release v2002 and results submitted/approved by MSC.Software. See prior slide for additional explanation of benchmark comparisons. All HP runs with HP-UX 11.x or Tru64.

intel. hp invent

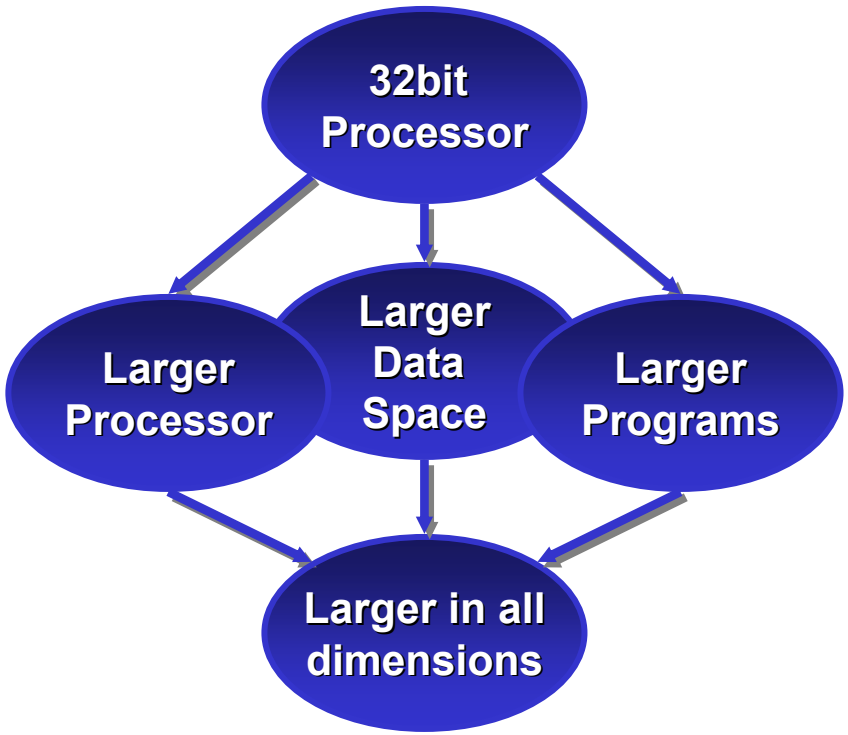
Dynamic Structural Analysis Application



Computation Fluid Dynamics Application

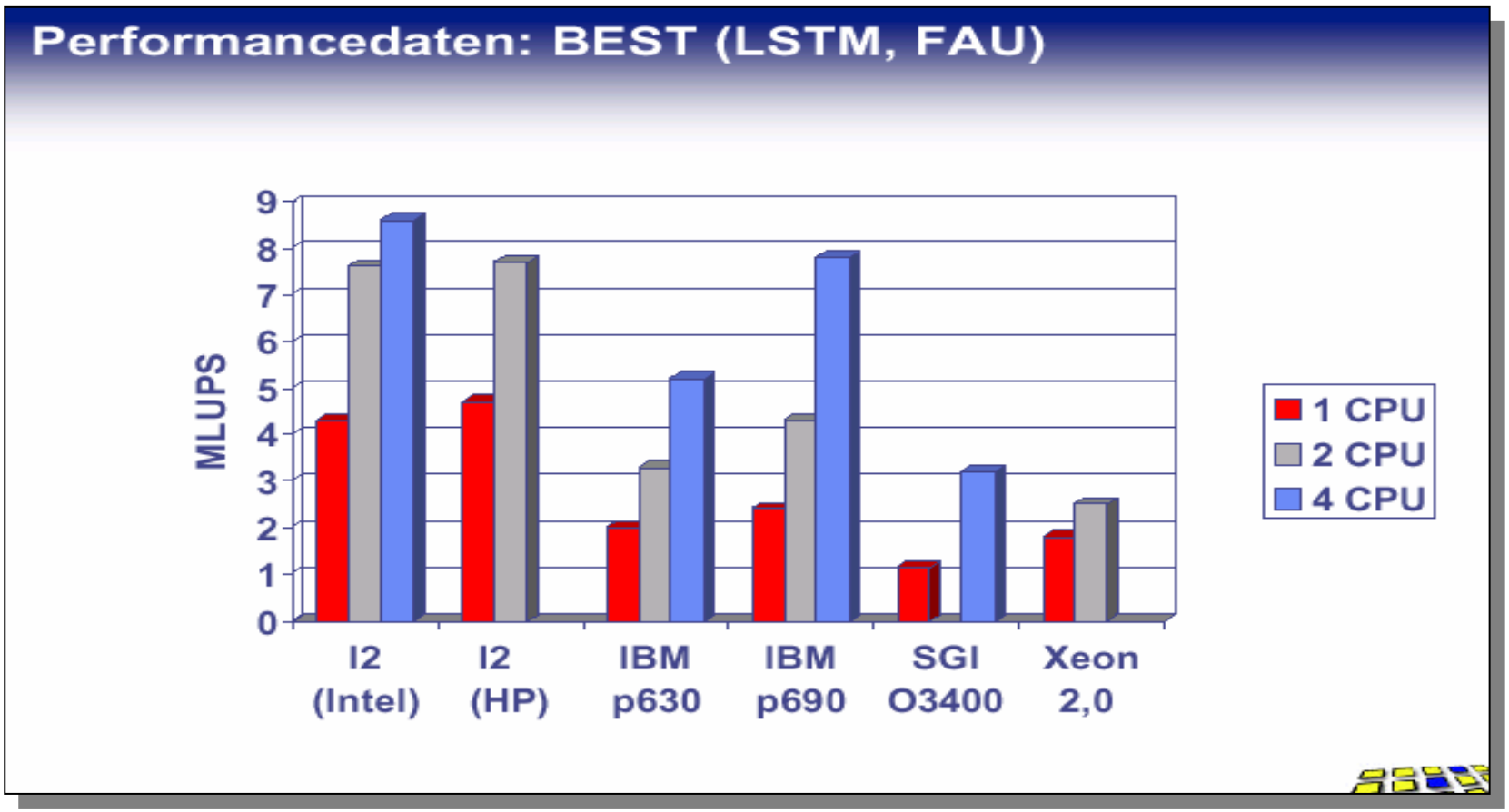


Outline – Improving Systems 3 Ways

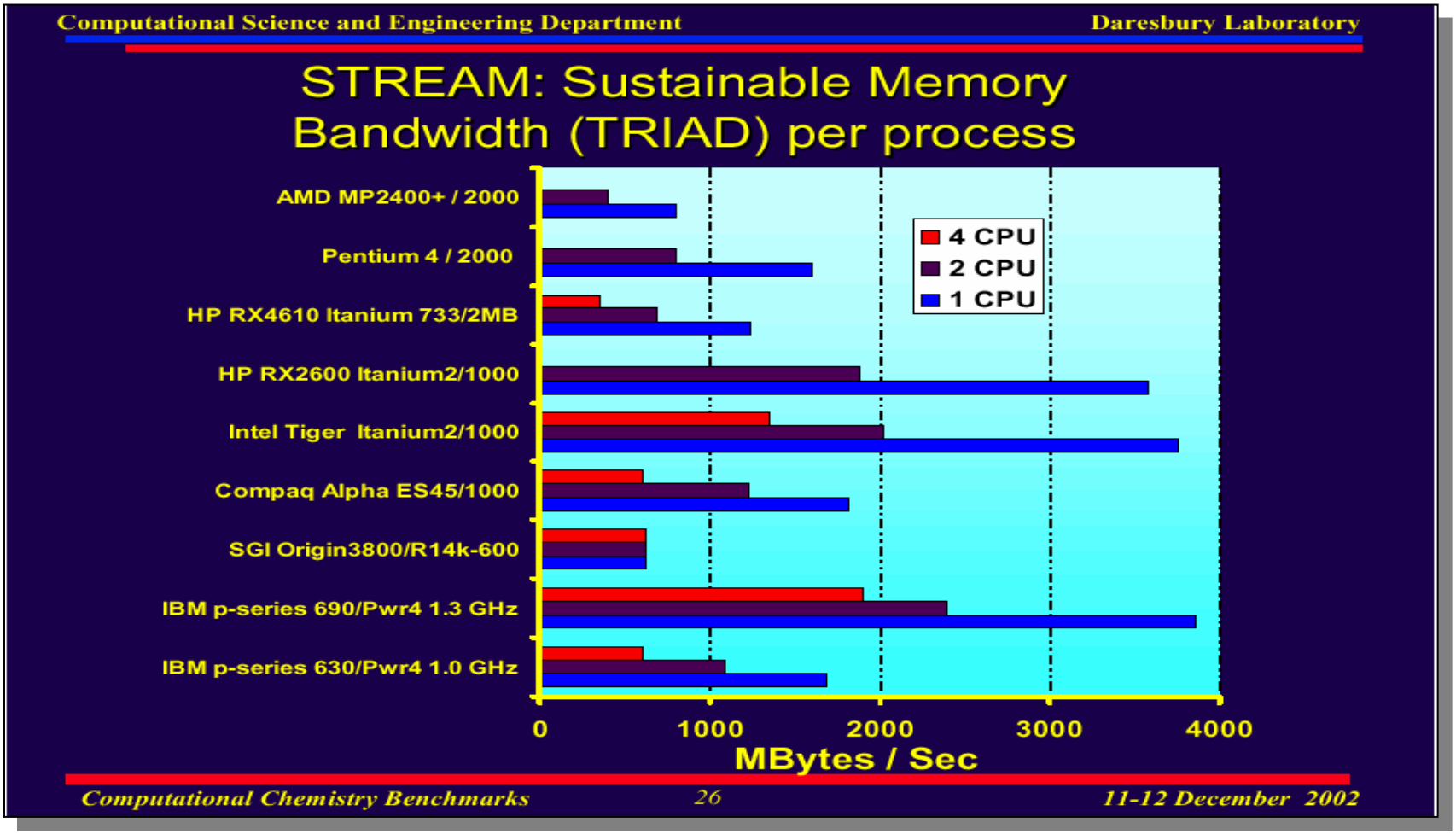


1. Larger processor
   - Itanium performance
2. Larger data space
   - SMPs
   - Interconnects
   - DVSM
3. Larger programs
   - Portals and DRMs
   - Multiphysics, Multichem

Erlangen U.: Boltzmann Equation Itanium 2, P=4



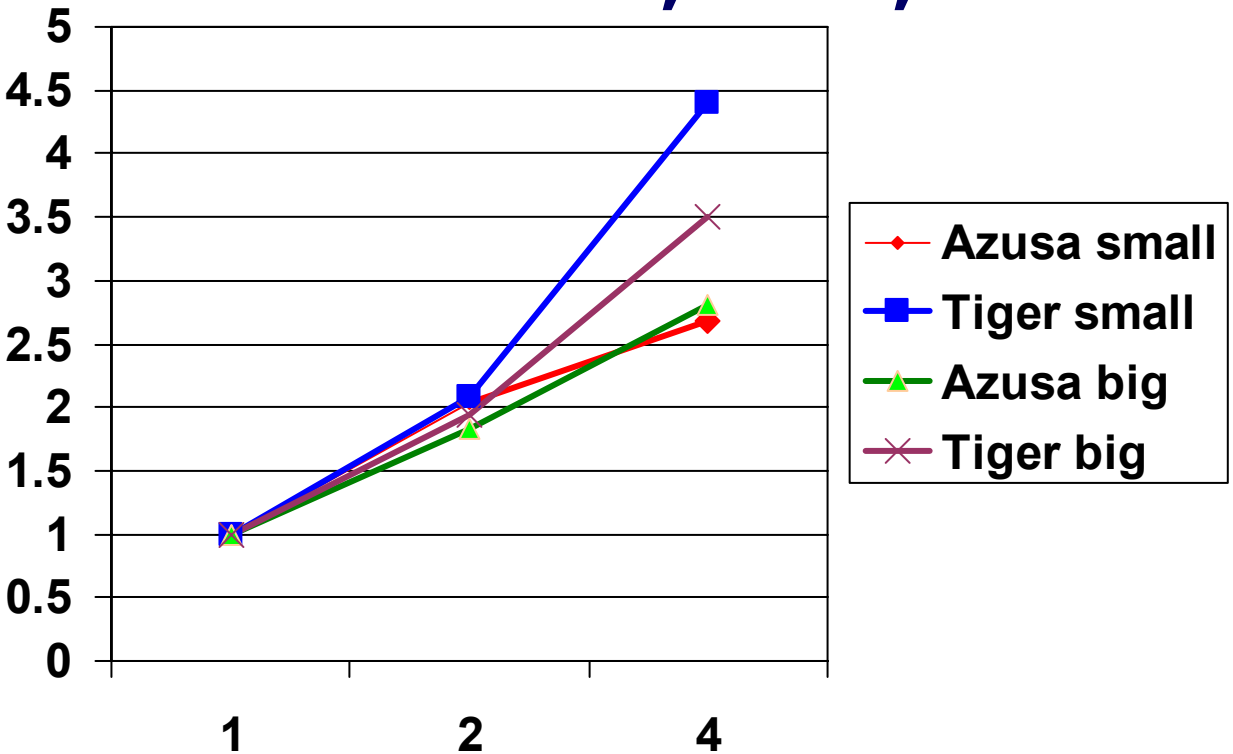
Daresbury Lab – Stream P=4



Compare BW/proc P=1,4: for Tiger (~.35) and Power 4 (~.5).

Martyn Guest
m.f.guest@dl.ac.uk

HLRS Stuttgart: StarCD Speedup Itanium 2, P=2,4



HLRS: Parallel Efficiency of Parapyr – Itanium 2, P=4

Scaled Problem

50x50 Grid Points per CPU

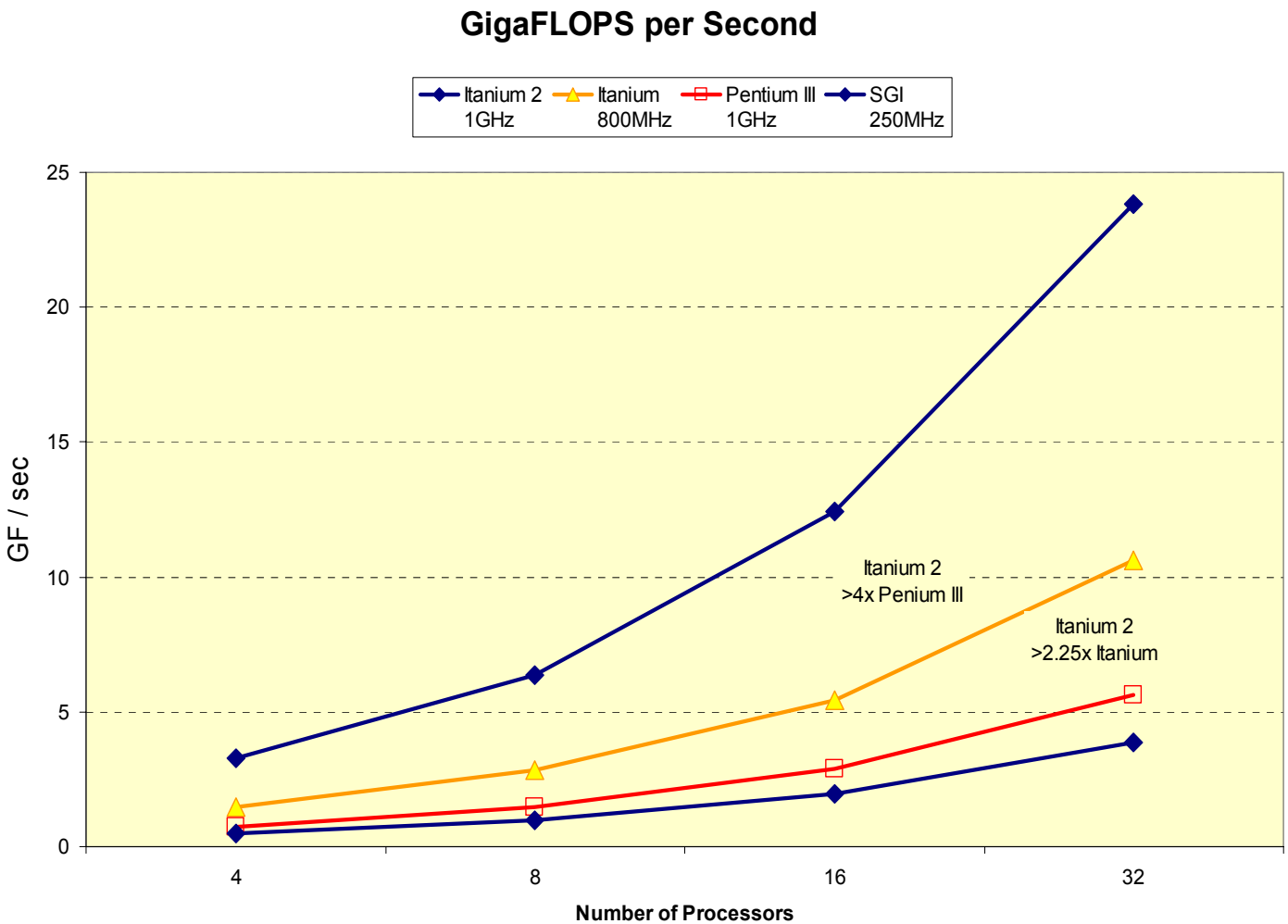
#CPU	4	16
T3E-1200		91.2
Azusa	80.5	78.3
Tiger	82.7	

Scaled Problem

100x100 Grid Points per CPU

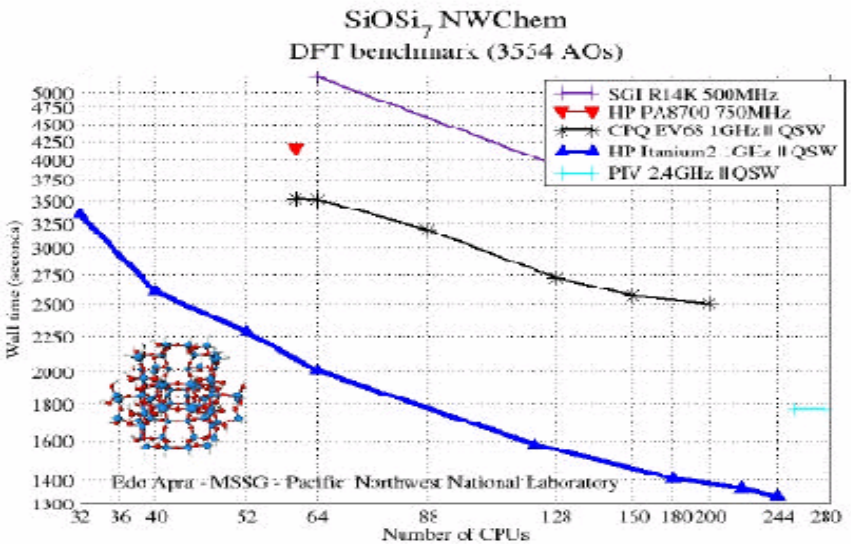
#CPU	4	16
T3E-1200		92.0
Azusa	83.3	78.0
Tiger	81	

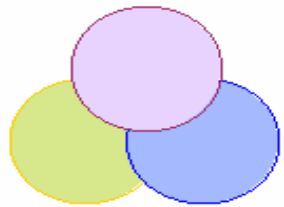
NCSA: NAMD Molec Dyn – Itanium 2, P=32



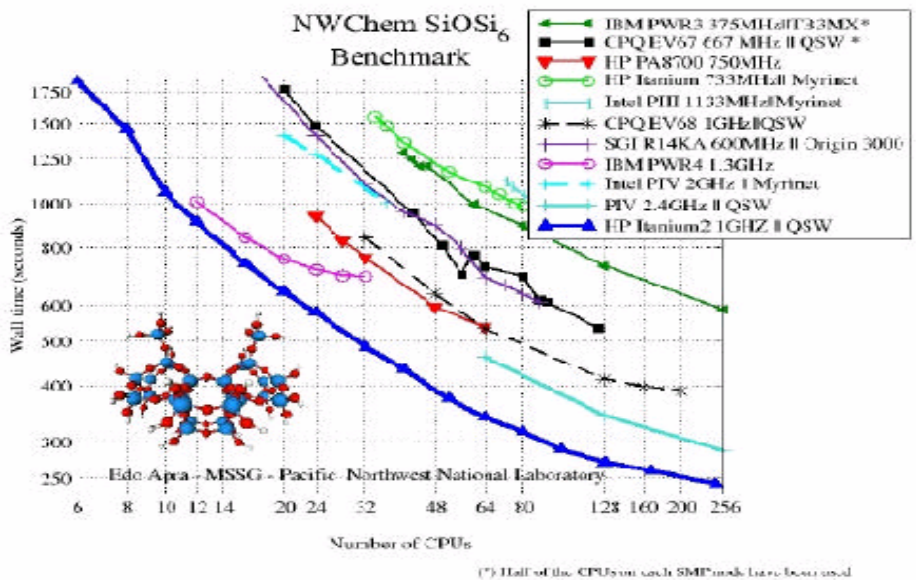
Src: J. Phillips, UIUC

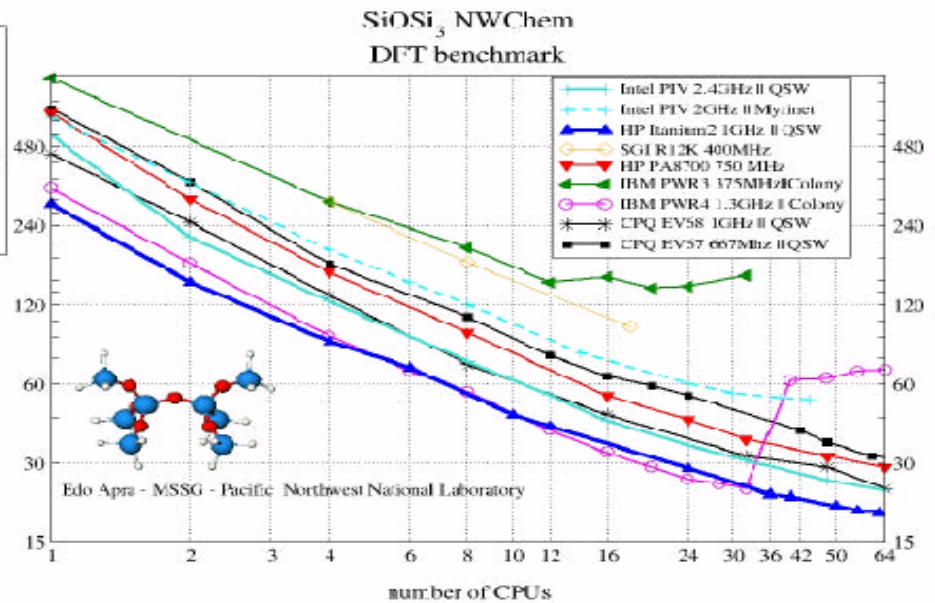
PNNL: NW Chem – Itanium 2 P=256



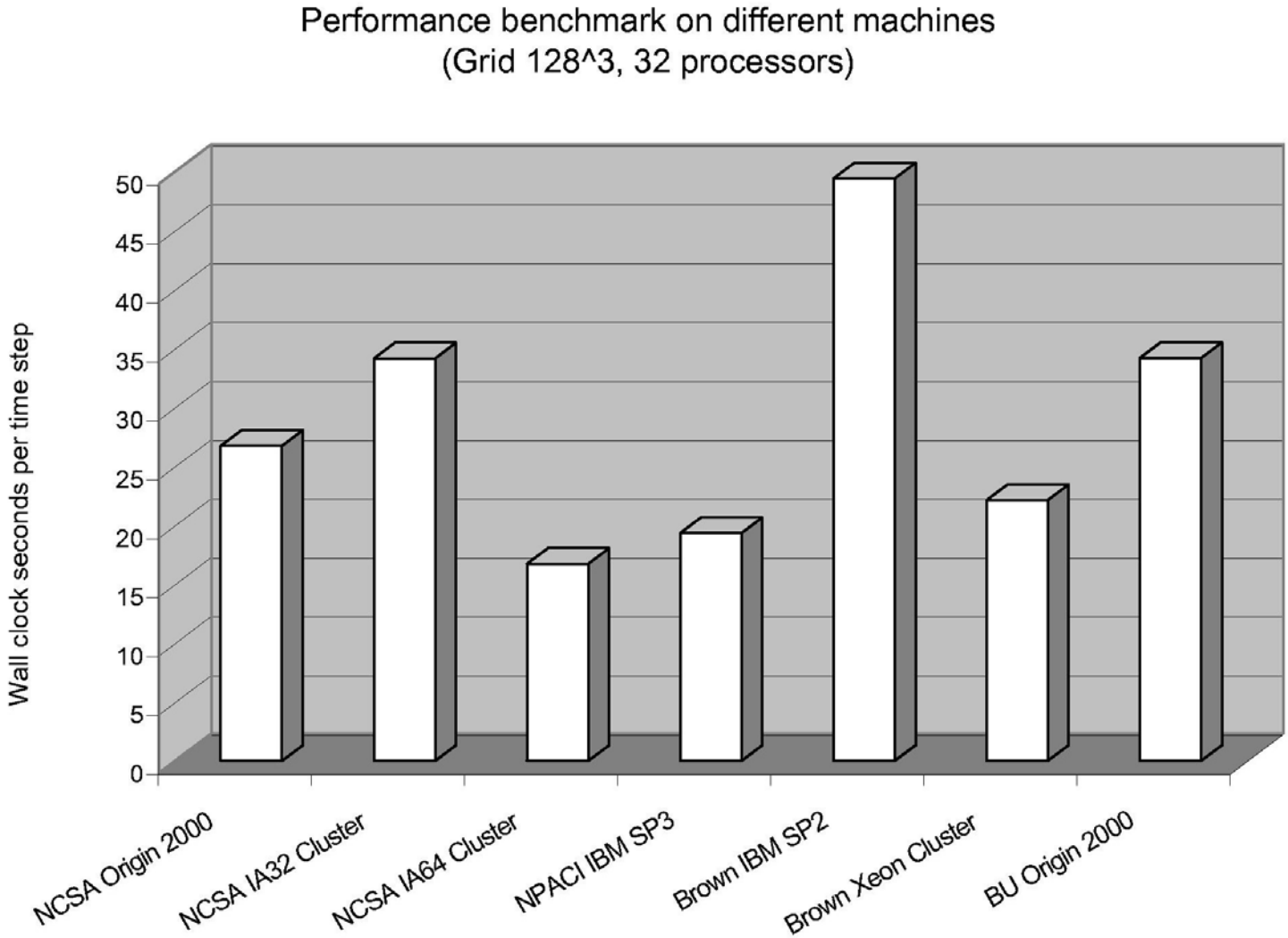


By using the **best in breed** hardware including HP's **ZX1 chipset**, Intel's **Itanium2®** processor and Quadrics **Elan3** we have a truly scalable system.





NCSA: Bubble Simulation – Itanium 1, P=32 Hybrid OMP/MPI code



Src: G. Karniadakis, Brown Univ.

Growth of HPC Clusters on IA

Government

TeraGrid
LANL
LLNL
NOAA
PNNL
Sandia
NCSA
Dutch MOD
China AMS
CERN
China Atmosphere
SNU Grid
Ohio SC
INRIA

Commercial

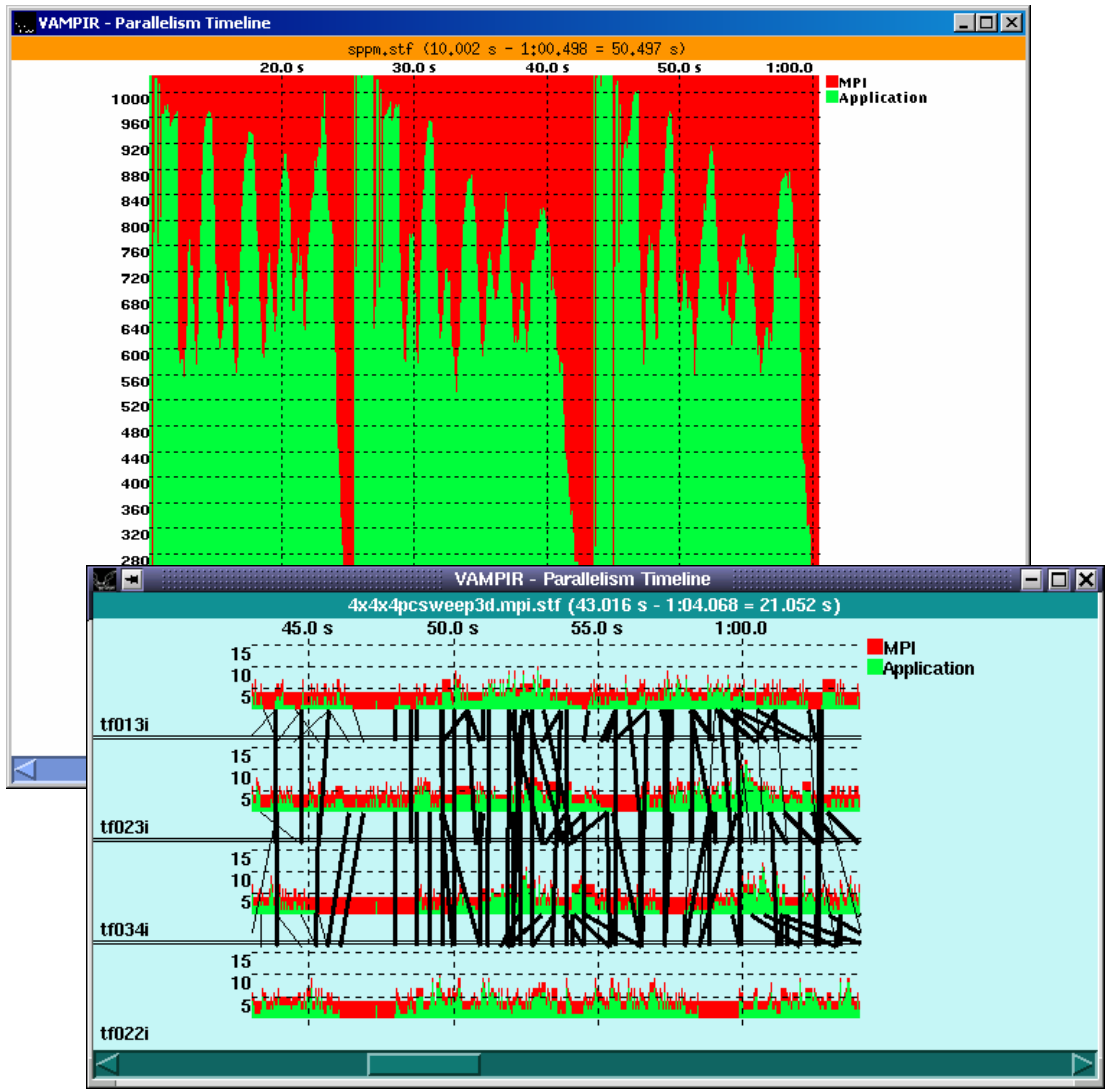
Western GECO
CGC
Shell
Saudi Aramco
PGS
WETA
Shell
Pemex
Google
Inpharmatica
Syrxx
Immunex
MDS Proteomics
Pixar
BP
Chrysler
SAS
Ifineon
Volvo

Academic

SUNY Buffalo
Denmark Scientific
Mississippi State
LSU
Brookhaven
Clemson
Utah
Cornell
Toronto
Virginia Polytech
OSU
Tsinghua
Imperial
NUS
Swinburne
SLAC
South Hampton
Valencia
Oxford
Johns Hopkins
Cal-Tech
Belfast
Zeijing
Princeton
Cambridge
Brandies

VGV – HPC Cluster Performance Analysis

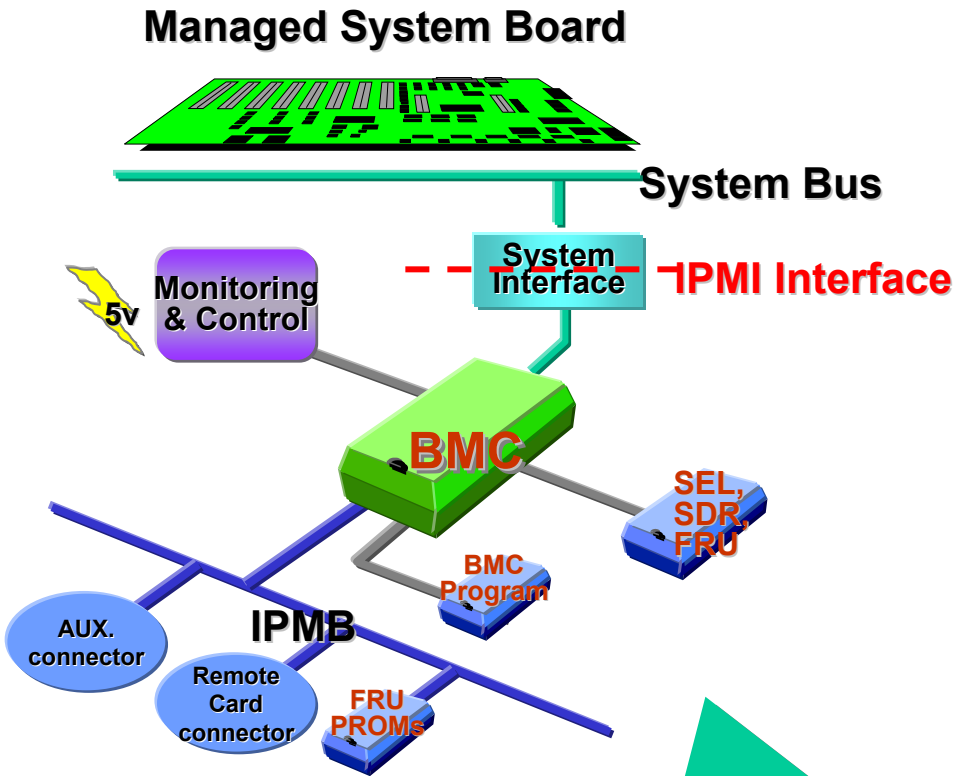
- *Unique in industry* performance analysis for MPI/OpenMP programs
- Many features –
  - MPI nodes messaging
  - OpenMP threading
  - Profiling
  - HPM
  - Object oriented
  - Dynamic control
- Tested for scalability
  - 1000 processor run shown here



"Hands-free" OSCAR

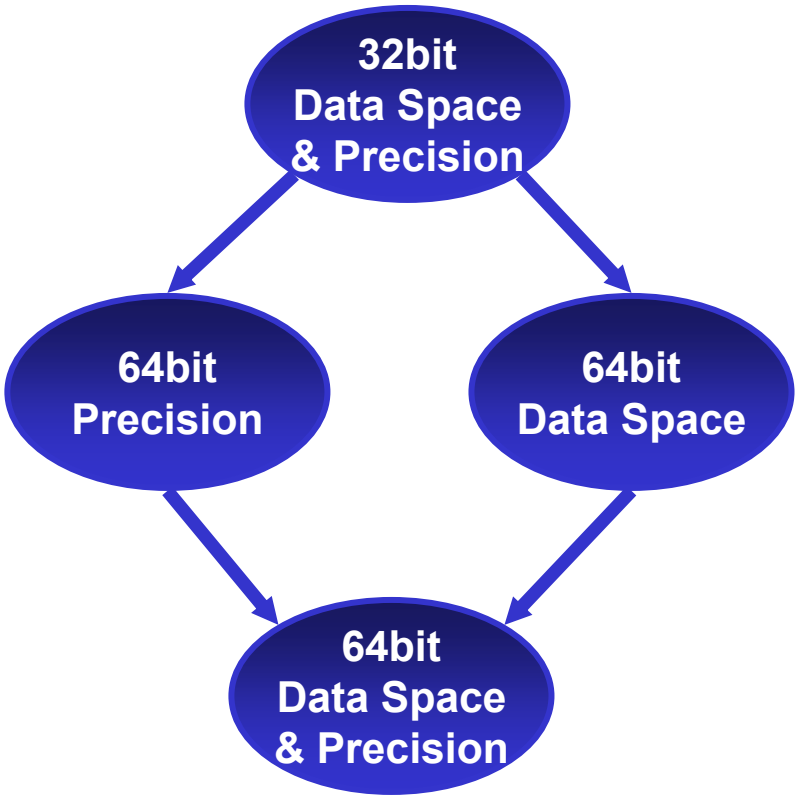
- Integrate ISM with a command line interface to the perl scripts within OSCAR.
  - Alarms to indicate pending hardware problems.
  - Remote control of cluster hardware
    - Power control
    - Flash bios
    - Load OS's and query hardware state

IPMI Hardware Support



- Software stack roundtrip –
  - IPMI – Raise signal
  - Clumon – Monitor
  - Application Checkpoint
  - Oscar – Reboot
  - PBS – Reschedule
  - Application Restart

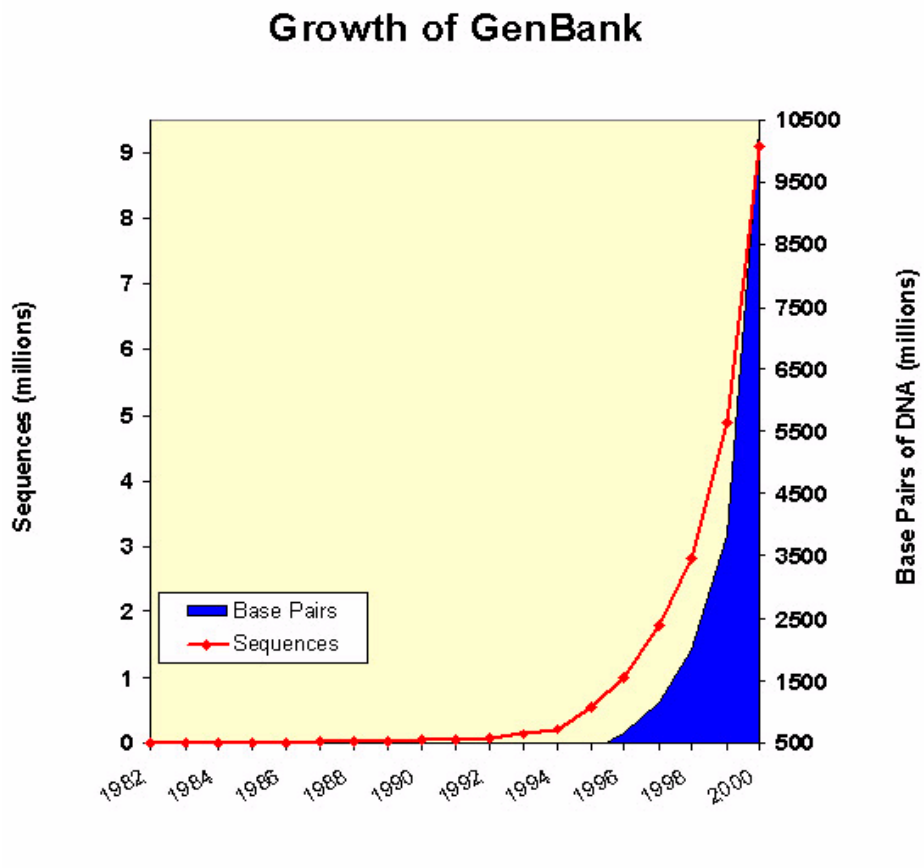
Baseboard provider must provide controller and NIC Also note, IPMI 1.5 was first that provides autonomous signaling



- 64bit precision
  - Very stiff problems
    - Quantum chemistry
    - Structural stiffness
- 64bit data space
  - Growing size of models
    - Structural dynamics
    - CFD
    - Life Science
    - Data Mining

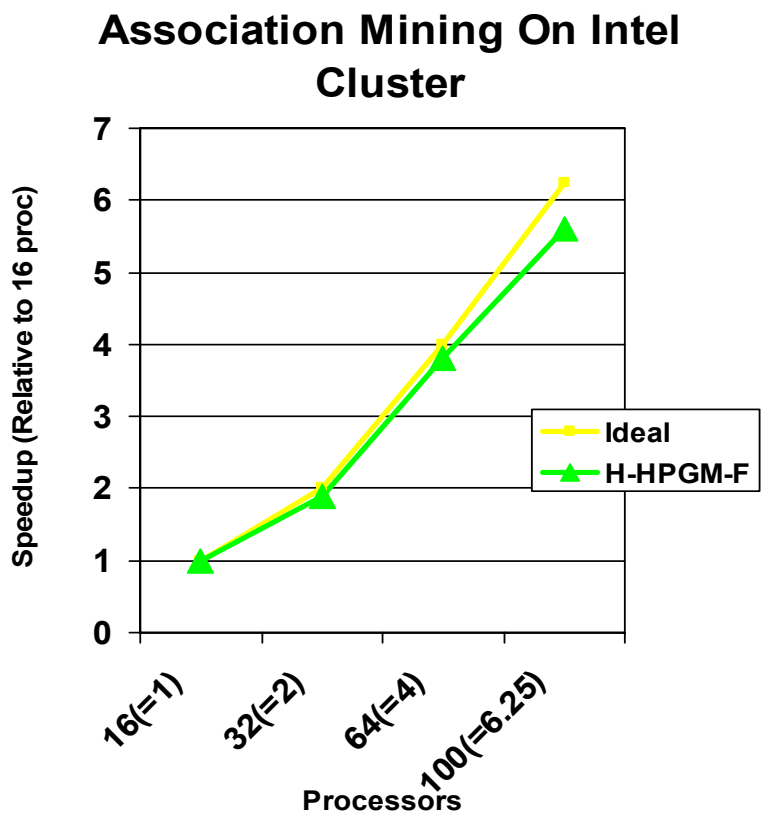
Bioinformatic's Tremendous Data Growth

- Tremendous growth of data in life sciences
  - And applications needed to process it
  - Need fast flexible programming models
  - Eg, Threading tools, OpenMP
  - Applications can rapidly grow out of 32bit data space, DVSM



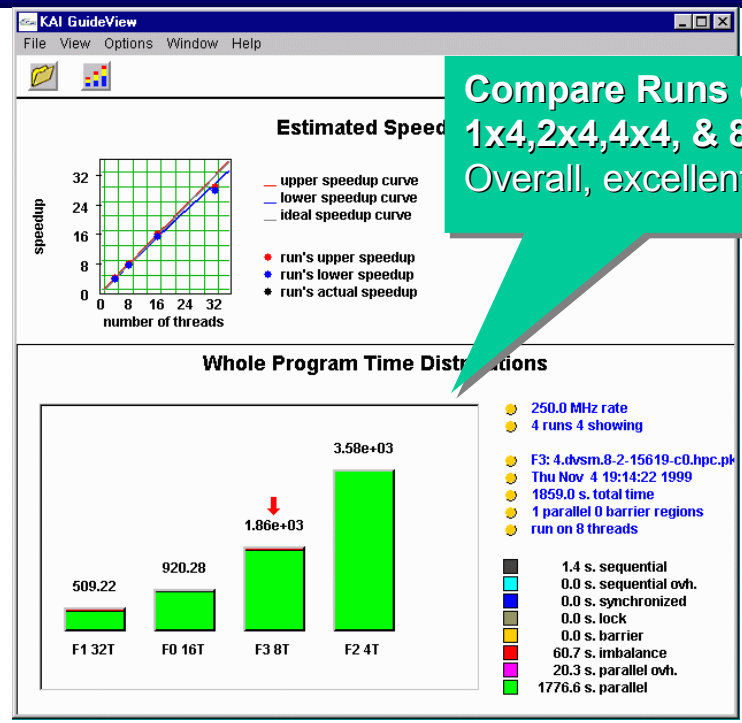
Large Scale Data Mining

- Distributed Association Mining
  - University of Tokyo, Shintani & Kitsuregawa
  - Pentium Pro cluster



DVSM OpenMP Eases Cluster Programming

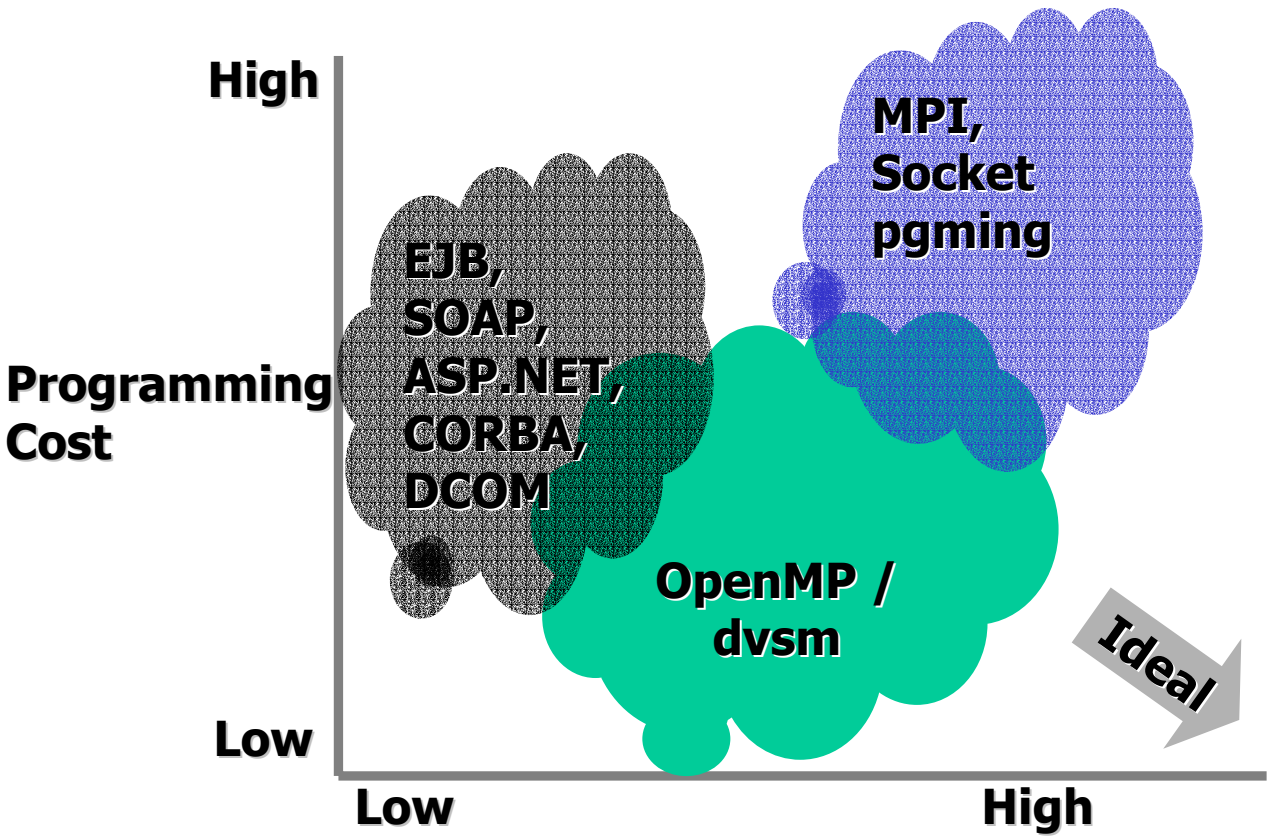
```
#pragma omp parallel
{
 #pragma for
   schedule(static,1)
 for( i=0; i<n; i++) {
  for(j=i+1; j<n; j++) {
   if (/*Compute d2*/)
     /*PrivateSimilar(
              SEQi, SEQj) */
 } }
 #pragma omp critical
 { /* Reduce PrivateSimilar to
   GlobalSimilar */ }
}
```



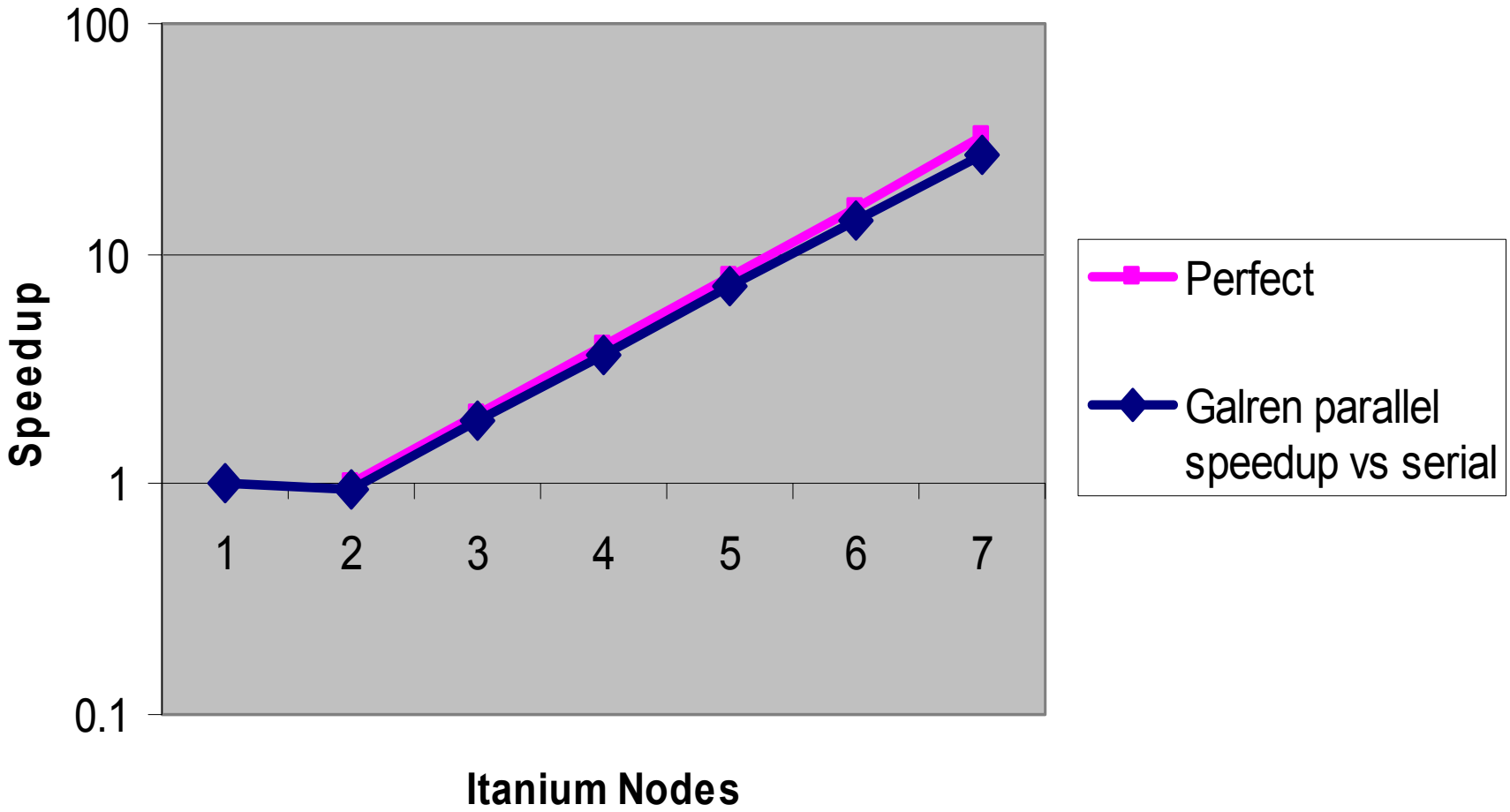
Compare Runs on 1x4,2x4,4x4, & 8x4
Overall, excellent scaling!

- Distributed Virtual Shared Memory --
  - Identifies all shared data with one new pragma
  - Provides perf data on where shared info is transmitted

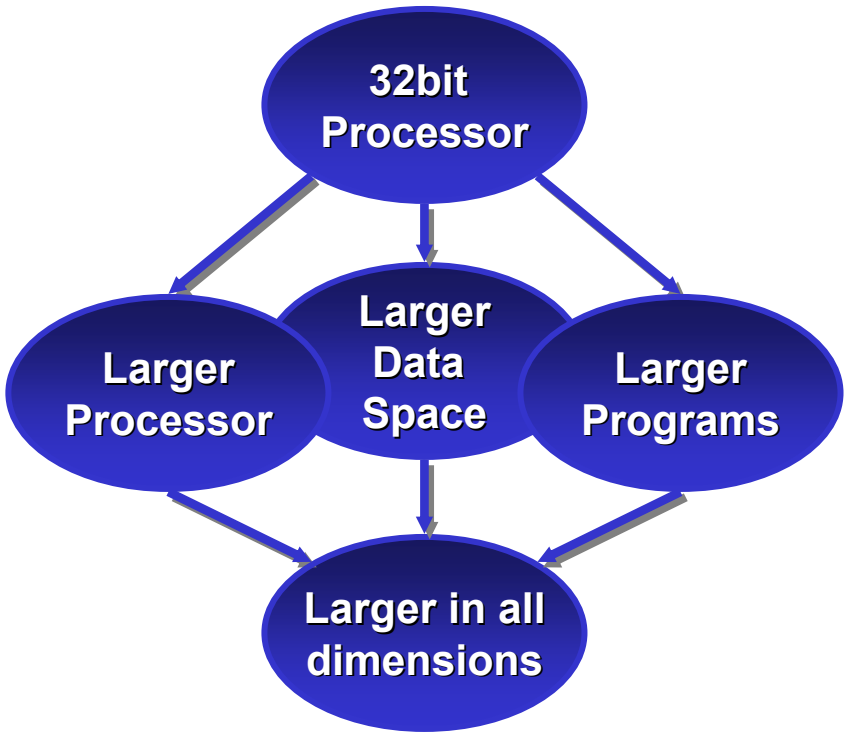
Performance / programming cost



Galren speedup on an Itanium Cluster



Outline – Improving Systems 3 Ways



1. Larger processor
   - Itanium performance
2. Larger data space
   - SMPs
   - Interconnects
   - DVSM
3. Larger programs
   - Portals and DRMs
   - Multiphysics, Multichem

Open Grid Services Architecture

End-Users and Admin Staff

S&TC e-business xSP etc

New classes of applications and Grid's

Common GUI

Install
Config
Deployment
Maintenance

OGSA Autonomic Management Services

eWorkload Management
eCluster Management
eUtility
eAutomation
eDisaster Recovery

OGSA Grid Middleware Services

Globus
Platform
Avaki
Entropia...

OGSA Meta-OS Services

Clustering
Policy
Data Access / Replication
Resource Instrumentation
Problem Determination
Logs

OGSA Domain Servic

Provisioning
Service Collections

OGSA Infrastructure

Discovery
Lifecycle
Registry
Manageability (serviceData+)
Factory
HandleMap
Notification

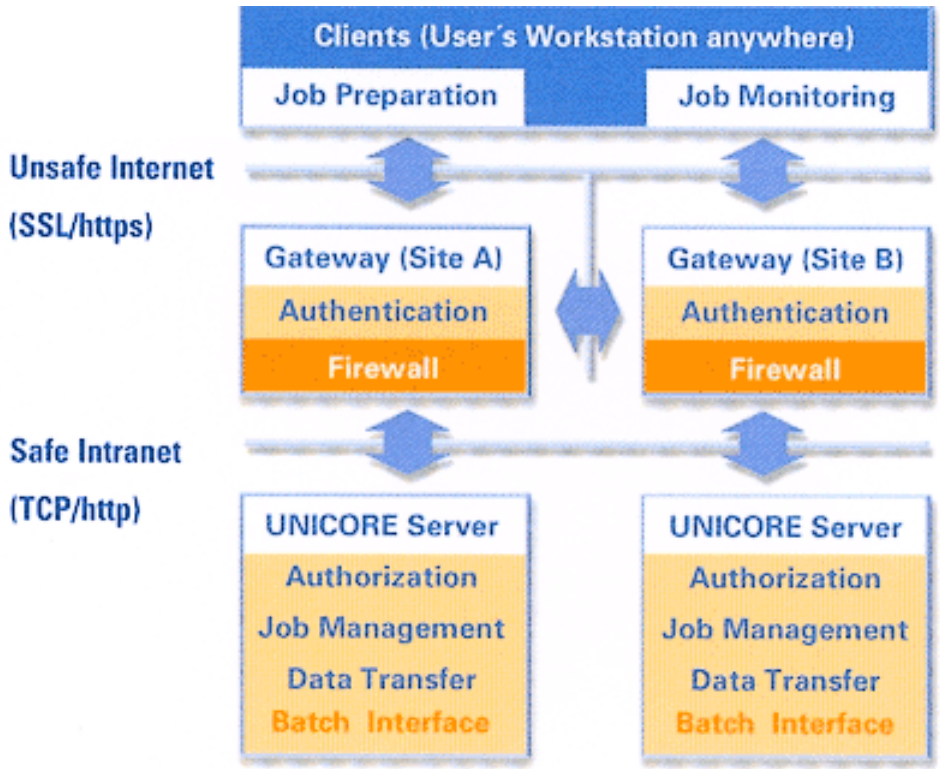
.Net JBOS WebSphere. software WS-Security

z/OS	AIX	OS/400	Linux	Windows	Storage	Solaris & HP/UX	Linux Windows	Network	Storage
zSeries	pSeries	iSeries	z,p,i,xSeries	xSeries	Shark	Sun & HP	Dell Compaq ...	Cisco Nortel ...	Veritas ...

Grid Portals Enabling Cluster Access From Workstations

Unicore – Example of Grid Portal – Joint project of:

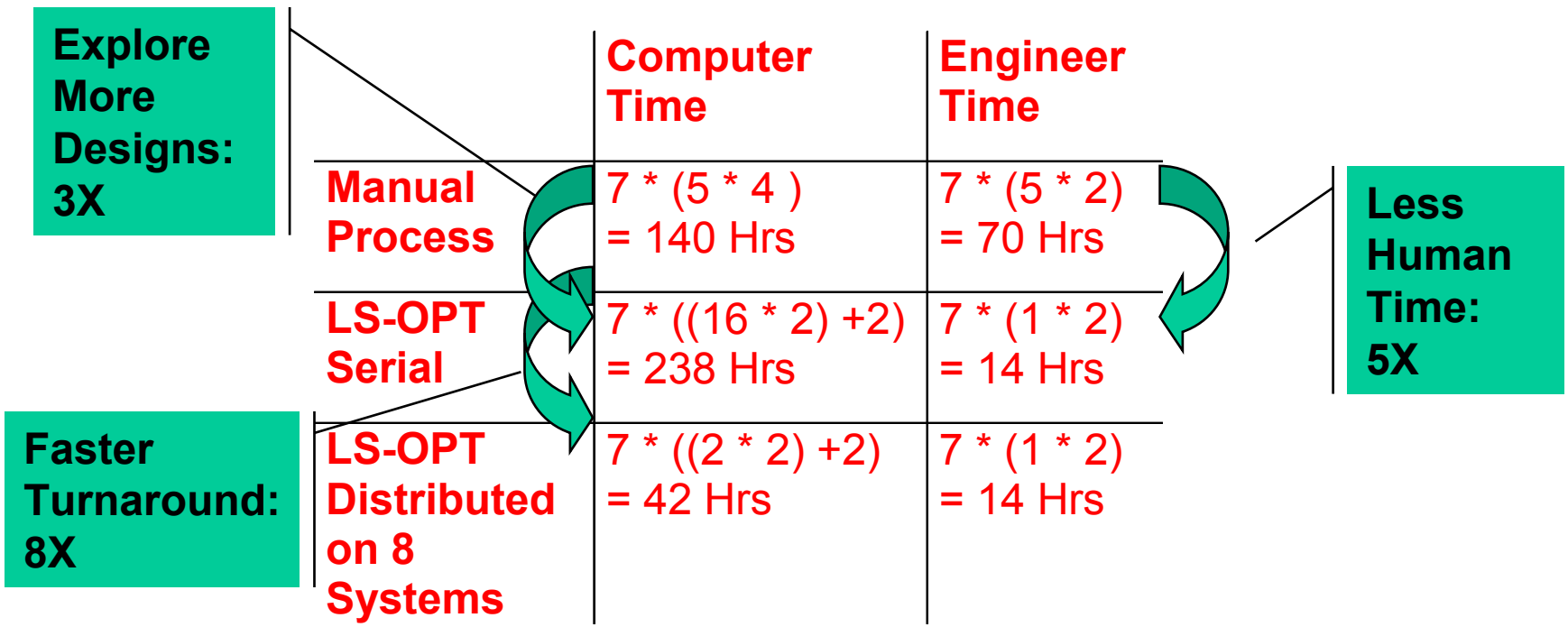
- Forschungszentrum Jülich
- Pallas GmbH
- Deutscher Wetterdienst
- Rechenzentrum der Universität Stuttgart
- Leibniz-Rechenzentrum München
- Universität-GH Paderborn
- University of Karlsruhe (TH), Computing Center
- Konrad-Zuse-Zentrum Berlin
- Technical University of Dresden



www.unicore.org

DRMAA – GGF Working Group on Distributed Resource Management

Dashboard Design Optimization (Ford, Visteon, LSTC)



Time Formula:
NumIterations * ((RunsPerIteration * TimePerRun) + QueuingDelay)

Key Messages

- Overall, Itanium systems have made major gains in performance
- Intel systems use is broadening
  - Clustered systems installations
- Most people know Intel for HW components
  - Processors, chipsets, interconnect
  - Parallelism available at all levels
- Intel produces software tools too
  - Compilers, performance, correctness tools