UKHEC

Application Performance on Beowulf Cluster Systems

Martyn F. Guest
CCLRC Daresbury Laboratory

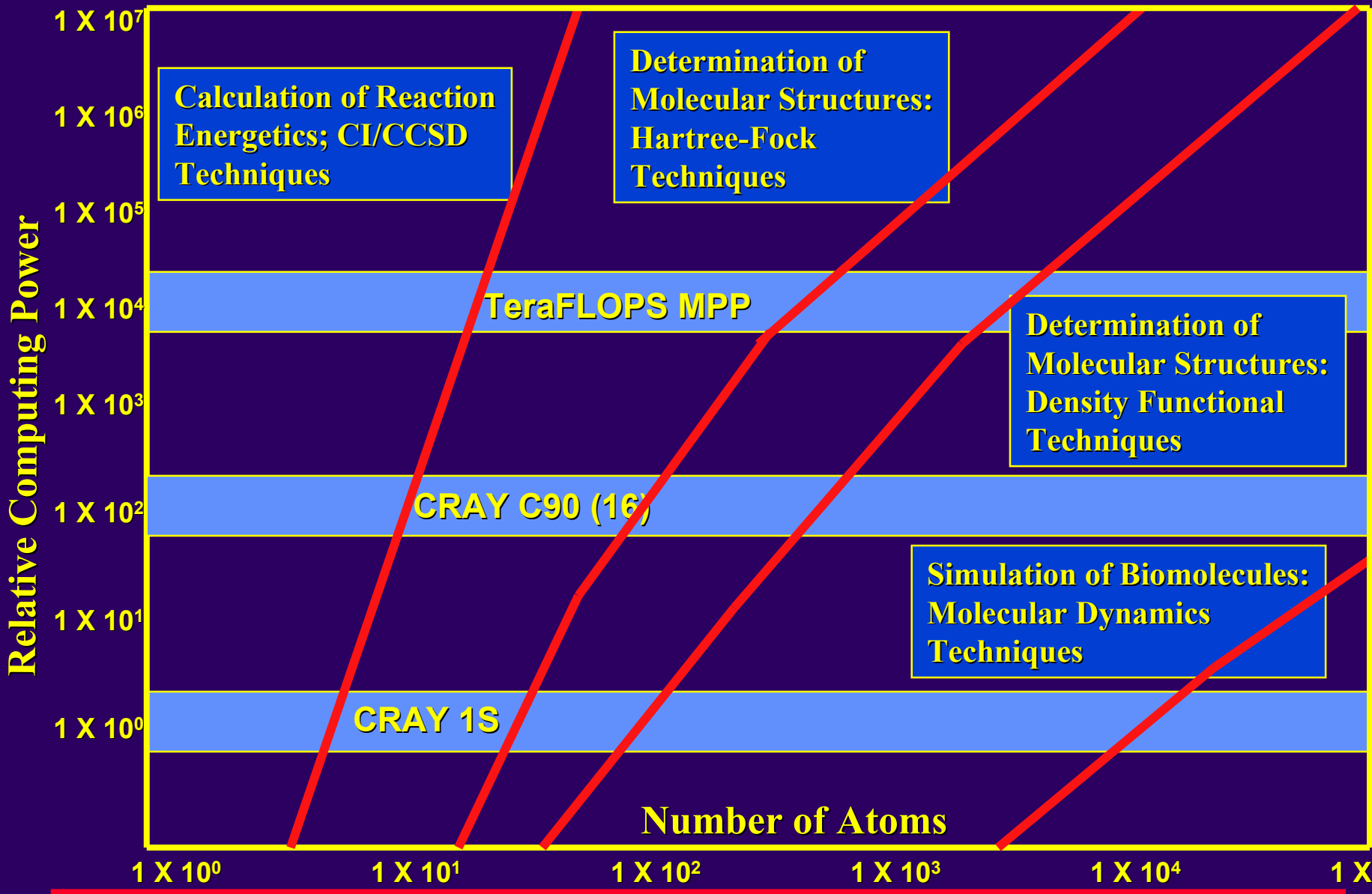
m.f.guest@daresbury.ac.uk

http://www.ukhec.ac.uk/publications

Outline

- "Design and Building of Beowulf Class Computers"
  - Technical Report of UKHEC Collaboration
- Cost-effective high-end, departmental and desktop computation
  - Capability (ASCI) vs. Throughput requirements
- Beowulf and High-end Systems
  - Single-node performance & the Interconnect bottleneck
  - Prototype Systems - 32-CPU comparisons with Cray T3E/1200E, IBM SP/WII-375 and Compaq Alpha Server SC
- High-End Computational Chemistry
  - Distributed data structures and Global Arrays (GAs) - NWChem and GAMESS-UK; Parallel Linear Algebra (PeIGS)
- Application performance on Commodity Computers
  - Electronic Structure and Molecular Modelling
    - **GAMESS-UK, NWChem, Turbomole, DL_POLY and CHARMM**
  - Materials **and Computational Engineering**
    - **CRYSTAL, CPMD & CASTEP and engineering (ANGUS & FLITE3D)**

Scaling of Molecular Computations



Scientific Computing - Cost Effectiveness & Dependencies

Specification	Usage	Cost Units	CPU	Memory	I/O
MPP/ASCI 500-1000 CPUs, T3E/SP3 (256 GB)	HPC community	3000	500-1000	250-500	20-100
			Beowulf Systems (1.5 x N ??)		
SMP 16-processor SGI Origin 3000, R12k (8GB RAM)	Department	60	20-30	20	10-20
			Beowulf Systems (1.5 x N)		
PC Pentium-III / 1GHz (512 MByte, 18 GB)	Desktop	1	1	1	1

Linux for HPC

High Performance Computing Summer School

Organised jointly between Daresbury (DisCo and UKHEC), MRCCS and CSAR, 4th-15th September 2000

http://www.man.ac.uk/summer_school/2000

Topics covered:

- Hands-on building of a Linux cluster
- Systems administration of a Linux cluster
- Tools for parallelisation of clusters (compilers, Pallas, CAPtools…)
- Vendor / User presentations (tools, H/W, interconnects, message passing…)

http://www.ukhec.ac.uk/publications/reports/beowulf_paper.pdf

Beowulf Systems as a High-End Resource

- Need for real workloads on large clusters. To date experimental systems, so failures are viewed as inevitable
- Mechanisms to explore options - these are bewildering
  - CPUS - Alphas, single, dual, quad Pentiums, caches, Xeon or old memory subsystems; interconnect (Ethernet, Myrinet, QSW ….)
  - In the absence of "marketing / technical support", how to perform benchmarks and evaluate systems and network components
- Support expectations need to be clarified
  - Linux - largest operating system development team worldwide
    - HPC Linux
  - Open Source software
    - Queuing Systems (LobosQ, PBS - LSF expensive solution)
    - GNU Performance Issues (g77 version 2.95)
  - Growing Involvement of Traditional Vendors and Products
    - SGI (ACE), Sun, IBM (Netfinity PC servers) ……
    - Compaq Fortran and C on Linux systems (Redhat on Alpha, CXML)

Beowulf Systems as a High-End Resource II.

- National Oceanic and Atmospheric Administration (NOAA)
  - HPTi win 5 year $15M procurement to provide system(s) for weather modelling;
  - 1st Alpha Linux Cluster chosen in a major supercomputing procurement (Oct. '99)
    - Phase I - 256 Alpha EV6 nodes with Myrinet interconnect, Red Hat Linux,l 100 TByte mass storage (November '99)
    - Phase II - add 256 Alpha EV6 nodes + enhanced Myrinet (November '00)
    - Phase III - upgrade all components (256 dual EV7 nodes, 250 TByte mass storage system)
- Sandia's expansion of their Alpha-based C-plant system
  - 400 EV5-based Compaq PWs connected with Myrinet (1998)
  - 800 XP1000 PWs, 16 DS20's, StorageWorks system (September '99)
- Cornell Theory Centre
  - 256-processor Dell / Intel /Microsoft Pentium III-based Cluster
  - largest "tightly-coupled" system with hardware switch using Cornell control software
- Maui HPCC LosLobos Linux Supercluster
  - 256 IBM Netfinity PC servers (dual 733 MHz Pentium III with Myrinet (March '00)
  - National Computational Sciences Alliance

Cray, IBM & Compaq High-End Solutions

- Cray T3E/1200E
  - 816 processor system at Manchester (CSAR service)
  - 600 Mz EV56 Alpha processor with 256 MB memory (1.2 Gflop/sec peak)
- IBM SP
  - 4-way Winterhawk2 SMP “thin nodes” with 2 GB memory
  - 375 MHz Power3-II processors with 8 MB L2 cache (1.5 Gflop/sec)
  - 1.6GB/sec node memory bandwidth (single bus)
  - Switch - 150 MB/sec unidirectional, 200 MB/s bidirectional bandwidth
  - 32 CPU (8-node) system at Daresbury
- Compaq AlphaServer SC
  - 4-way ES40 SMP nodes with 2 GB memory
  - 667 MHz Alpha 21264a (EV67) CPUs with 8 MB L2 cache (1.3 Gflop/sec)
  - 5.2 GB/sec node memory bandwidth (dual bus)
  - Quadrics “fat tree” interconnect (5 usec latency, 150 MB/sec bandwidth)
  - 256 CPU system (64 nodes) in Compaq

EPSRC's Strategic Equipment Initiative (SEI)

- £30+M for University Equipment (41 bids for expt. & computing kit)
- Previous JREI competitions dominated by SMP servers (SGI - Origin 2000, Onyx2, VR).
- Emerging role of Beowulf Systems ...

Equipment Category	Number of Proposals	Funds requested from SEI (£K)
SMP Servers	7	1,970
Beowulf Systems	11	1,200
VR, visualisation	8	880
TOTAL	26	£5M+

Beowulf Chemistry Sites

- http://www.wulfpack.med.jhmi.edu (Grossfield)
  - CHARMM
- http://www.lobos.nih.gov (Brookes et al., Pentiums)
  - CHARMM, GAMESS
- http://www.hpti.com/clusterweb (Lonergan, 34 XP1000s + Myrinet)
  - CHARMM and GAUSSIAN
- http://www.soton.ac.uk/~chemphys/jessex/beowulf.html
  - MC and MD simulations.
- http://www.ccr.buffalo.edu (Furlani, 64 Ultra 5 CPUs)
  - GAMESS and CRYSTAL 95
- http://www.dhpc.adelaide.edu.au/projects/beowulf/perseus.html
- Burger, Zurich (16 PII Cluster)
  - Turbomole, DMOL, ADF
- http://www.t12.lanl.gov/~mchallacombe (26 CPU PII cluster)
  - variety of QC software
- http://zinc10.chem.ucalgary.ca (94 Alpha EV56 Compaq PW 500 au)
  - ADF and PAW

Technical Progress in 1999/2000

Hardware and Software Evaluation:

- CPU
  - PC systems - Intel 866 MHz Pentium III, Athlon (AMD K7) 650, 850, 1000 MHz
  - Alpha systems - DS10, DPC-264, DS20, ES40, XP1000, UP2000
- Networks
  - Fast Ethernet options, cards, switches, channel-bonding, ….
  - 100Mbit switch,
  - Myrinet interconnect (3 X dual-CPU UP2000 evaluation system, from Compusys)
- System Software
  - message passing S/W (LAM MPI, LAM MPI-VIA (100 us to 60 us), MPICH),libraries (ATLAS, NASA), compilers (Absoft, PGI, GNU/g77), GA tools (PNNL)
  - resource management software (LobosQ, PBS, Beowulf, **LSF** etc.)

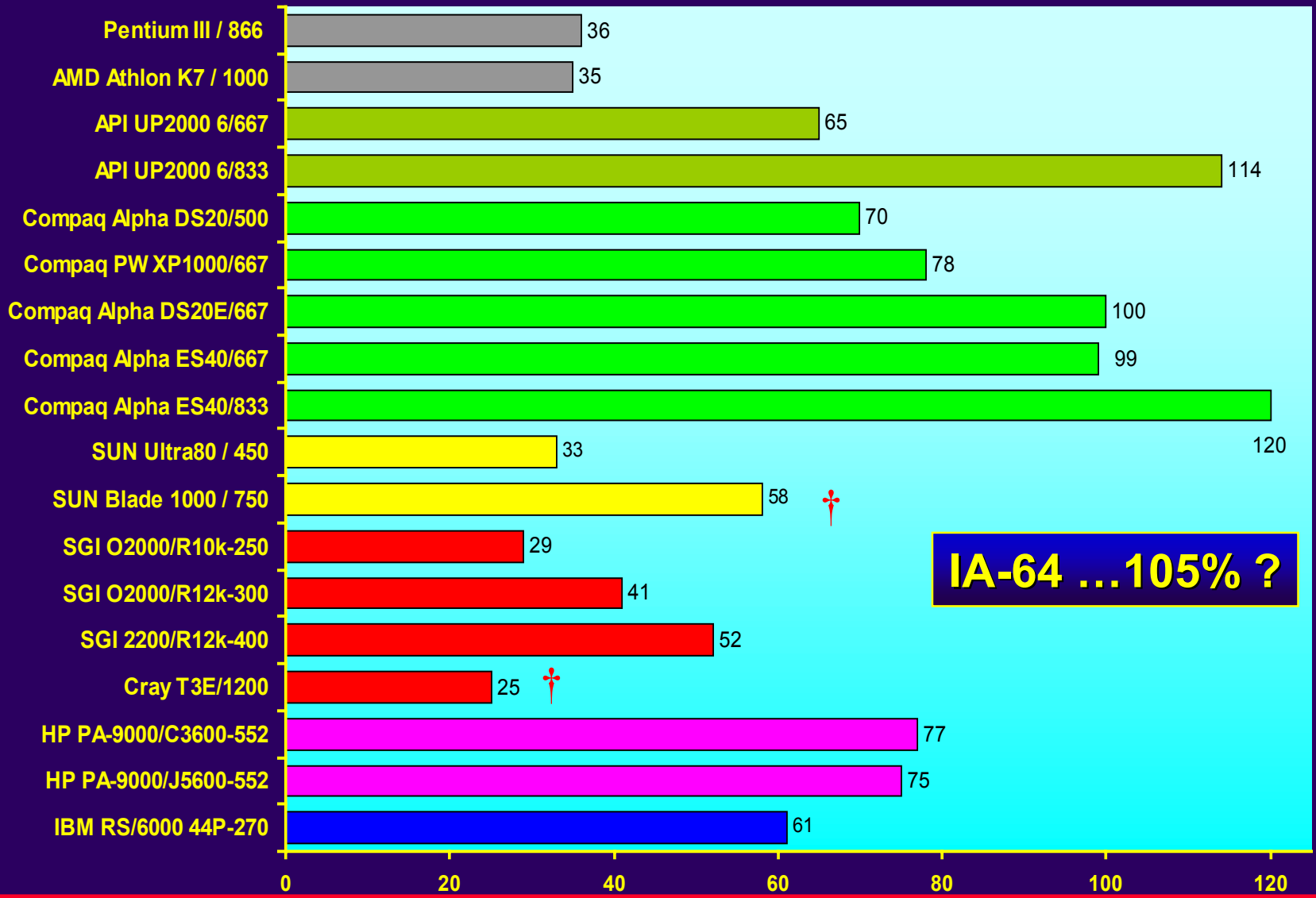
Technical Progress in 1999/2000 II.

- Application Porting and Performance Analysis
  - Computational Chemistry and Materials
  - Computational Engineering
  - Computational Biology
  - Climate Modelling
- Performance Modelling
  - Collaboration with Pallas GMBH (Vampir and Dimemas)
- Vendor Interactions:
  - Integration Companies - Workstations UK Ltd., Compusys/InSiliCo Ltd
  - Compaq, QSW (QsNet), IBM, Packet Engines, Extreme Networks
  - Sychron Ltd., Platform Computing
  - Pallas GMBH

Beowulf Prototype / Evaluation Systems

Evaluation Systems	Location	Configuration
Beowulf II	Daresbury	32 Pentium III / 450 MHz; fast ethernet (EPSRC)
Beowulf III (Loki)	Daresbury	17 X dual UP2000/EV67-667, QSNet Alpha/LINUX cluster
Beowulf IIIa	Daresbury	16 X dual UP2000/EV67-667, QSNet Alpha/LINUX cluster
Beowulf IV (Wulfgar)	RAL	16 X Athlon AMD K7 850MHz; myrinet interconnect
Beowulf V (Bobcat)	EPCC	16 X Athlon AMD K7 650MHz; fast ethernet
Protoype Systems		
Beowulf I	Daresbury	10 CPUS, Pentium II/266
Compusys/Insilico	Daresbury	3 X dual UP2000/EV67, myrinet Interconnect
Athlon system	RAL	networking test-bed of 5 AMD K7 Athlon CPUs

SPECfp95. Values relative to Compaq Alpha DS20E/667 (83.6)



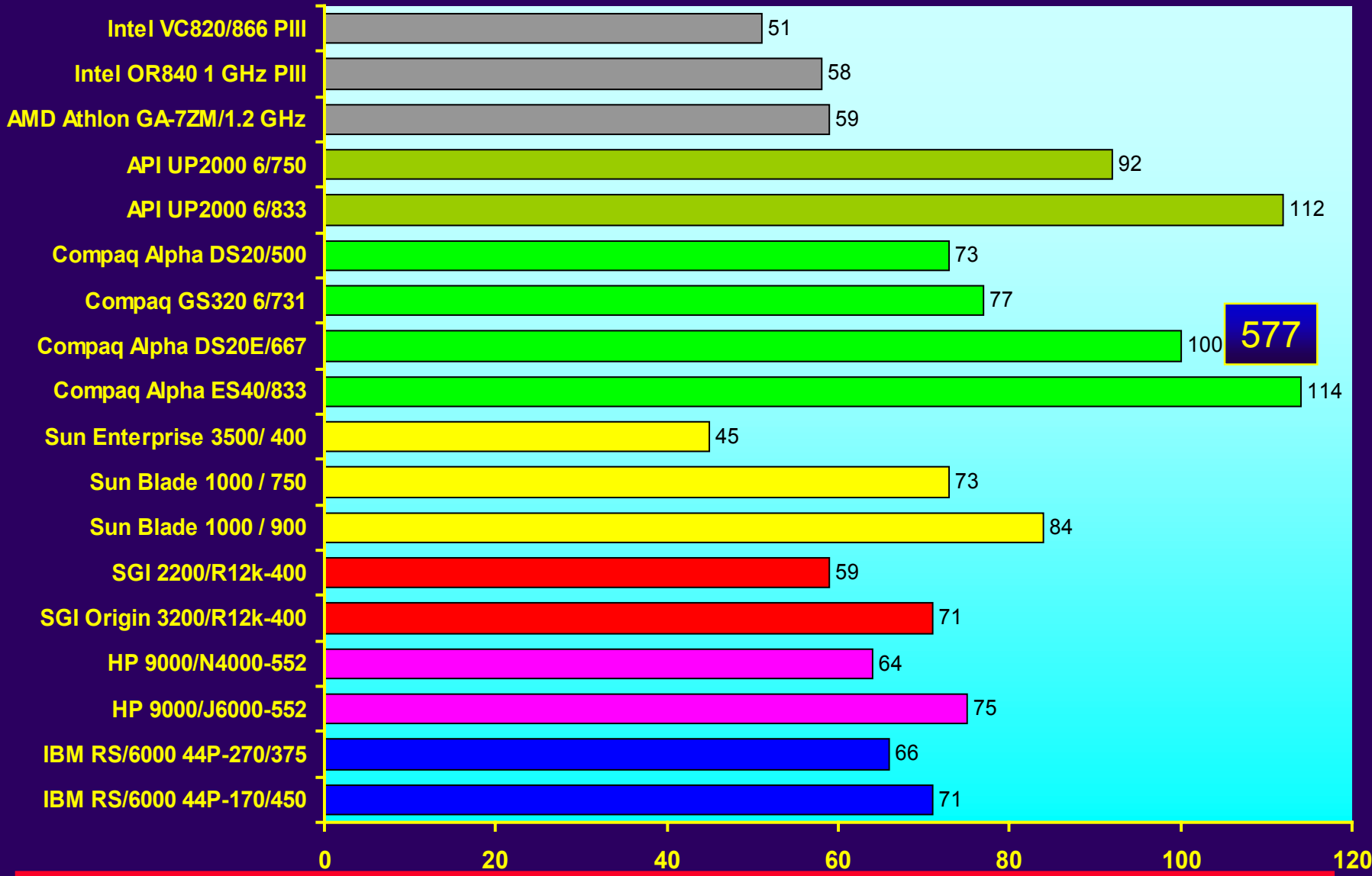
SPEC CPU 2000 - Floating point Benchmark Suite (SPECfp2000)

Benchmark	Language	Description
168.wupwise	F77	Physics: Quantum chromodynamics
171.swim	F77	Shallow water modelling
172.mgrid	F77	Multigrid solver: 3D potential field
173.applu	F77	Partial differential equations
177.mesa	C	3D graphics library
178.galgel	F90	Computational fluid dynamics
179.art	C	Image recognition / neural networks
183.equake	C	Seismic wave propagation simulation
187.facerec	F90	Image processing: Face recognition
188.ammp	C	Computational chemistry
189.lucas	F90	Number theory / primality testing
191.fma3d	F90	Finite-element crash simulation
200.sixtrack	F77	Nuclear physics accelerator design
301.apsi	F77	Metereology: Pollutant distribution

Reference: 300 MHz Ultra 5/10 = 100

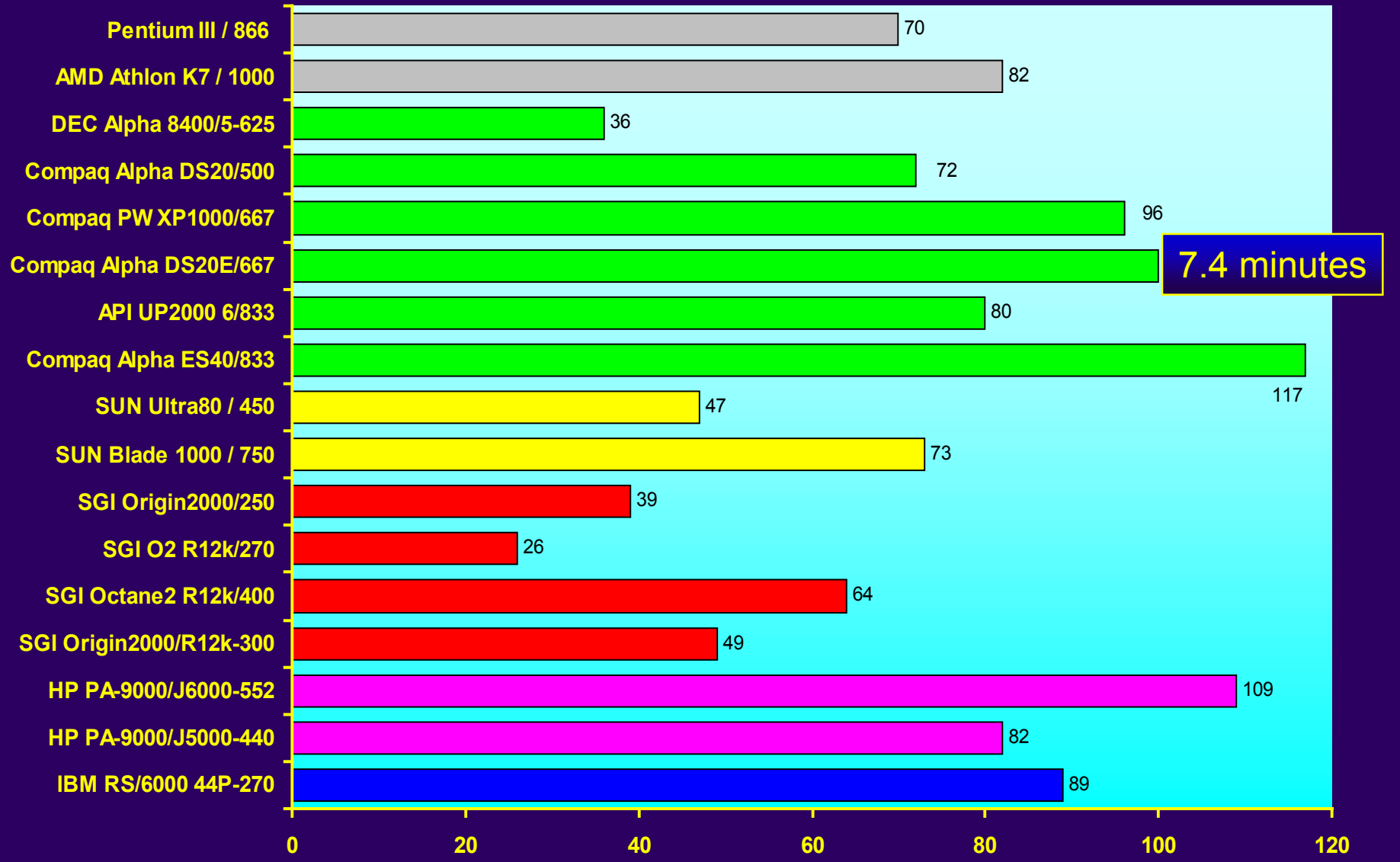
SPEC CPU 2000 - SPECfp2000

Values relative to Compaq Alpha DS20E/667

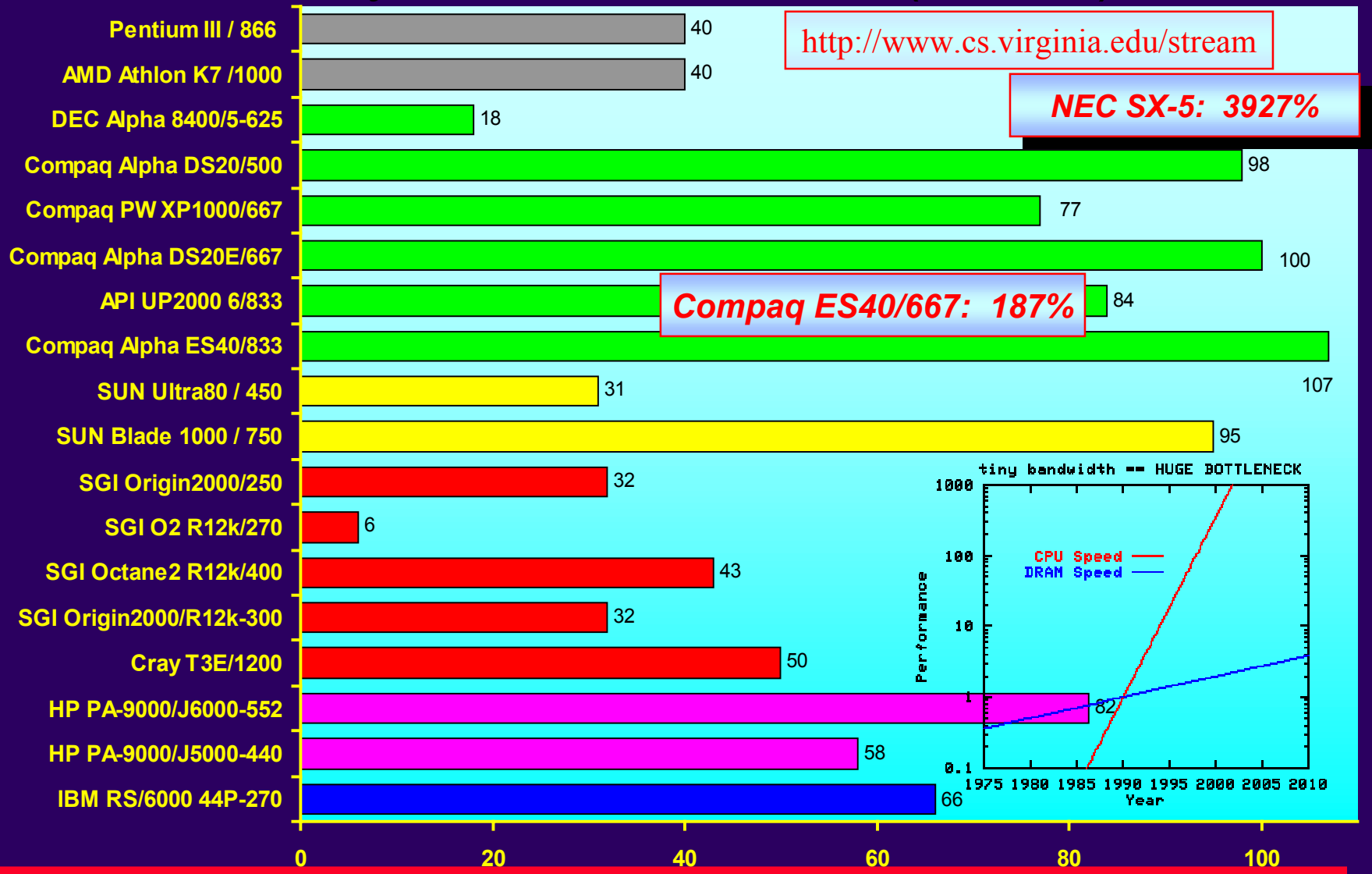


The GAMESS-UK Benchmark

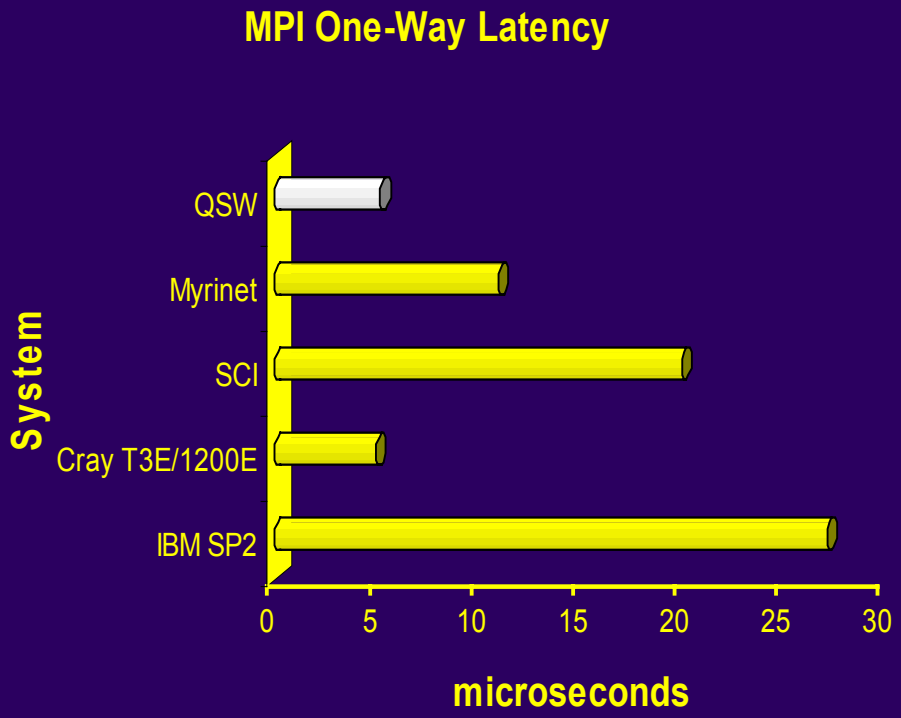
CPU Performance relative to the Compaq Alpha DS20E/667

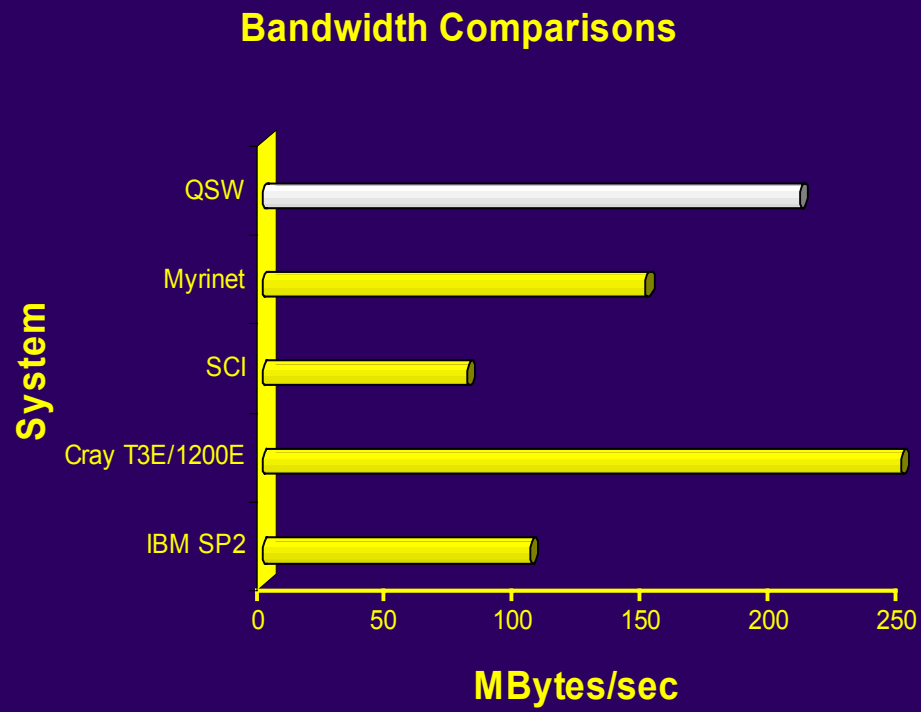


STREAM: Measured Sustainable Memory Bandwidth in HPC (TRIAD)



Interconnect Comparison - MPI





References
SP Switch Performance (IBM Document) 1998
Fujitsu April 1997
HLRS Germany 1997
NASA Ames Laboratory 1997
QSW 1998

Communications Benchmark
PMB: Pallas MPI Benchmark Suite (V2.2)

Point-to-Point (Mbytes/sec)

PingPong; PingPing; Sendrecv; Exchange

Collective Operations (Time - usec) - as function of no.of CPUs

Allreduce; Reduce; Reduce_scatter; Allgather; Allgatherv; Alltoall; Bcast, Barrier

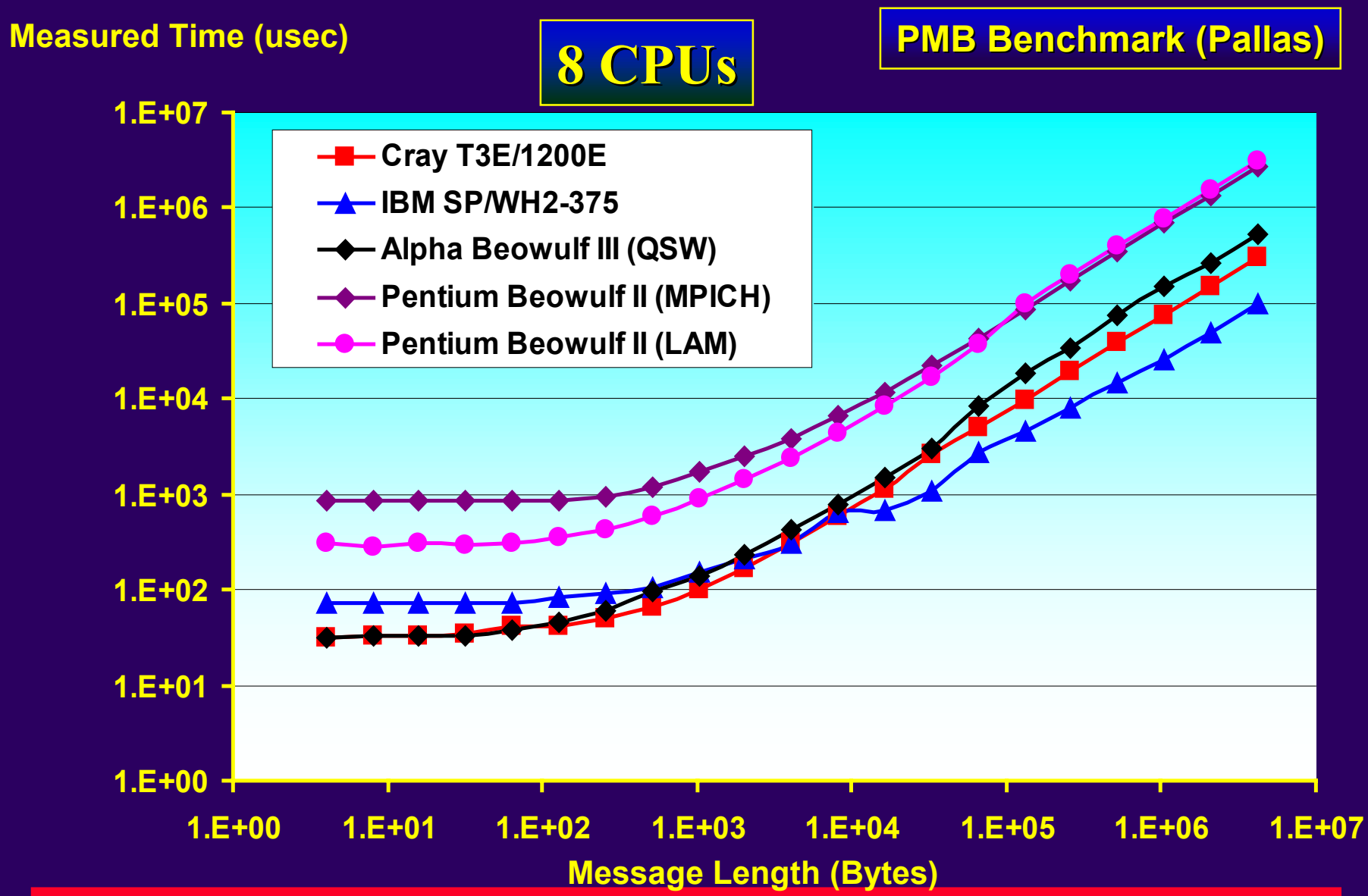
Message Lengths:

0 to 4194304 Bytes

Systems Investigated

- Cray T3E/1200E
- IBM SP/WH2-375 - using both 2- & 4- CPUs per quad-SMP node
- Pentium Beowulf II (450 MHz) - MPICH and fast ethernet
- AMD Athlon/850 MHz / Myrinet Beowulf IV (MPICH)
- Linux Alpha Beowulf III - both single and dual CPUS per UP2000 node

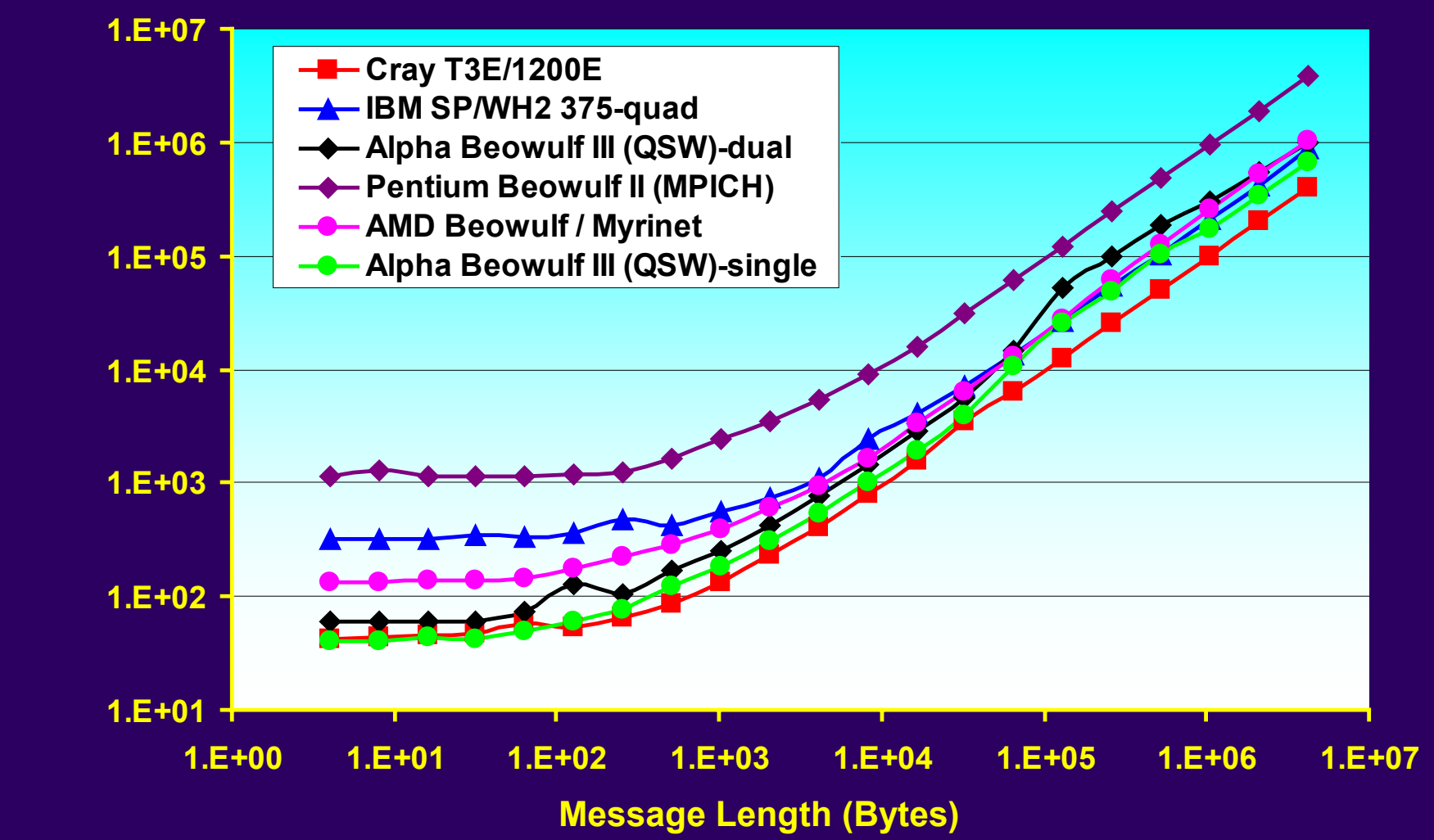
MPI_allreduce Performance



16 CPUs

MPI_allreduce Performance

PMB Benchmark (Pallas)



Beowulf Prototype Systems - I and II

- Collaboration with LoBoS project at NIH
- Prototype Systems - LAM 6.2 (MPI), compilers (GCC 2.95, PGI, Absoft etc.)
  - Beowulf I (~ £13K):
    - Master Node
      - Pentium II-266, 128 MByte, 9 GByte UW-SCSI disk, 4 MByte graphics cards, monitor, network card, keyboard, CD-ROM
    - Compute Nodes (8)
      - 6 x Pentium-266 (128 MB), 2 x dual Pentium II-266 (256 MB); all with 4 GByte UW-SCSI disk, 3 network cards, minimal graphics. 12 port 100Base-T Fast Ethernet switch
  - Beowulf II (~ £50K):
    - Master node
      - Pentium II-266, 256 MB; 4 GB system disk, 9 GB UW-SCSI user disk
    - Compute nodes (32)
      - Pentium III-450, 256 MB ECC SDRAM, 10 GB UDMA disk
      - connected to the master via a Fast Ethernet switch
      - All nodes are also connected to each other via a second Fast Ethernet switch dedicated to the parallel communications traffic.

www.cse.clrc.ac.uk/Activity/DisCo

Steven Andrews *s.andrews@dl.ac.uk*

Barry Searle *b.searle@dl.ac.uk*

Beowulf Prototype Systems - III

- Master node
  - UP2000 with 2 X 667 MHz Alpha 21264A (4 Mbytes L2 cache), 2 GB SDRAM; 36 GB UW-SCSI user disk
- 16 X Compute nodes (UP2000)
  - UP2000 with 2 X 667 MHz Alpha 21264A, 1 GB SDRAM, 9 GB SCSI disk
  - QSNet PCI Adaptor + 2 X 100 Mb/s Ethernet NIC (optional Beowulf + private management network)
- High Performance Interconnect:
  - 16-way QsNet Elite switch
  - 2 X 24-way Ethernet switch (Beowulf)
- System Software
  - RH 6.2 LINUX, Compaq C/Fortran compilers, libraries etc., GCC, RMS for LINUX, MPICH, IP and SHMEM libs.
- QUASI Application system
  - Above system to be duplicated

QSNet Alpha/LINUX Cluster Installation - May '00

QSNet Alpha/LINUX Cluster

- Performance of Collective MPI operations
- Shared Memory Performance:
  - Cost effectiveness of UP2000 stems from use of commodity SDRAMS (c.f. STREAM figures) and 1/2 number of paths to memory as Compaq DS20 platform (effect on inter-node MPI performance);
  - Expected to impact on applications with heavy memory usage
- Effect of variations in messaging schemes:
  - Programming model for QLC-alpha uses flat MPI/SHMEM model implemented across clustered SMP solutions.
  - Intra node comms. Supported as either direct memory copies or via Elan NIC. Default for intra-node in MPI is shared memory while in SHMEM interface default is to use Elan. Complex - messages currently poll ...
- LINUX Issues and Performance:
  - Page colouring: support for multi-way associative caches (L2) to provide optimum strategy for cache reuse.

Application Codes

Performance comparisons between Beowulf systems and both high-end MPP (CSAR Cray T3E/1200E) and ASCI-style SMP-node platforms (IBM SP / Winterhawk II-375, Compaq AlphaServer SC (ES40/6-667 + QSW):

- Chemistry and Materials
  - **DL_POLY - parallel MD code with many applications**
  - **GAMESS-UK, NWChem & Turbomole - Ab initio Electronic structure codes**
  - **CRYSTAL, VASP, CASTEP** - UKCP Car-Parrinello total energy materials code
- Molecular Biology
  - **CHARMM (NIH and BASF), DL_POLY, …**
- Engineering
  - **ANGUS** - regular-grid domain decomposition engineering code with conjugate-gradient and multi-grid solvers
  - **FLITE3D** - finite-element irregular-grid engineering code
- Climate Modelling
  - **Unified Model** (Atmospheric Climate Modelling)

Performance Metric (% 32-node Cray T3E)

DL_POLY: A Parallel Molecular Dynamics Simulation Package

- Developed as CCP5 parallel MD code by W. Smith and T.R. Forester
- UK + International user community
- Adopted by Materials Consortium 1995

Boundary Conditions

- None (e.g. isolated macromolecules)
- Cubic periodic boundaries
- Orthorhombic periodic boundaries
- Parallelpiped periodic boundaries
- Truncated octahedral periodic boundaries
- Rhombic dodecahedral periodic boundaries
- Slabs (i.e. x,y periodic, z nonperiodic)

Target Systems

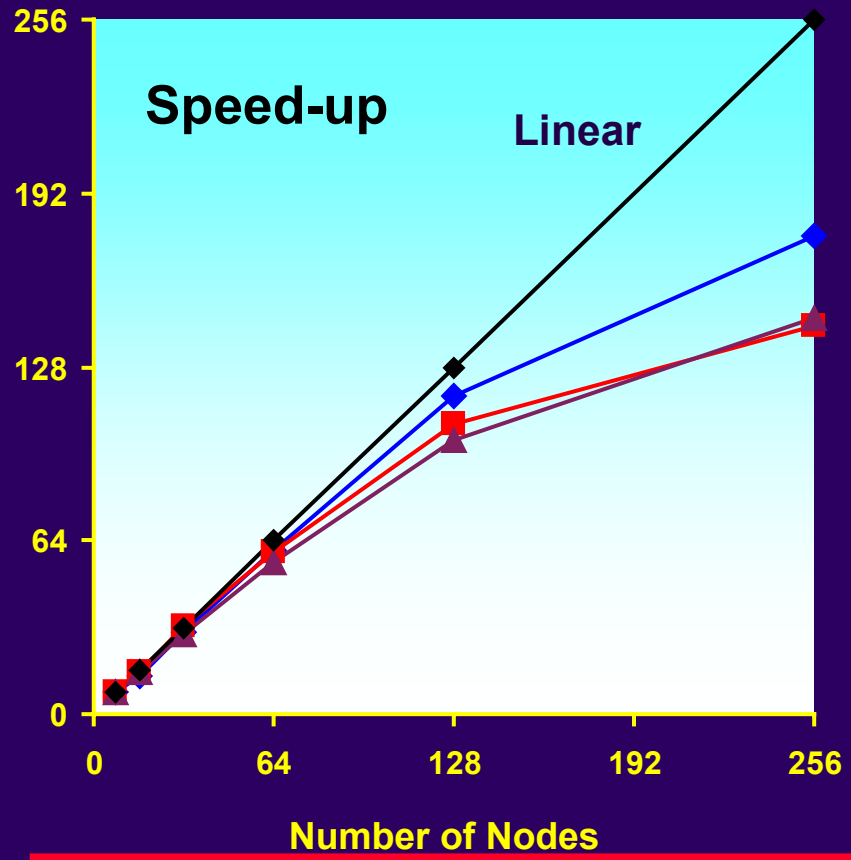
- Atomic systems & mixtures (Ne, Ar, etc.)
- Ionic melts & crystals (NaCl, KCl etc.)
- Polarisable ionics (ZSM-5, MgO etc.)
- Molecular liquids & solids (CCl_4, Bz etc.)
- Molecular ionics (KNO_3, NH_4Cl, H_2O etc.)
- Synthetic polymers ($[PhCHCH_2]_n$etc.)
- Biopolymers and macromolecules
- Polymer electrolytes, Membranes,
- Aqueous solutions, Metals

MD Algorithms/Ensembles

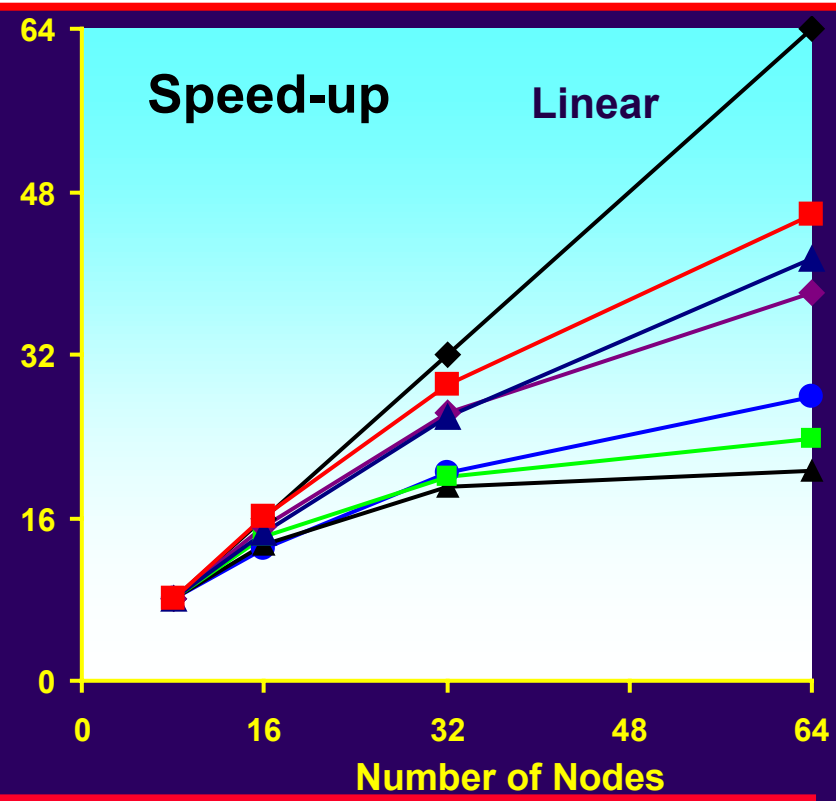
- Verlet leapfrog, Verlet leapfrog + RD-SHAKE
- Rigid units with FIQA and RD-SHAKE
- Linked rigid units with QSHAKE
- Constant T (Berendsen) with Verlet leapfrog and with RD-SHAKE
- Constant T (Evans) with Verlet leapfrog and with RD-SHAKE
- Constant T (Hoover) with Verlet leapfrog

DL_POLY Parallel Benchmarks (Cray T3E/1200)

4. NaCl; MTS Ewald, 27,000 ions
5. NaK-disilicate glass; 8,640 atoms, Ewald
8. MgO microcrystal: 5,416 atoms



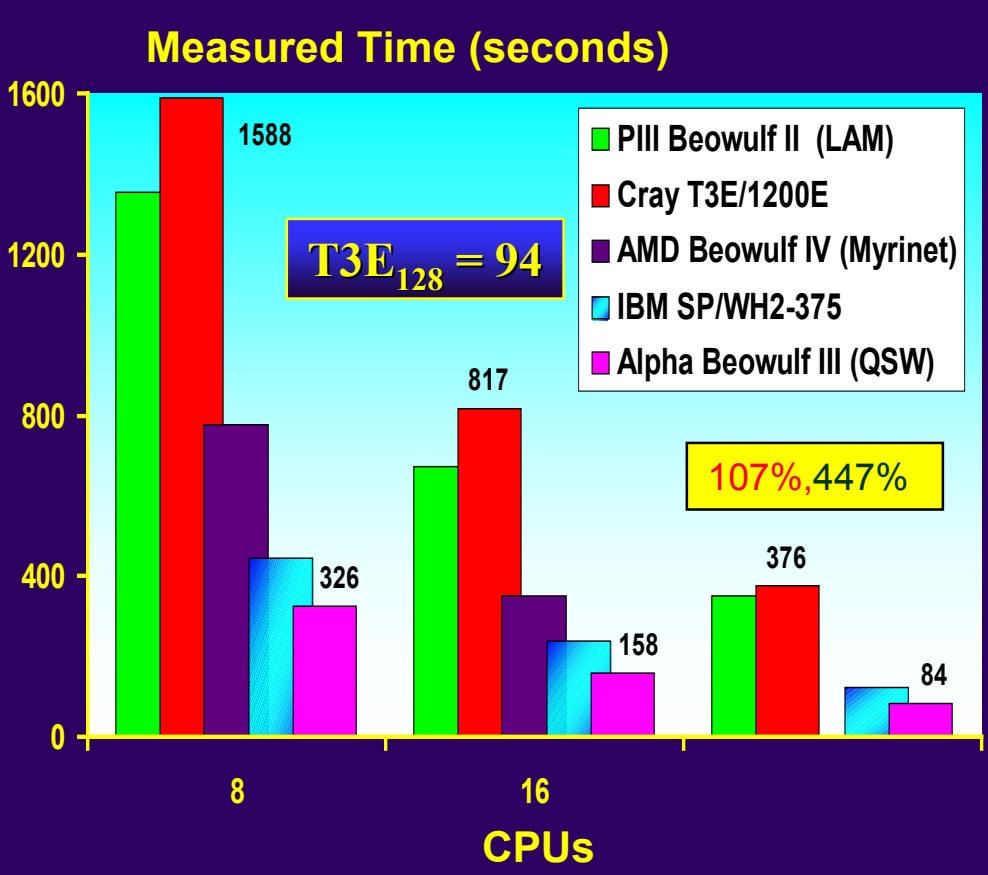
9. Model membrane/Valinomycin (MTS, 18,886)
7. Gramicidin in water (SHAKE, 13,390)
6. K/valinomycin in water (SHAKE, AMBER, 3,838)
1. Metallic Al (19,652 atoms, Sutton Chen)
3. Transferrin in Water (neutral groups + SHAKE, 27,593)
2. Peptide in water (neutral groups + SHAKE, 3993).

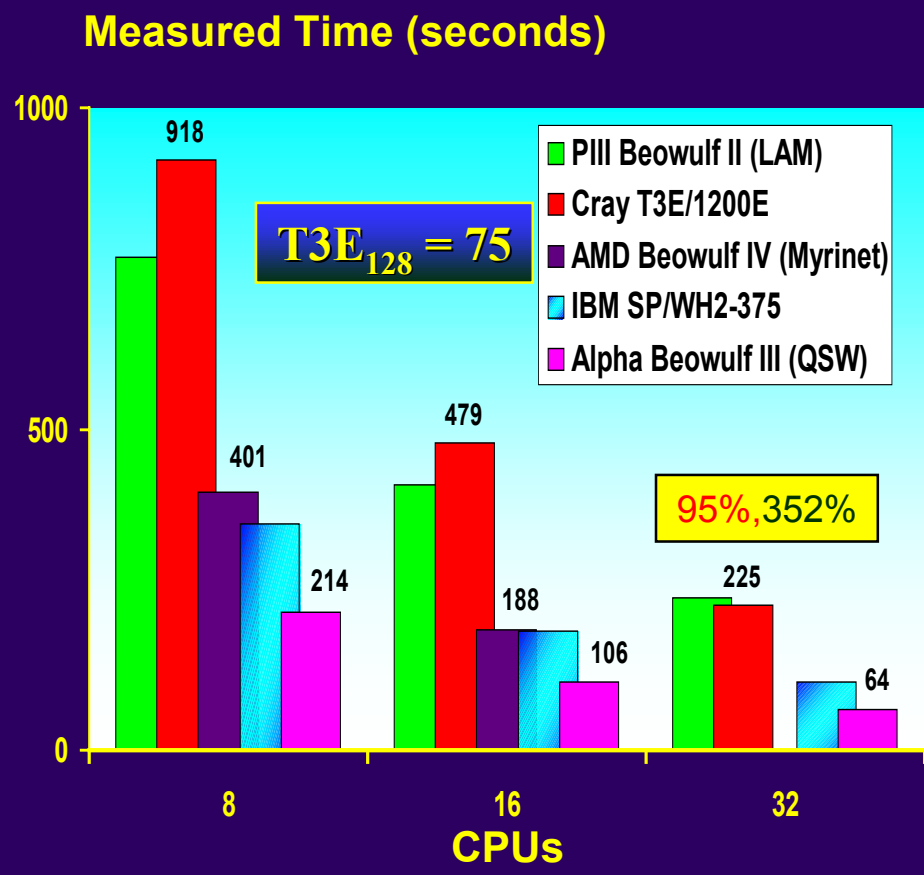


DL_POLY: Cray/T3E, IBM SP & Beowulf Systems

Bench 4. Sodium Chloride;
MTS Ewald, 27,000 ions, 200 time steps, Cutoff = 24Å

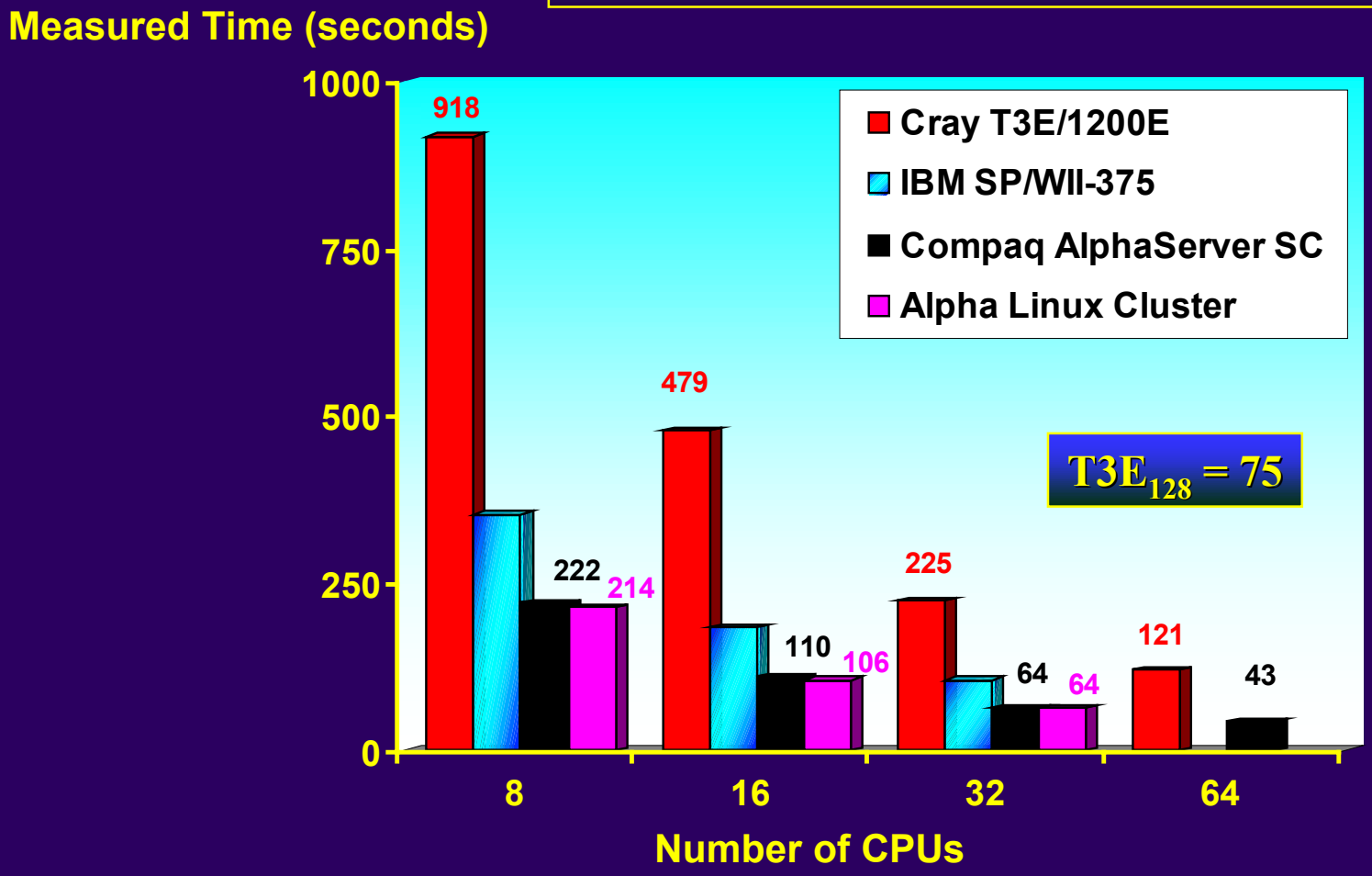
Bench 5. NaK-disilicate glass;
8,640 atoms, Ewald, 270 time steps, Cutoff=12Å





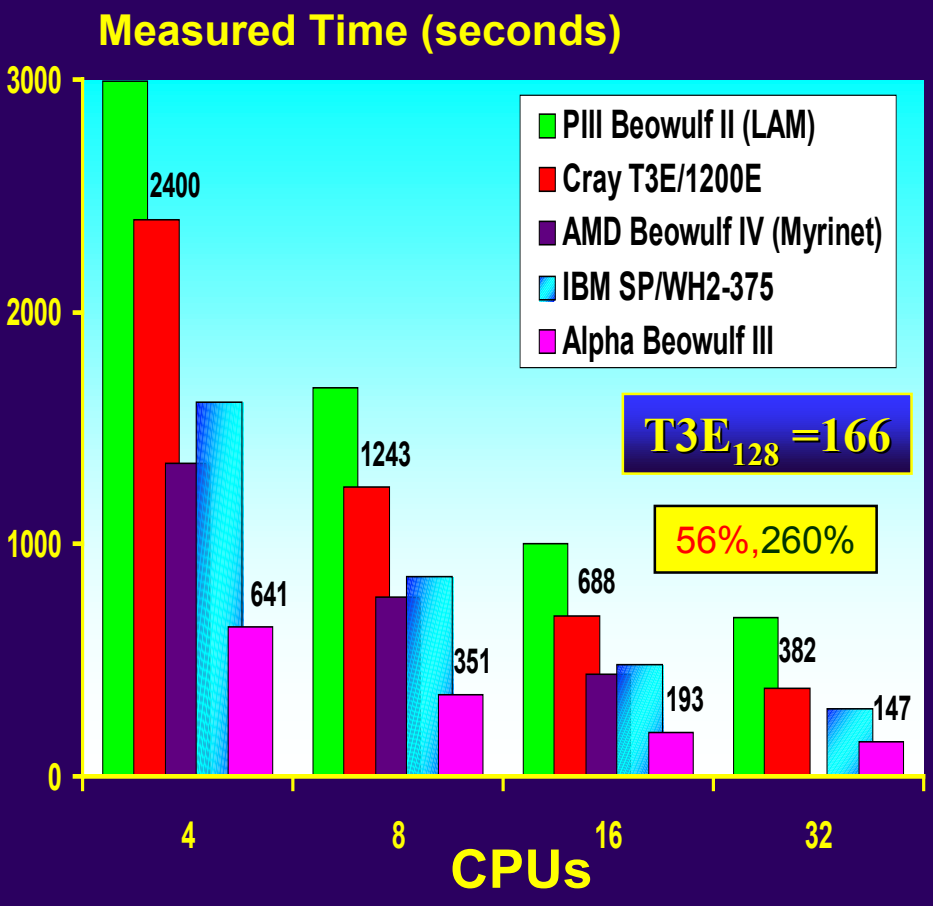
DL_POLY: Alpha Linux Cluster and Compaq AlphaServer SC

Bench 5. NaK-disilicate glass; 8,640 atoms, Ewald

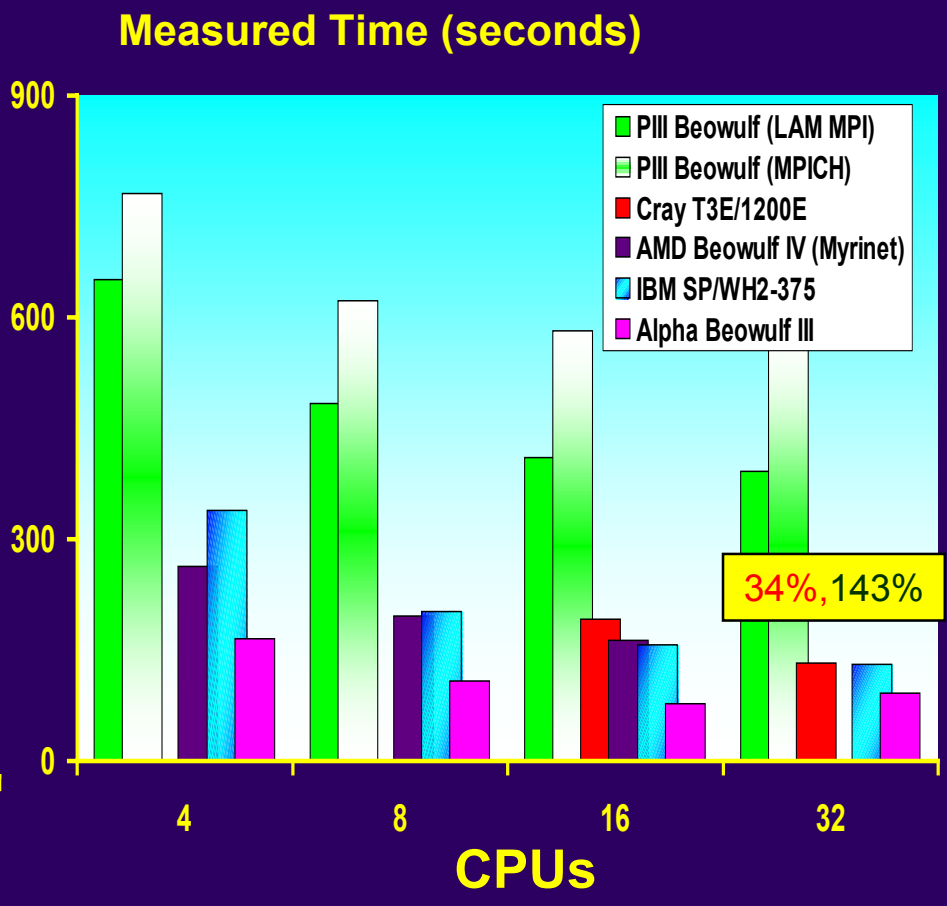


DL_POLY: Cray/T3E, IBM SP & Beowulf Systems

Bench 7. Gramicidin in water;
rigid bonds and SHAKE, 13,390 atoms, 500 time steps

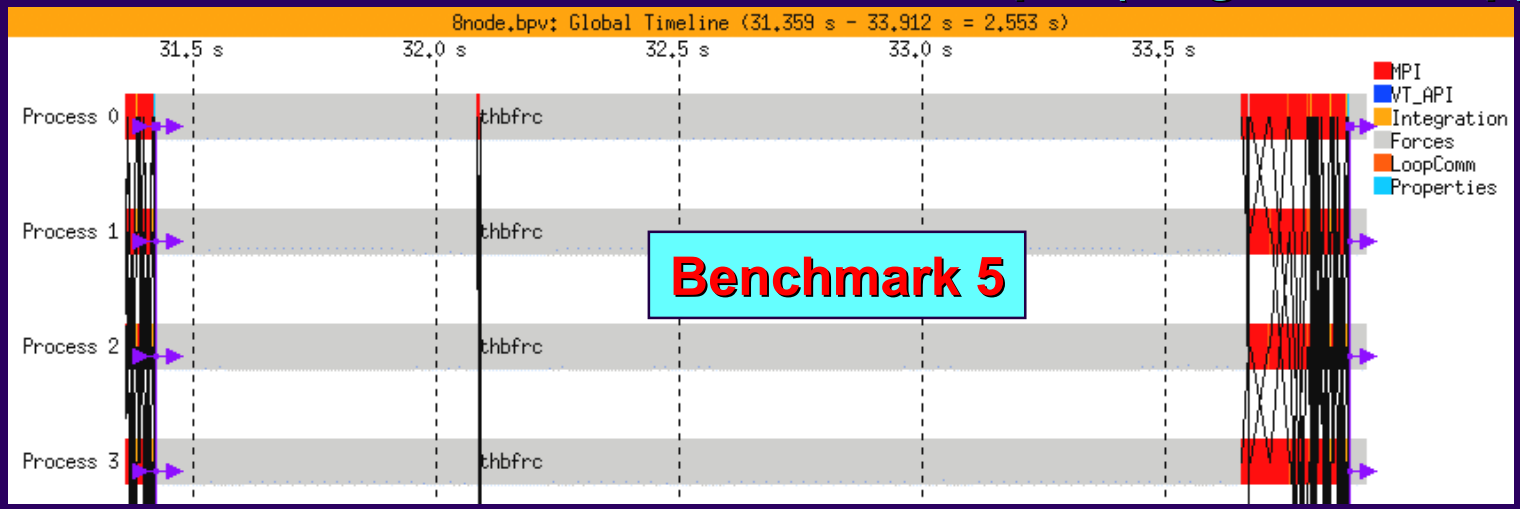


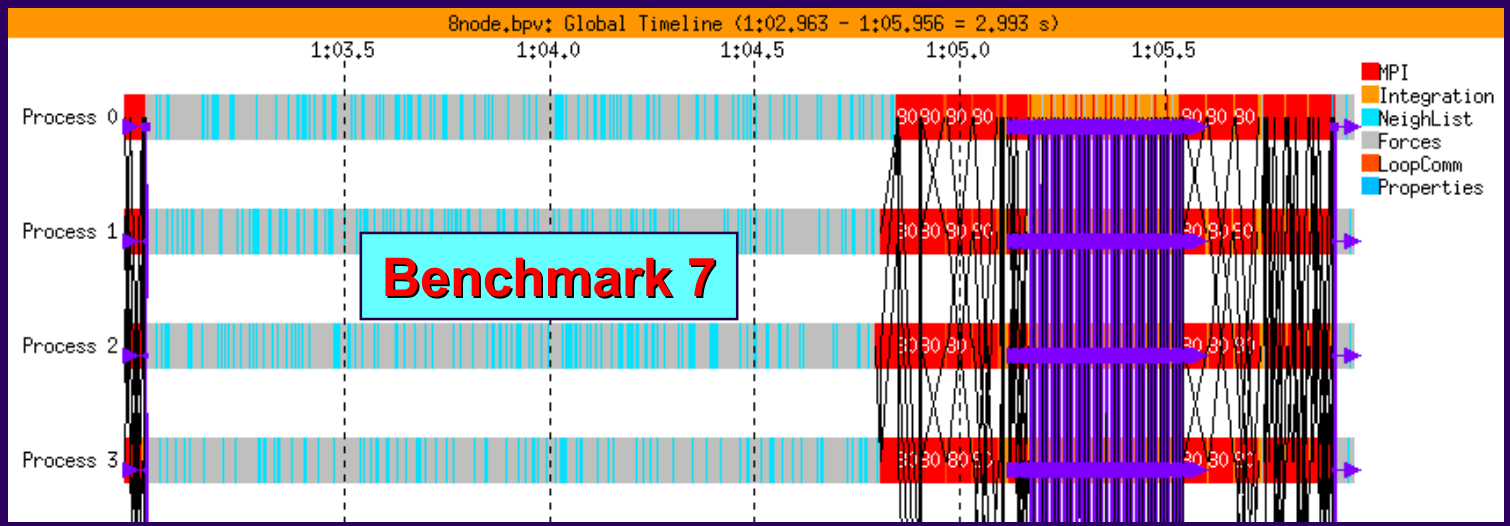
Bench 3. Transferrin in Water;
neutral groups + SHAKE, 27,593 atoms, 250 time steps



Vampir

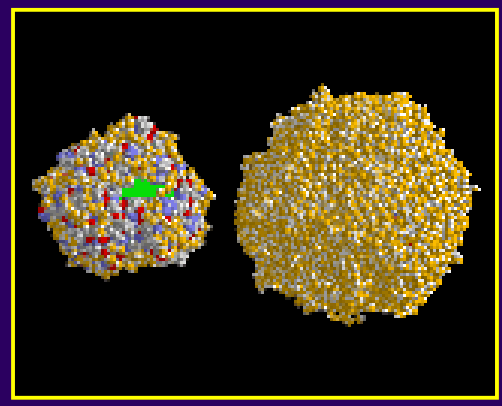
Beowulf II- 8 node DLPOLY trace output (single timestep)





CHARMM

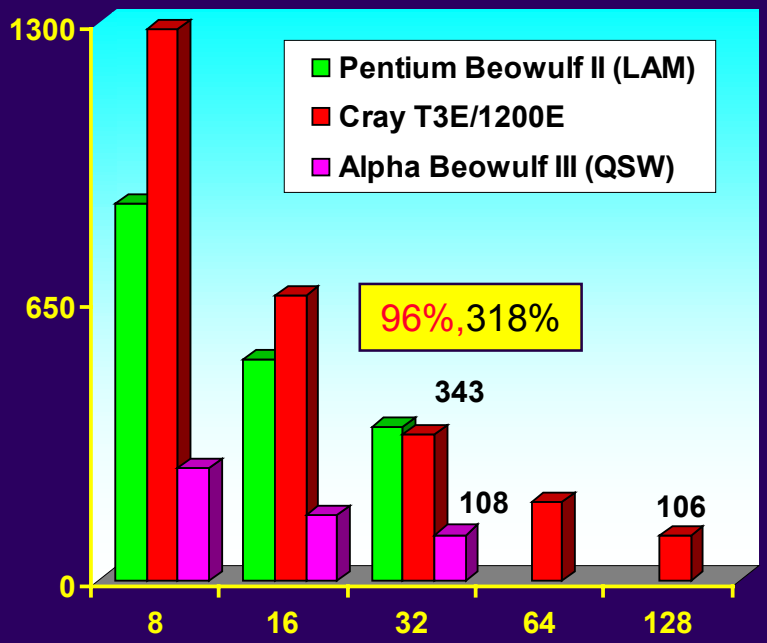
CHARMM (c26b2) is a general purpose molecular mechanics, molecular dynamics and vibrational analysis package for modelling and simulation of the structure and behaviour of molecular systems.

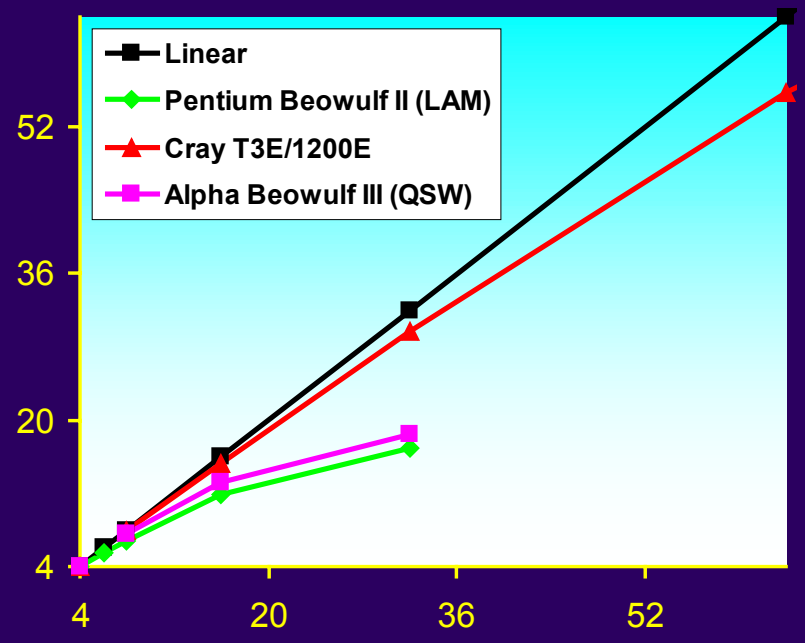


Benchmark MD Calculation of Carboxy Myoglobin (MbCO) with 3830 Water Molecules.

14026 atoms, 1000 steps (1 ps), 12-14 A shift using standard parallel CHARMM

Measured Time (seconds)

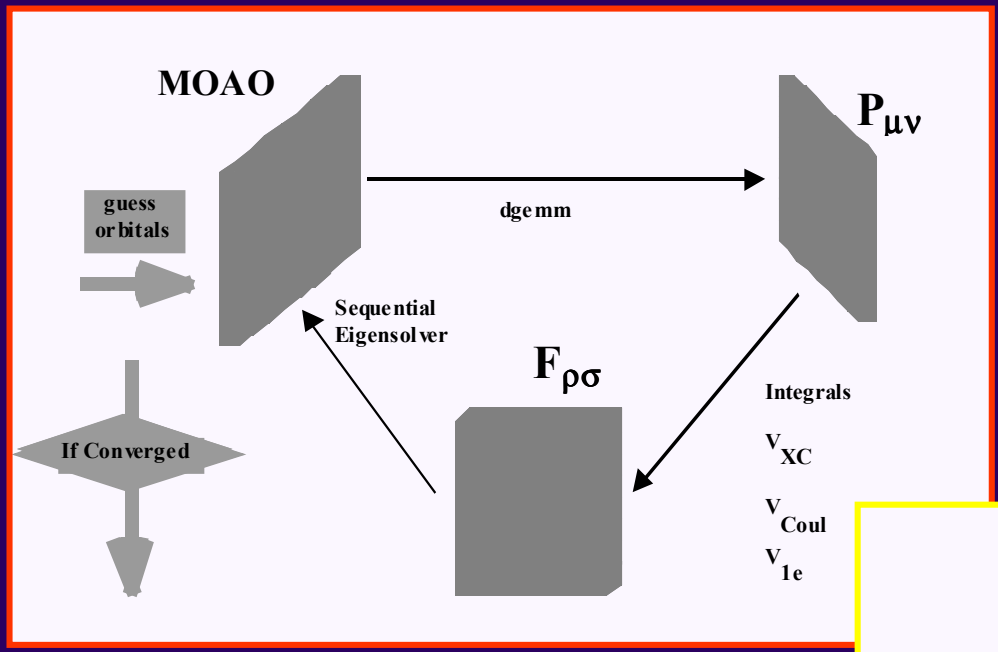




High-End Computational Chemistry The NWChem Software

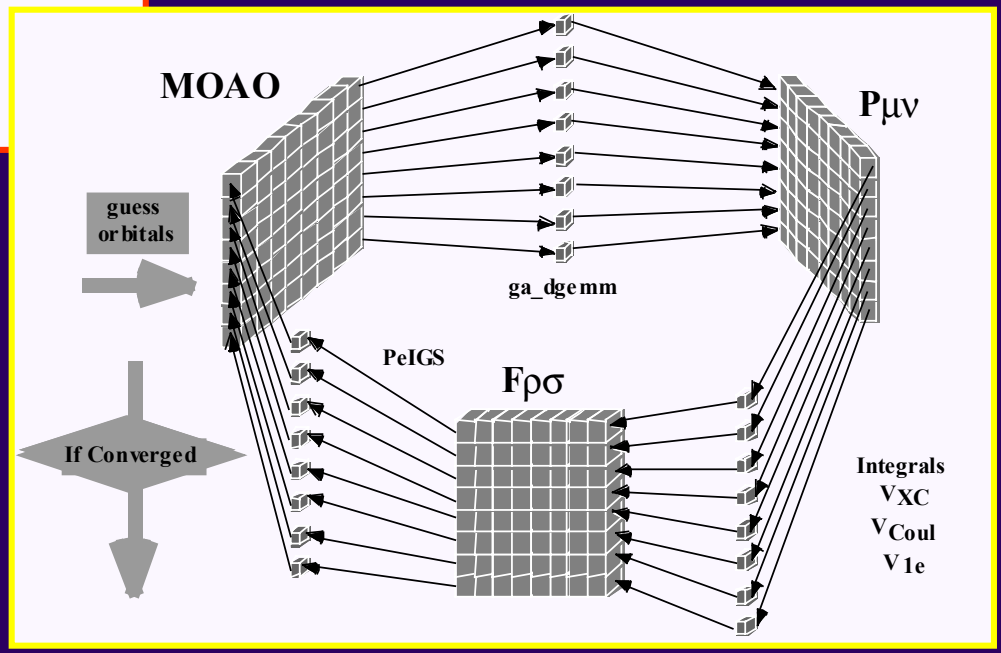
- Developed as part of the construction of the Environmental Molecular Sciences Laboratory (EMSL) at PNNL.
- Funded to be used as an integrated component in solving DOE's grand challenge environmental restoration problems
- Designed and developed to be a highly efficient and portable MPP computational chemistry package, providing computational chemistry solutions which are scalable with respect to chemical system size as well as MPP hardware size
- Extensible framework supporting development of new methods in computational chemistry; NWChem Architecture
  - Object-oriented design
    - abstraction, data hiding, handles, APIs
  - Parallel programming model
    - non-uniform memory access, global arrays (GAs)
  - Infrastructure
    - **Global Arrays (GA)** , Parallel I/O, RTDB, MA, **Linear algebra (PeiGS)** ...

Distributed Data SCF



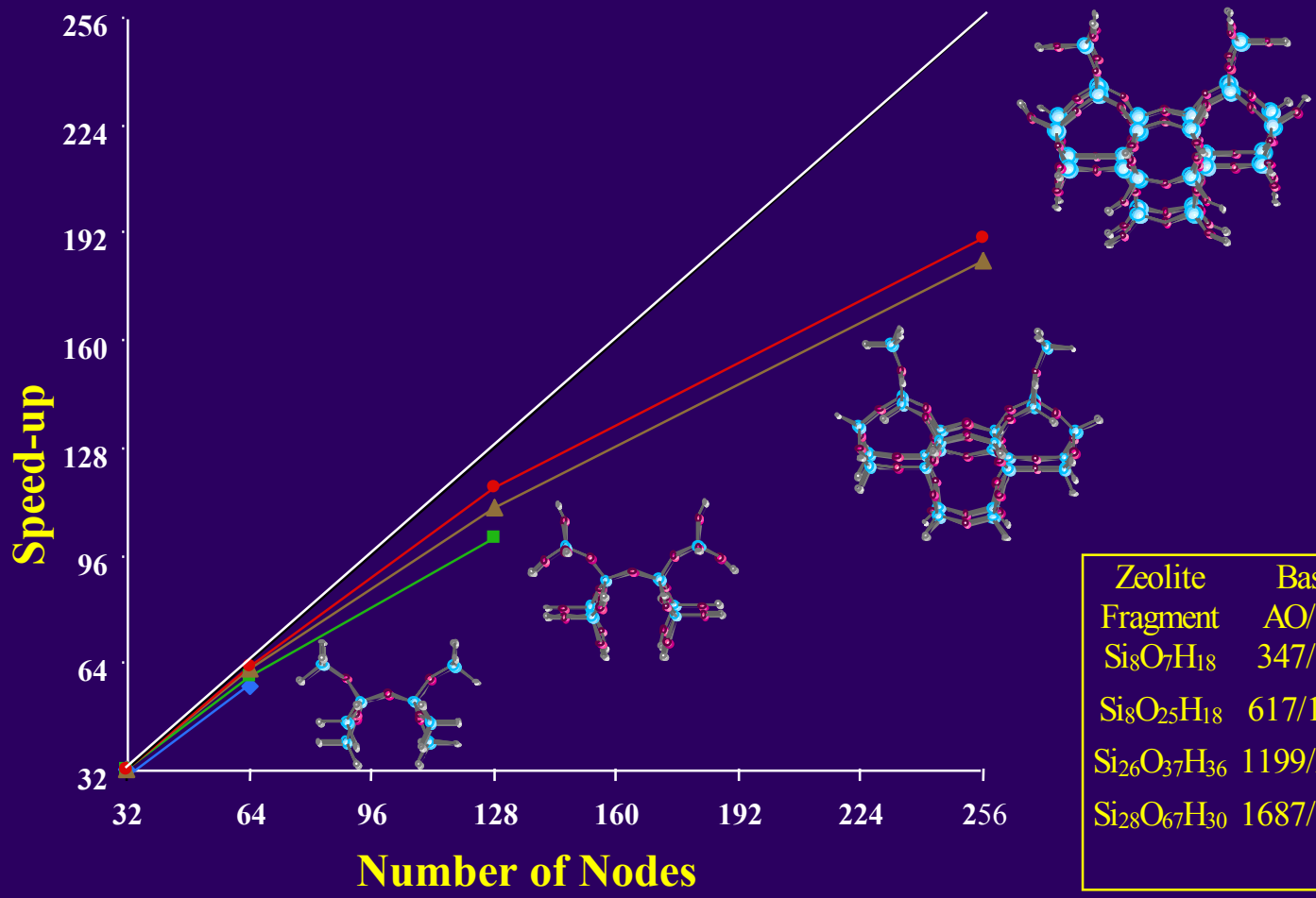
Pictorial representation of the iterative SCF process in (i) a sequential process, and (ii) a distributed data parallel process: **MOAO** represents the molecular orbitals, **P** the density matrix and **F** the Fock or Hamiltonian matrix

Sequential



Distributed Data

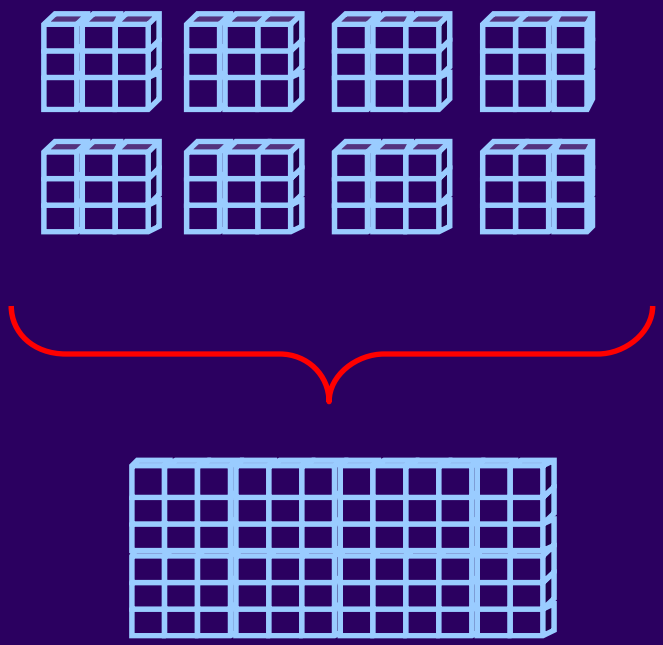
Measured Parallel Efficiency for NWChem - DFT on IBM-SP; Wall Times to Solution for SCF Convergence



Zeolite Fragment	Basis AO/CD	Number of Nodes	Wall Time to Solution
$Si_8O_7H_{18}$	347/832	64	238s
$Si_8O_{25}H_{18}$	617/1444	128	364s
$Si_{26}O_{37}H_{36}$	1199/2818	256	1137s
$Si_{28}O_{67}H_{30}$	1687/3928	256	2766s

Global Arrays

Physically distributed data



Single, shared data structure

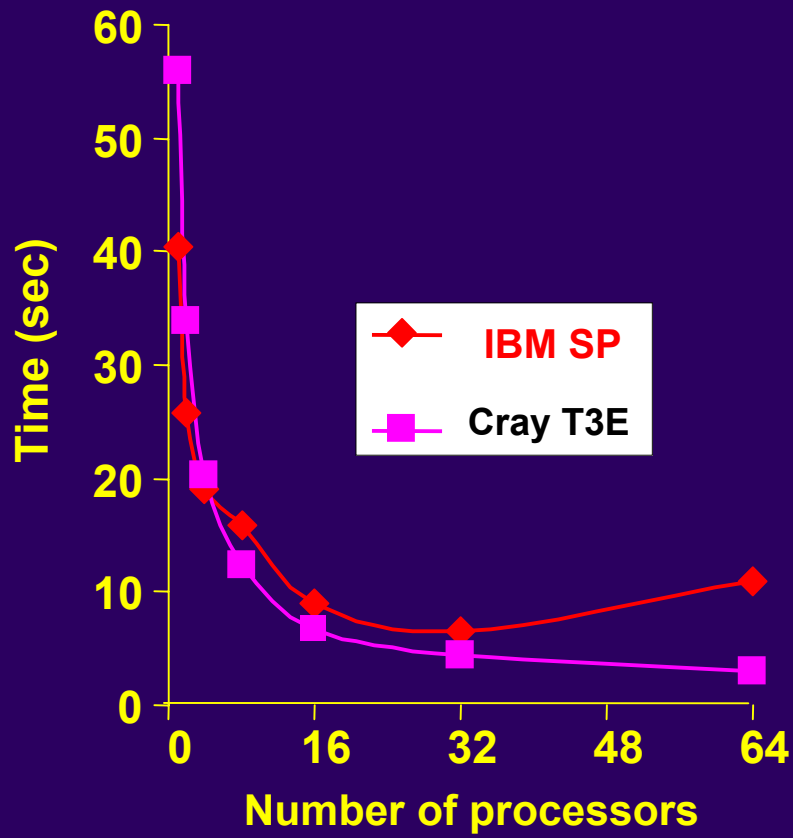
- Shared-memory-like model
  - Fast local access
  - NUMA aware and easy to use
  - MIMD and data-parallel modes
  - Inter-operates with MPI, …
- BLAS and linear algebra interface
- Ported to major parallel machines
  - IBM, Cray, SGI, dusters,...
- Originated in an HPCC project
- Used by 5 major chemistry codes, financial futures forecasting, astrophysics, computer graphics

PeIGS 3.0 Parallel Performance

(Solution of real symmetric generalized and standard eigensystem problems)

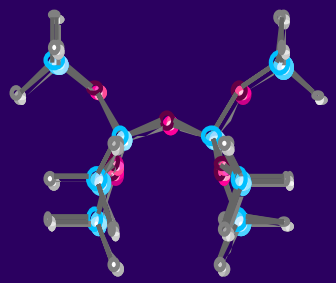
Features (not available elsewhere):

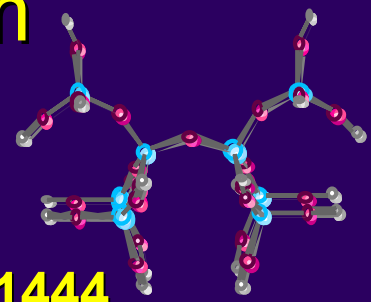
- **Inverse iteration using Dhillon-Fann-Parlett's parallel algorithm (fastest uniprocessor performance and good parallel scaling)**
- **Guaranteed orthonormal eigenvectors in the presence of large clusters of degenerate eigenvalues**
- **Packed Storage**
- **Smaller scratch space requirements**



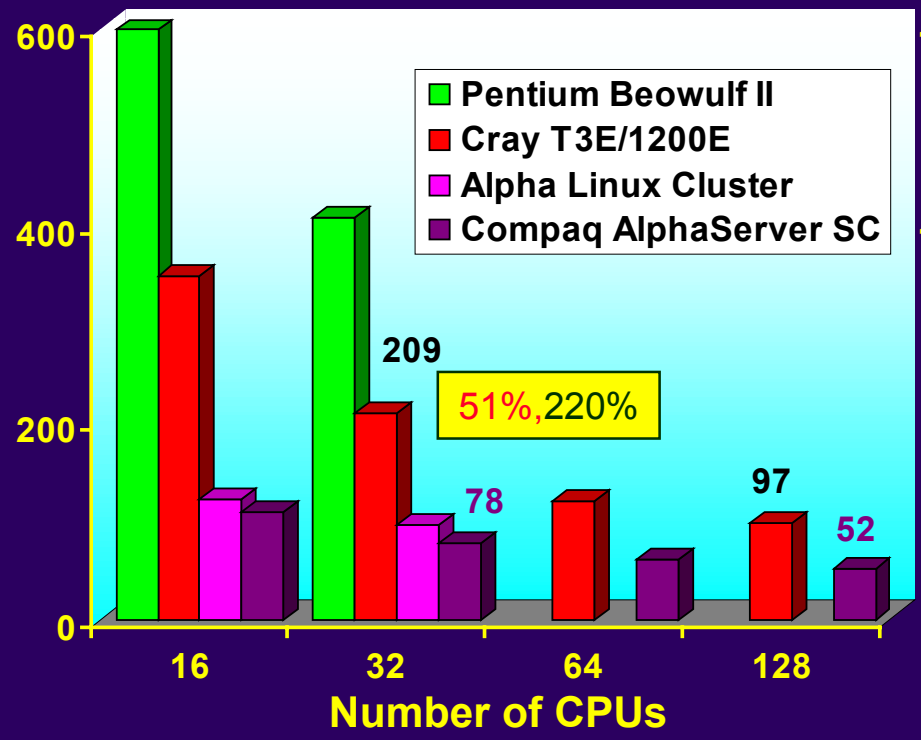
Full eigensolution performed on a matrix generated in a charge density fitting procedure (966 fitting functions for a fluorinated biphenyl).

DFT Coulomb Fit - NWChem

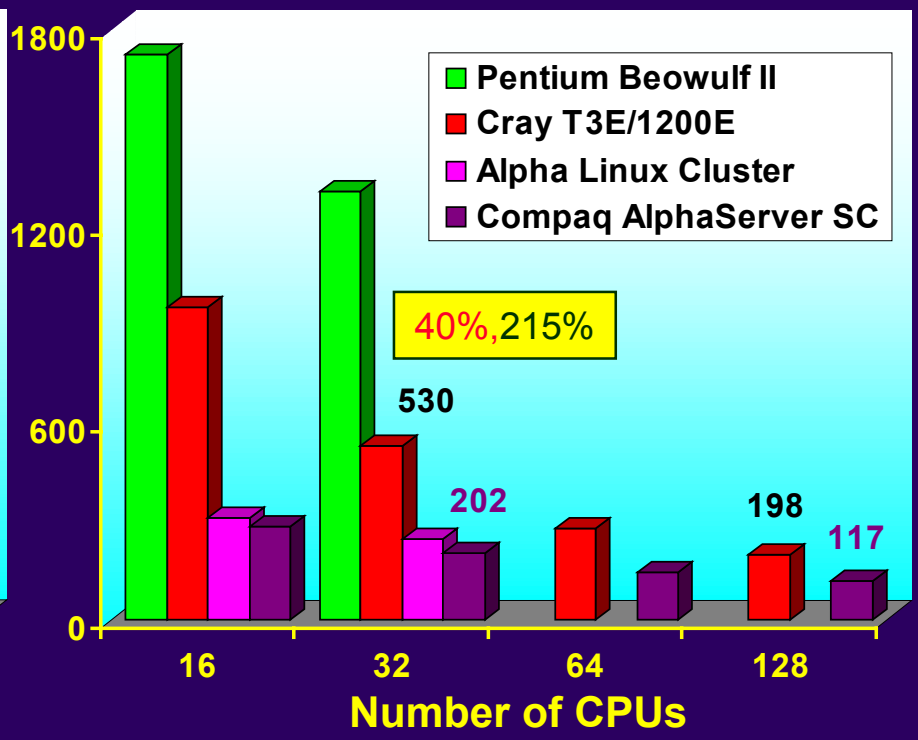


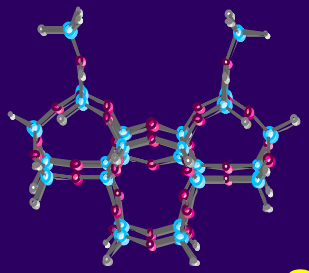


$Si_8O_7H_{18}$ 347/832

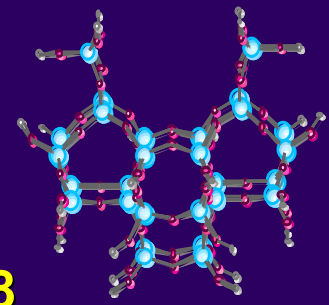


$Si_8O_{25}H_{18}$ 617/1444





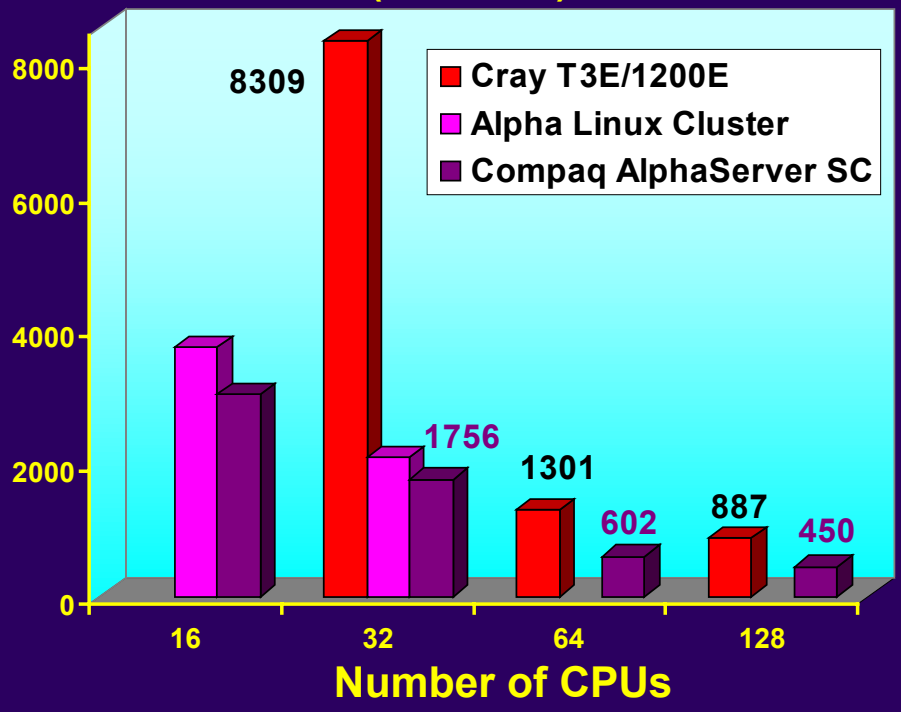
DFT Coulomb Fit - NWChem

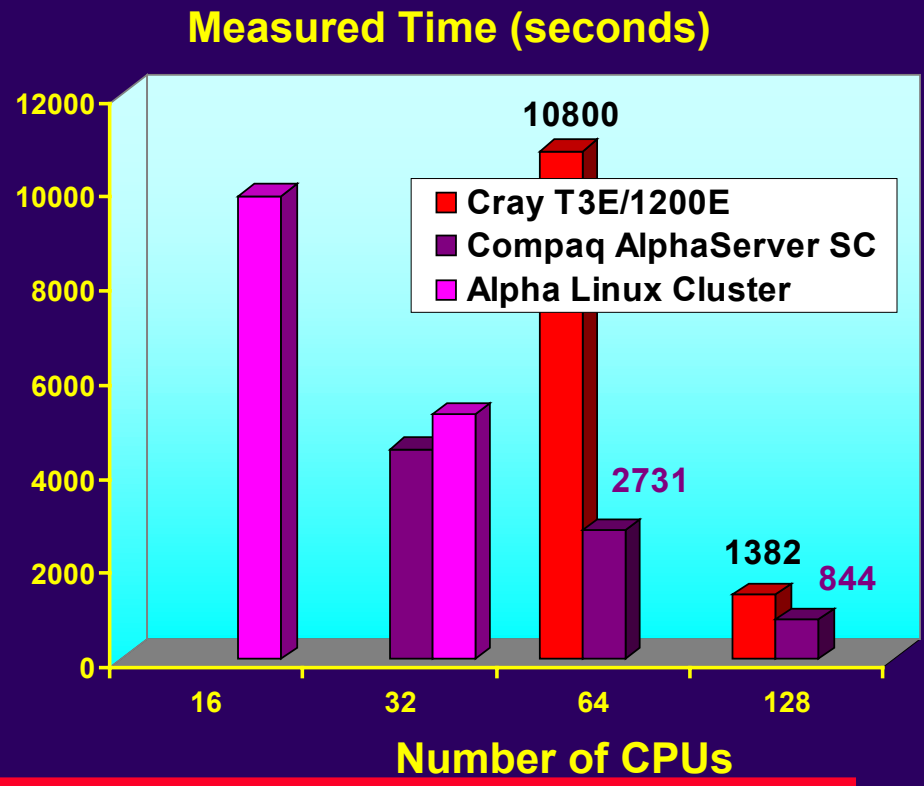


$Si_{26}O_{37}H_{36}$ 1199/2818 $Si_{28}O_{67}H_{30}$ 1687/3928

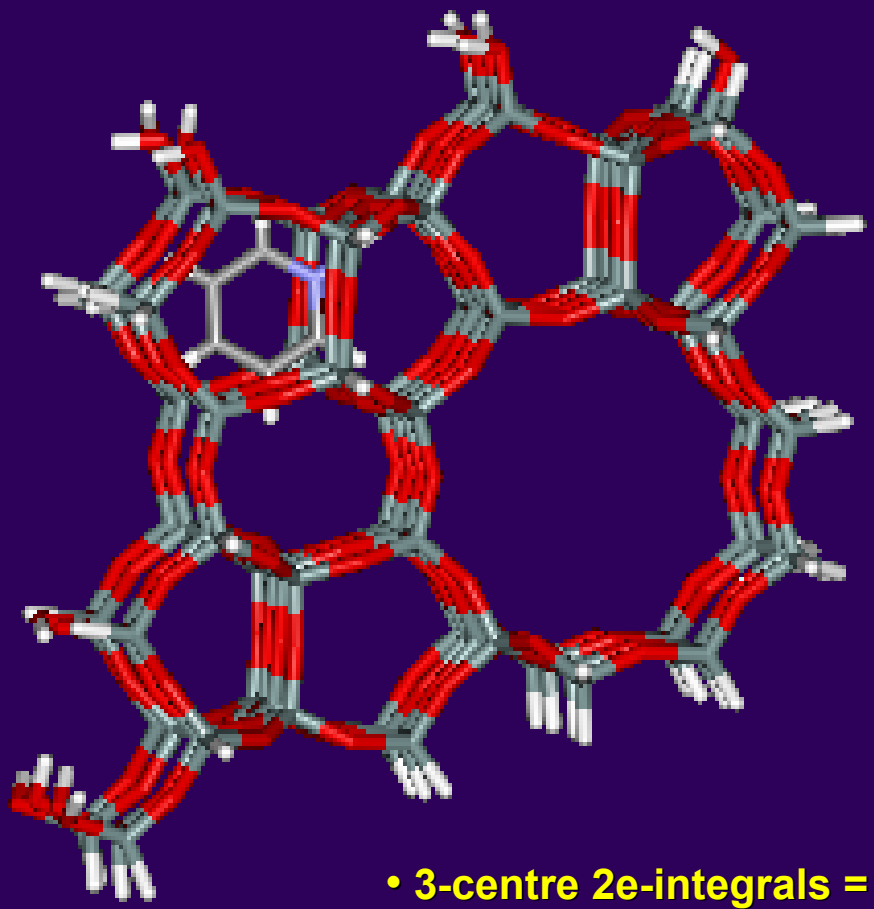
$T_{IBM\text{-}SP/P2SC\text{-}120}(256) = 1137$

$T_{IBM\text{-}SP/P2SC\text{-}120}(256) = 2766$





NWChem - DFT (LDA) Performance on the Compaq AlphaServer SC



Pyridine in Zeolite ZSM-5

- DZVP Basis (DZV_A2) and Dgauss A1_DFT Fitting basis:

  AO basis: 5457
  CD basis: 12713

- 256 EV67/6-667 CPUs (64 Compaq AlphaServer SC nodes)

Wall time (10 SCF iterations) on 256 CPUs = 11,960 seconds (60% efficiency)

- 3-centre 2e-integrals = 3.79×10^{12}
- Schwarz screening = 2.81×10^{11}
- % 3c 2e-ints. In core = 1.66%

GAMESS-UK

GAMESS-UK is the general purpose ab initio molecular electronic structure program for performing SCF-, MCSCF- and DFT-gradient calculations, together with a variety of techniques for post Hartree Fock calculations.

- The program is derived from the original GAMESS code, obtained from Michel Dupuis in 1981 (then at the NRCC), and has been extensively modified and enhanced over the past decade.
- This work has included contributions from numerous authors†, and has been conducted largely at the CCLRC Daresbury Laboratory, under the auspices of the UK's Collaborative Computational Project No. 1 (CCP1). Other major sources that have assisted in the on-going development and support of the program include various academic funding agencies in the Netherlands, and ICI plc.

Additional information on the code may be found from links at:

http://www.dl.ac.uk/CFS

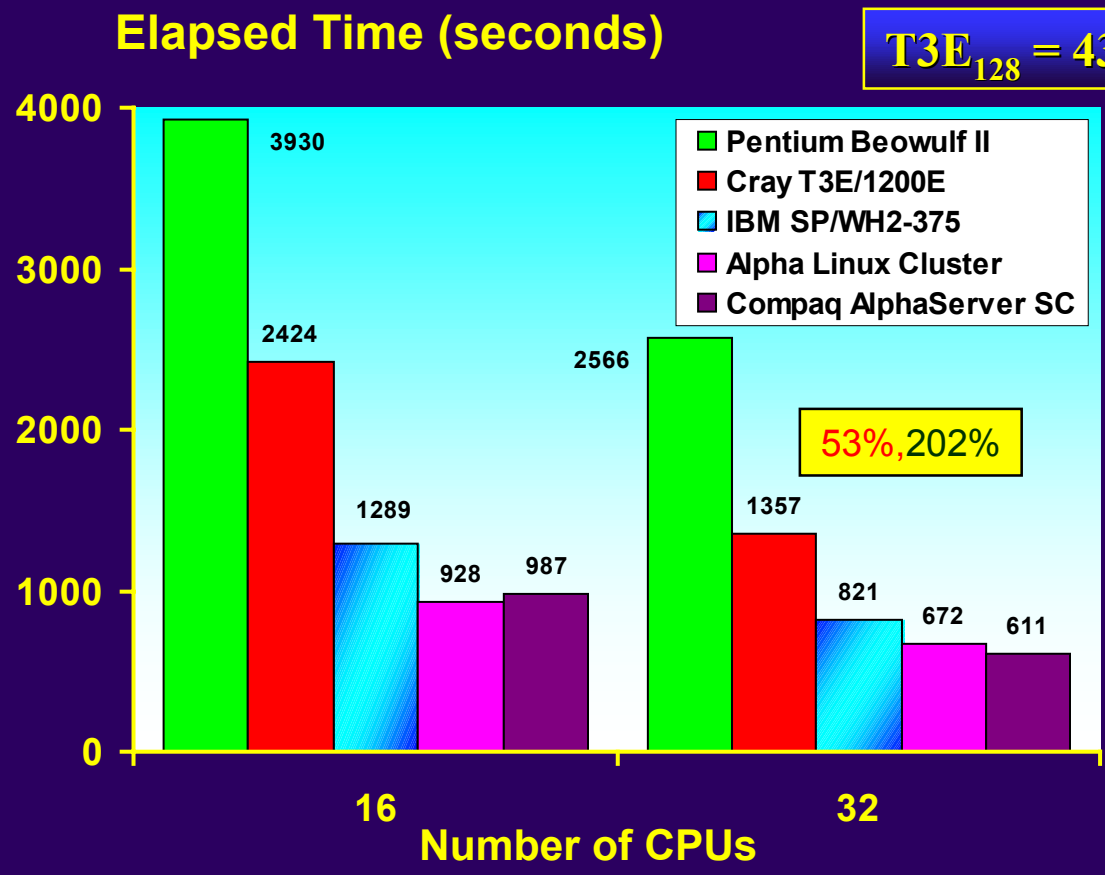
† *M.F. Guest, J.H. van Lenthe, J. Kendrick, K. Schoffel & P. Sherwood, with contributions from R.D. Amos, R.J. Buenker, H.H. van Dam, M. Dupuis, N.C. Handy, I.H. Hillier, P.J. Knowles, V. Bonacic-Koutecky, W. von Niessen, R.J. Harrison, A.P. Rendell, V.R. Saunders, A.J. Stone and D.Tozer.*

Parallel Implementations of GAMESS-UK

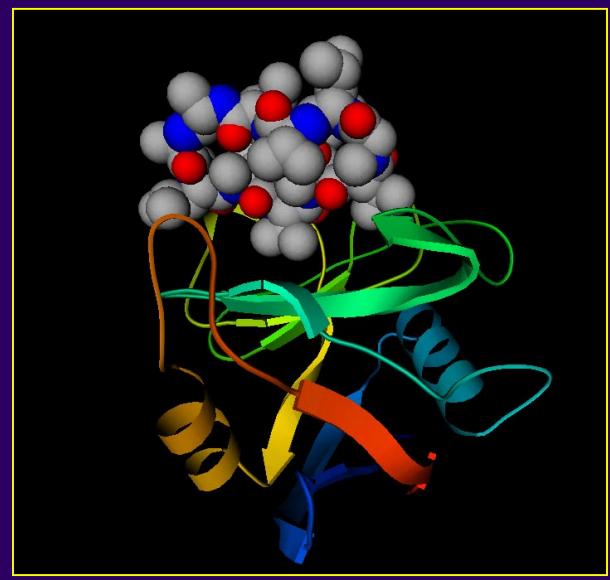
- Extensive use of Global Array (GA) Tools and Parallel Linear Algebra from NWChem Project (EMSL)
- SCF and DFT energies and gradients
  - Replicated data, but …
  - GA Tools for caching of I/O for restart and checkpoint files
  - Storage of 3-centre 2-e integrals in DFT Jfit
  - Linear Algebra (via PeIGs, DIIS/MMOs, Inversion of 2c-2e matrix)
- SCF second derivatives
  - Distribution of <vvoo> and <vovo> integrals via GAs
- MP2 gradients
  - Distribution of <vvoo> and <vovo> integrals via GAs

GAMESS-UK ΔSCF Performance

Cray T3E/1200E, IBM SP/WH2-375, Compaq AlphaServer SC & Beowulf Systems



$T3E_{128} = 436$



Cyclosporin:(3-21G Basis, 1000 GTOS)

Impact of Serial Linear Algebra:

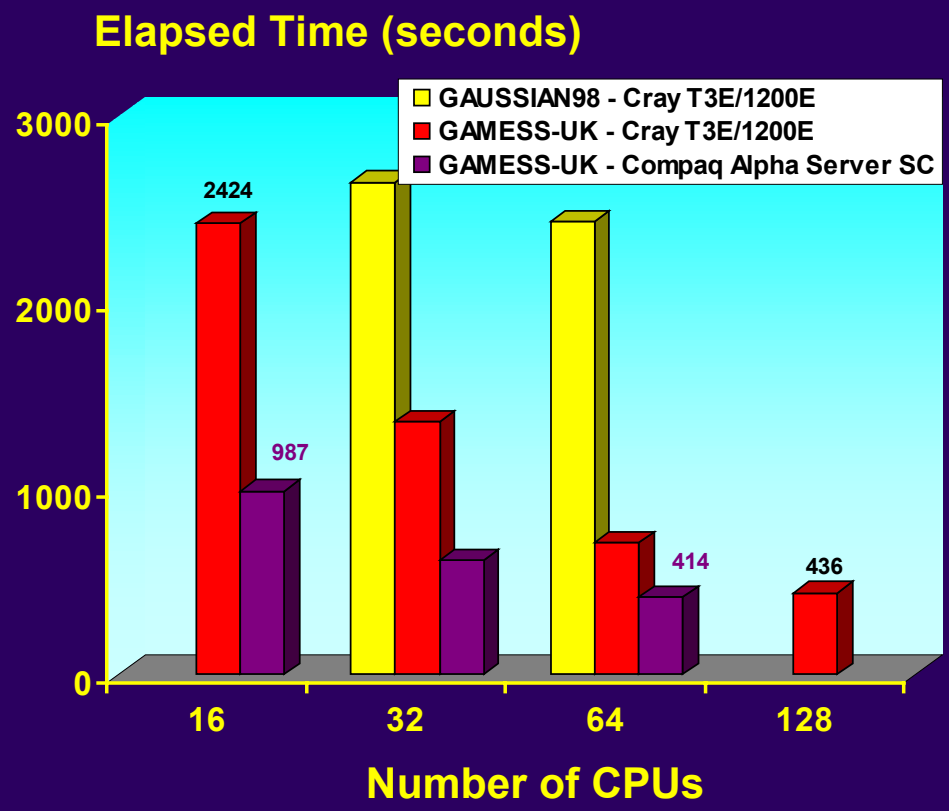
$T_{IBM-SP}(16) = 2656$ [1289]

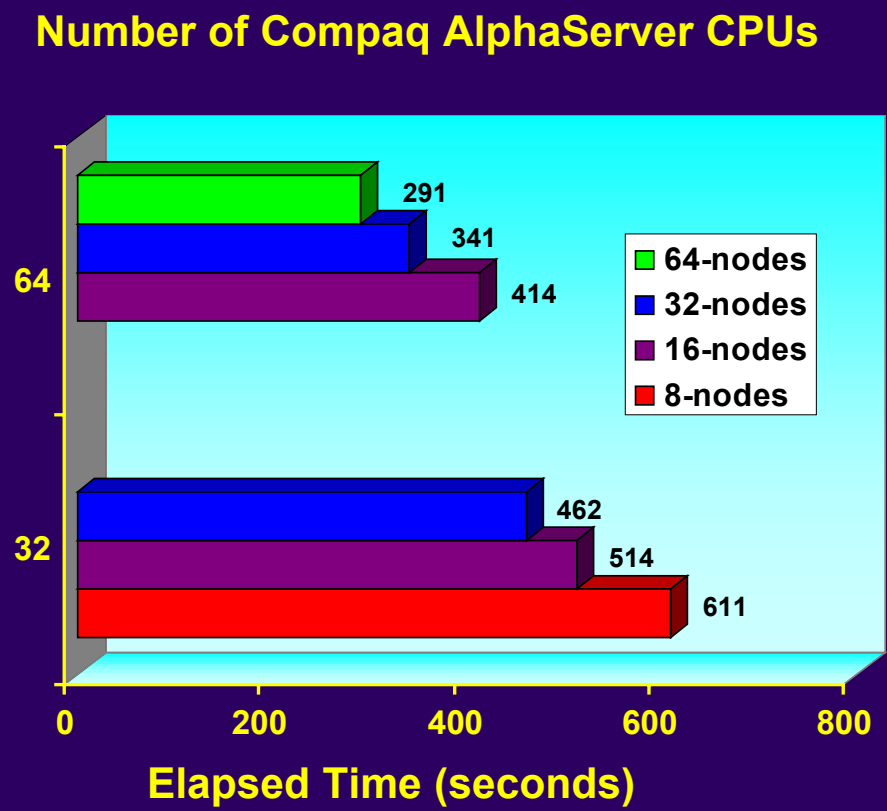
$T_{IBM-SP}(32) = 2184$ [821]

ΔSCF Performance - Cray T3E/1200E and Compaq AlphaServer SC

Cyclosporin:(3-21G Basis, 1000 GTOS)

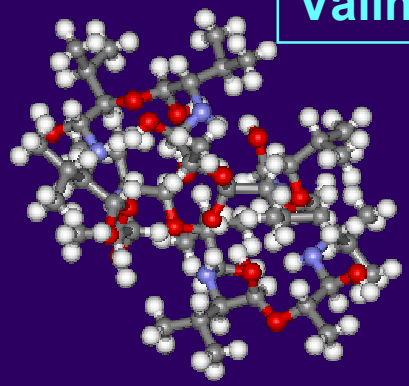
Gaussian 98
Serial: L302 - 90 secs;
L401 - 292 secs.
Serial linear algebra





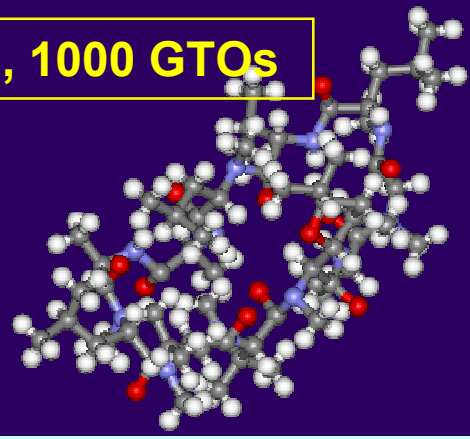
GAMESS-UK. DFT B3LYP Performance

Cray T3E, IBM SP, Compaq AlphaServer SC and Beowulf Systems



Valinomycin, 882 GTOs

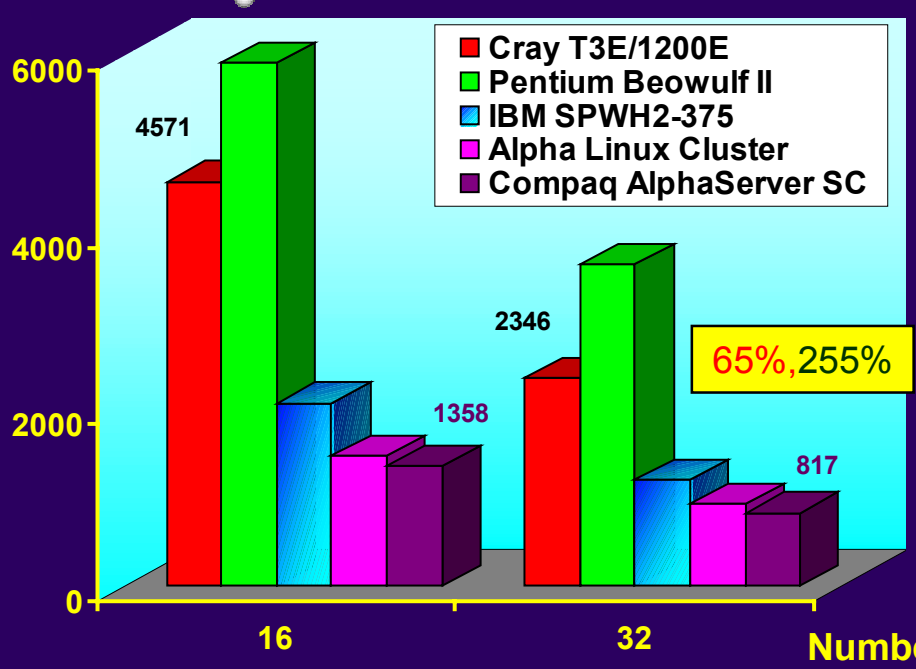
Cyclosporin, 1000 GTOs

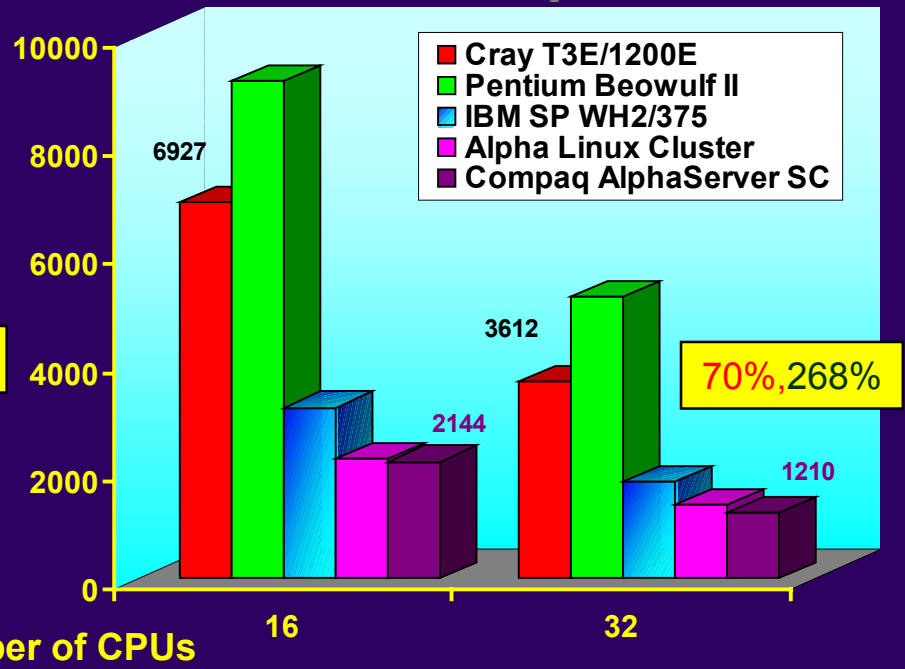


$J_{EXPLICIT}$

Basis: 6-31G

Elapsed Time (seconds)



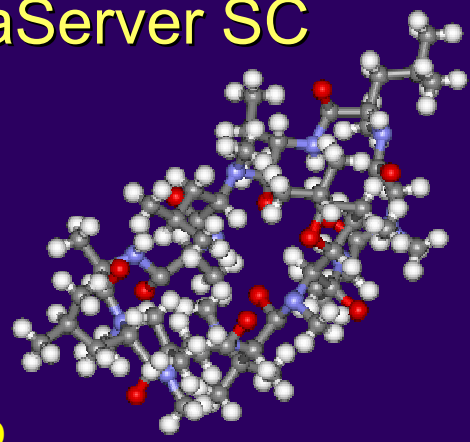


GAMESS-UK. DFT B3LYP Performance

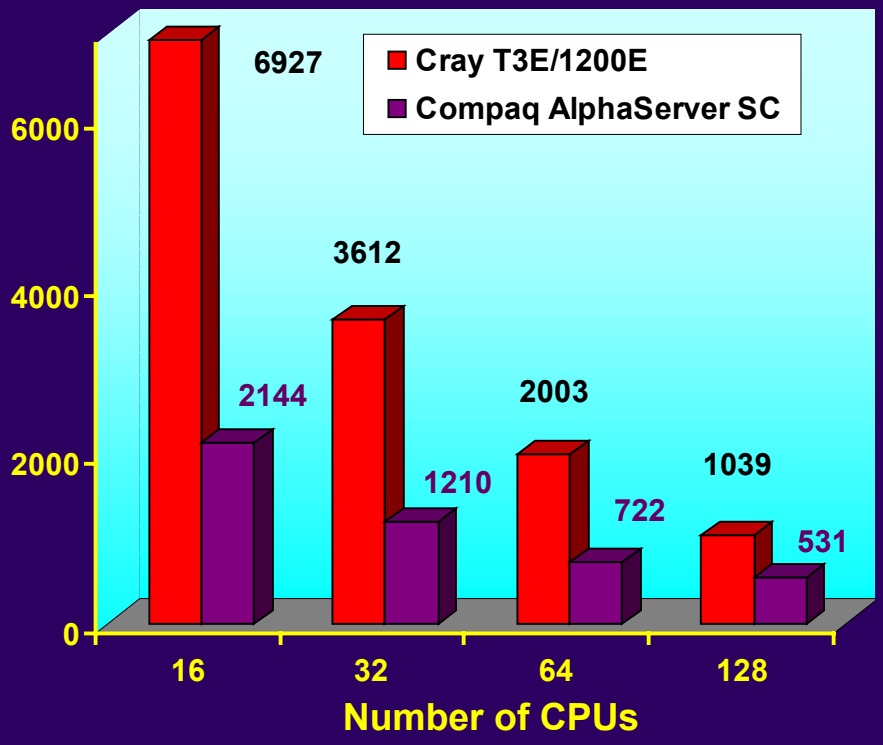
The Cray T3E/1200 and Compaq AlphaServer SC

Cyclosporin, 1000 GTOs

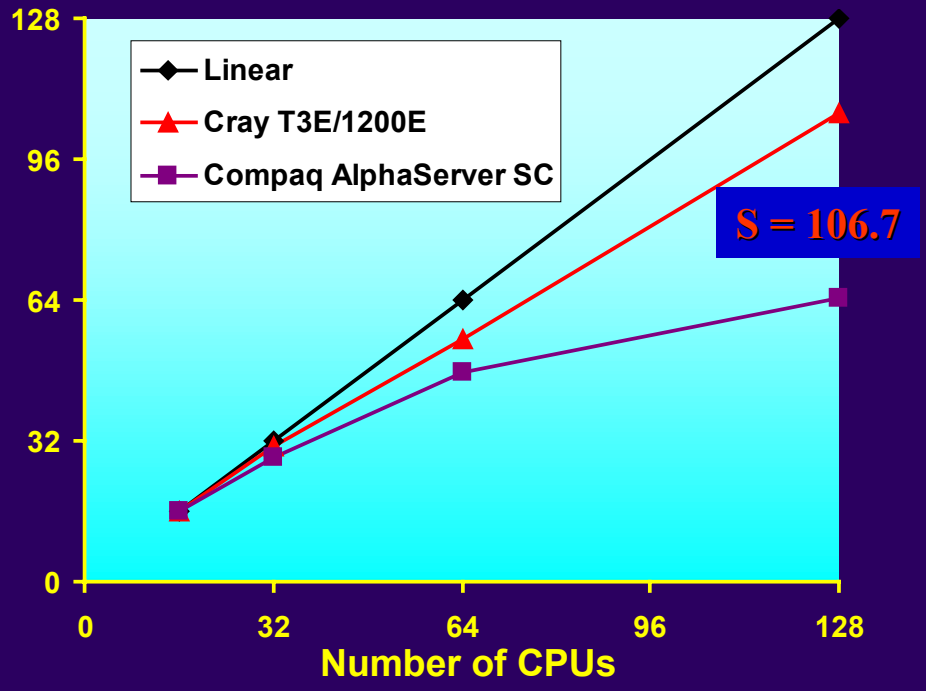
Basis: 6-31G



Elapsed Time (seconds)



Speed-up



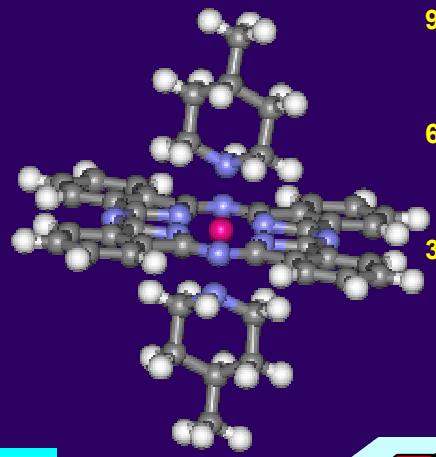
GAMESS-UK: - DFT B3LYP

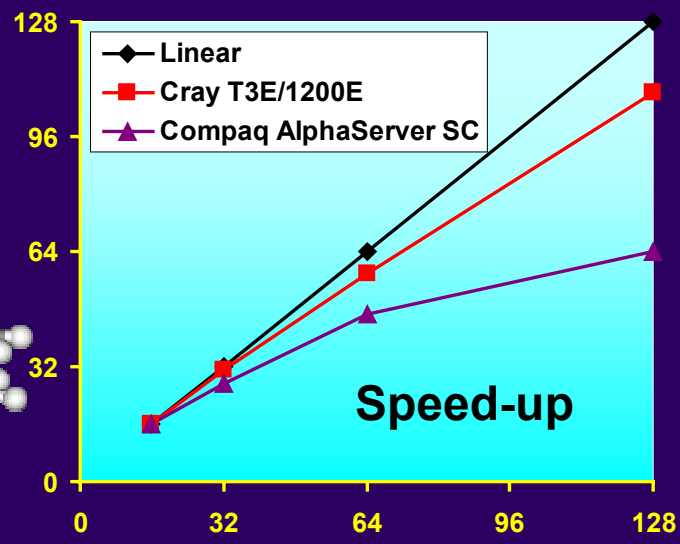
PcFe(4-Me.pip)2 (814 GTOs)

Basis

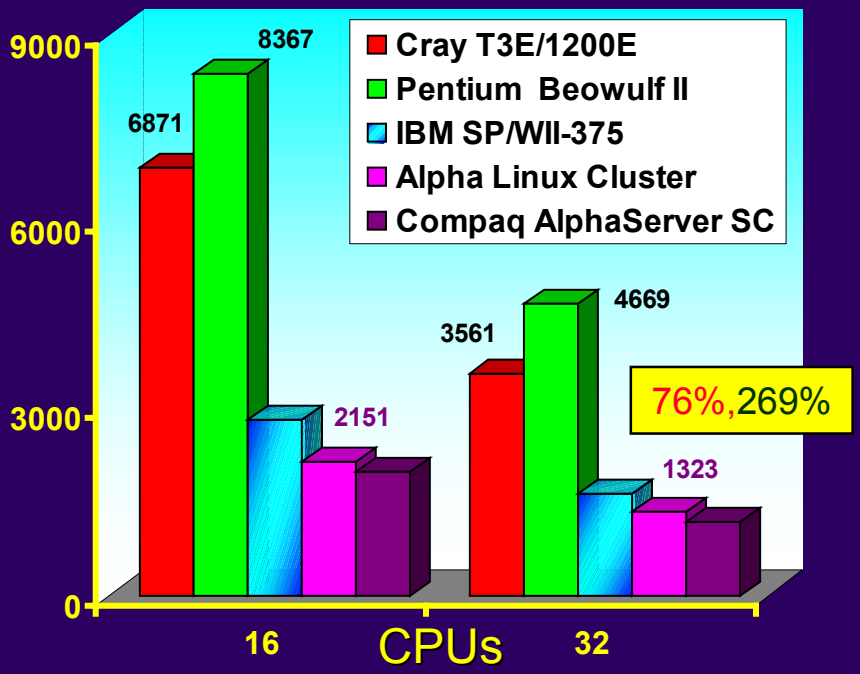
TZV Fe, TZVP (N), 6-31G*(C), 4-31G (H)

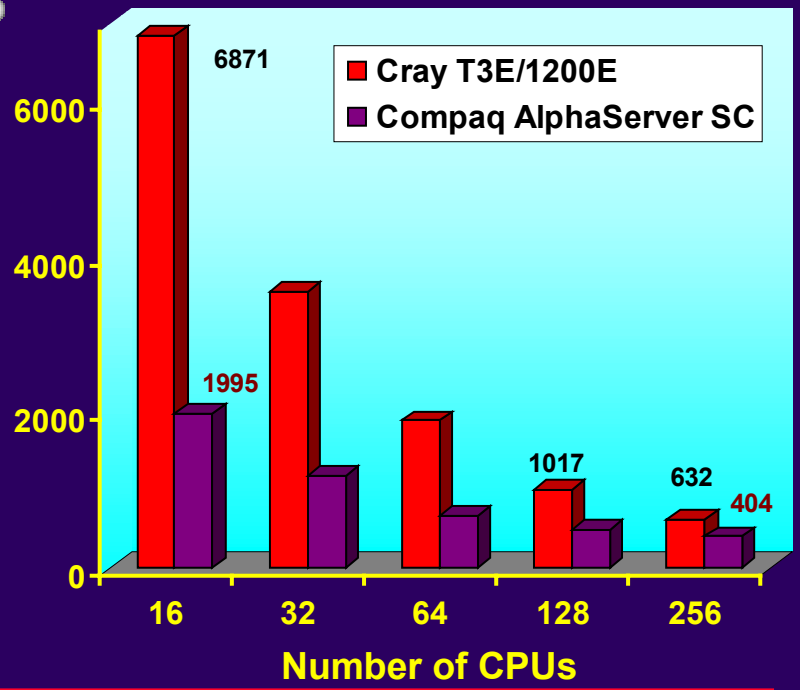
Compaq DS20/6-500 : 10.3 hours
SGI Origin 2000: 12 CPUS 4.3 hours

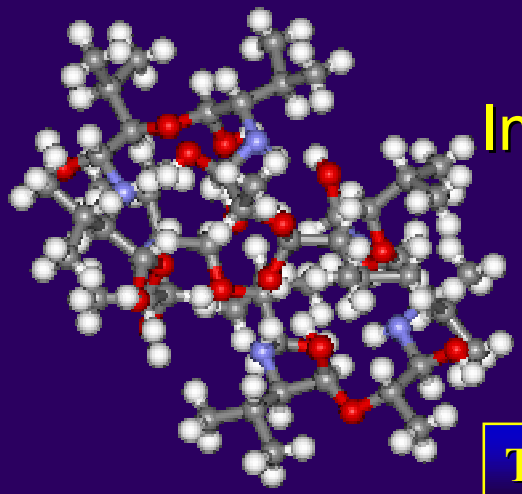




Elapsed Time (seconds)







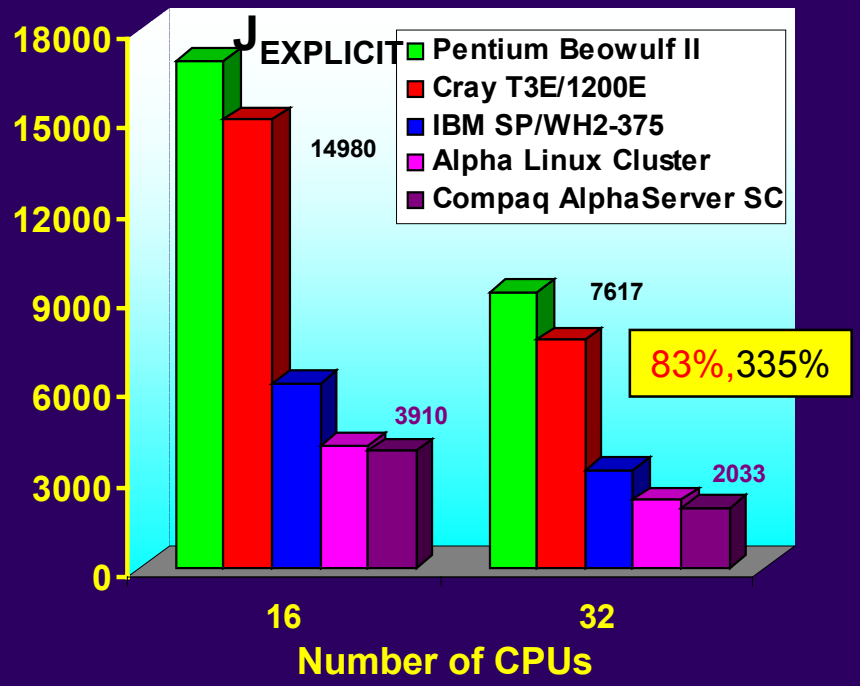
GAMESS-UK: DFT HCTH on Valinomycin.

Impact of Coulomb Fitting: Cray T3E/1200, IBM SP, Compaq AlphaServer SC & Beowulf Systems

Basis: DZV_A2 (Dgauss)
A1_DFT Fit: 882/3012

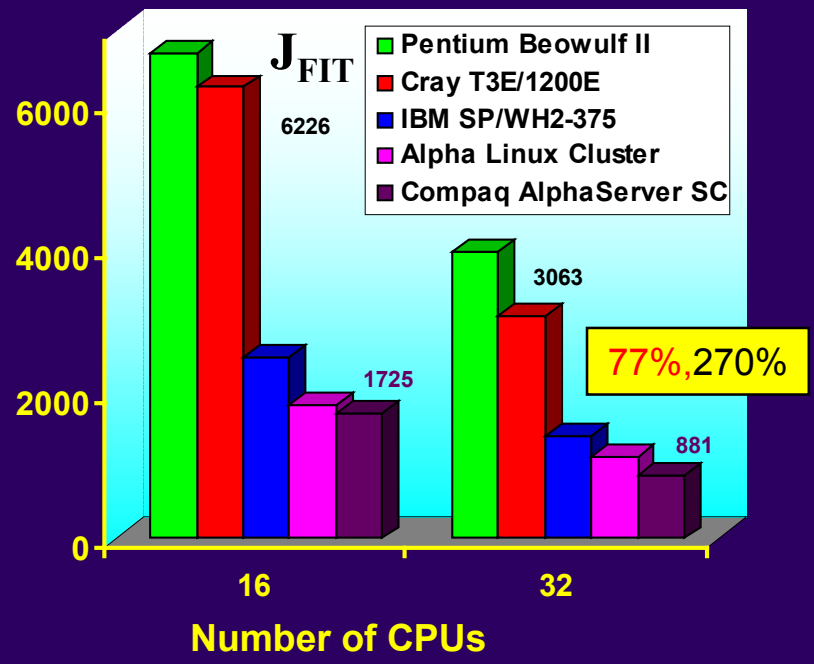
$T_{T3E/1200E}(128) = 2139$

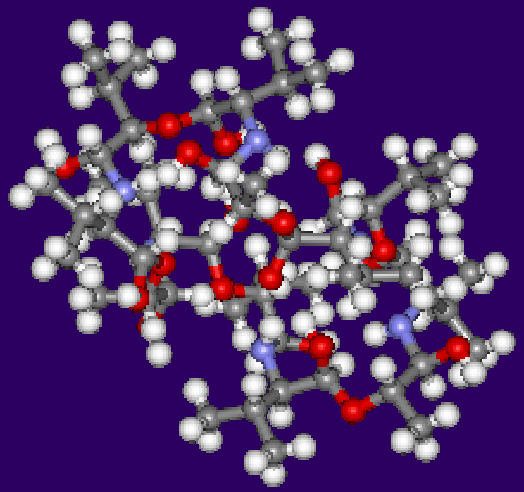
Measured Time (seconds)



$T_{T3E/1200E}(128) = 995$

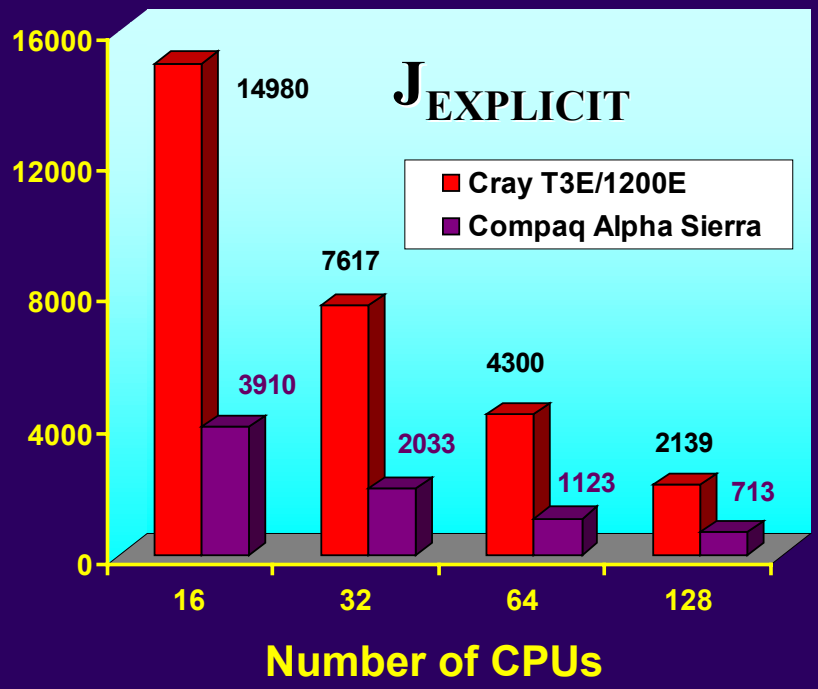
Measured Time (seconds)

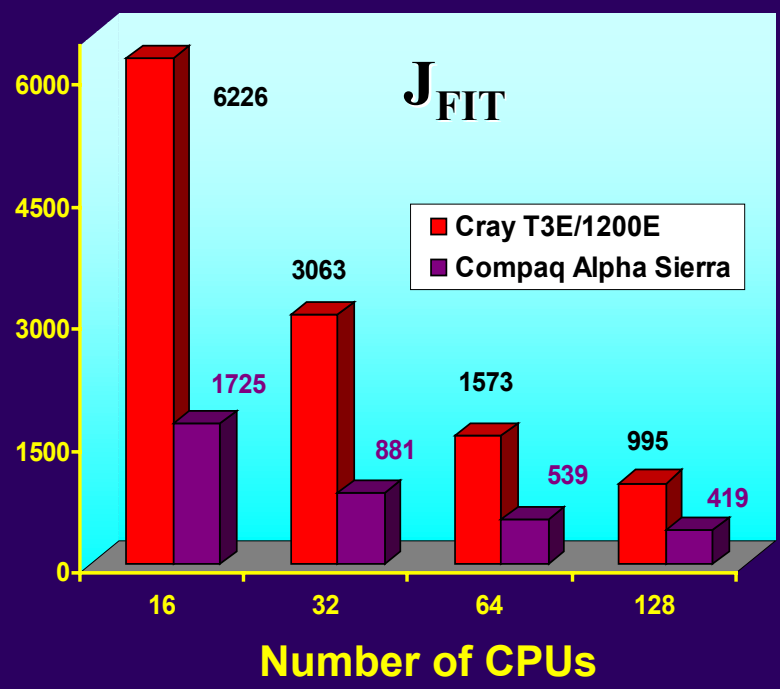




GAMESS-UK: DFT HCTH on Valinomycin. Impact of Coulomb Fitting: Cray T3E/1200, and Compaq AlphaServer SC

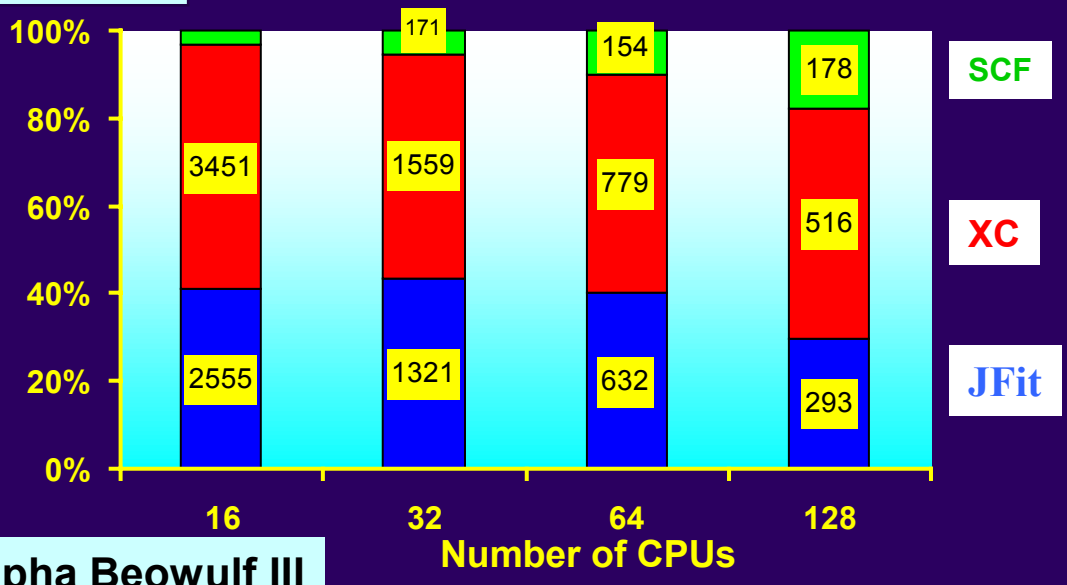
Basis: DZV_A2 (Dgauss)
A1_DFT Fit: 882/3012

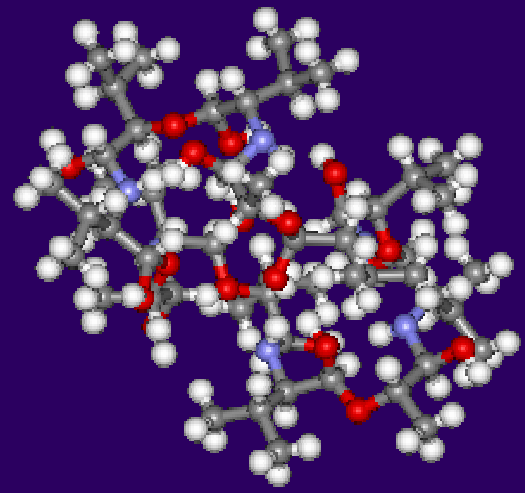




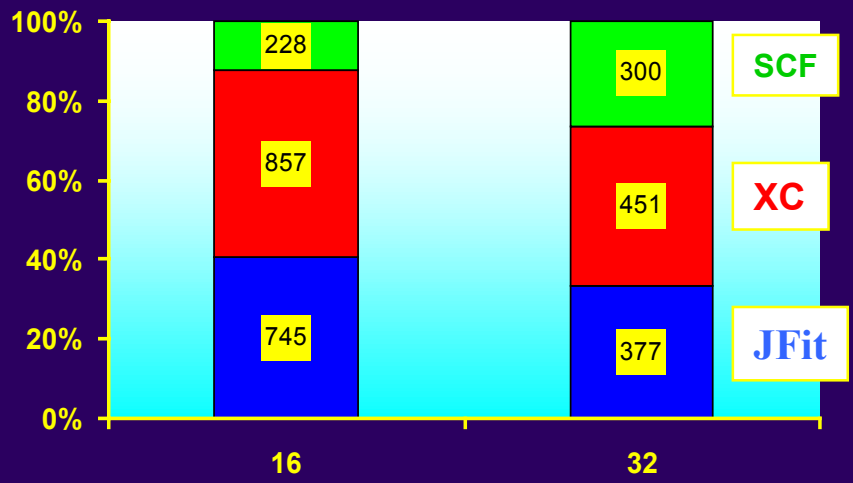
DFT HCTH Performance : Valinomycin

Cray T3E/1200E

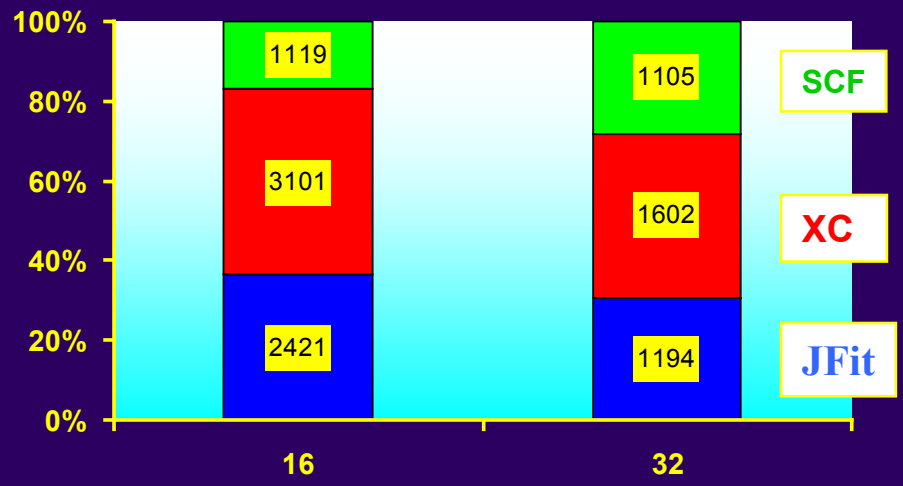




Alpha Beowulf III



Pentium Beowulf II

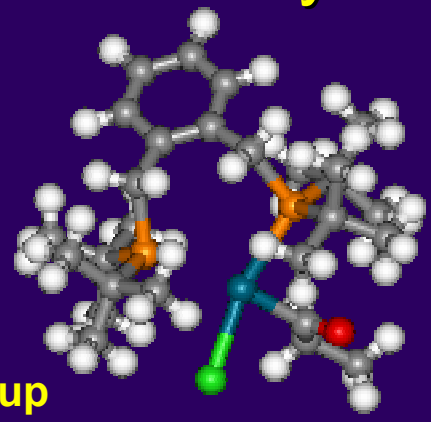


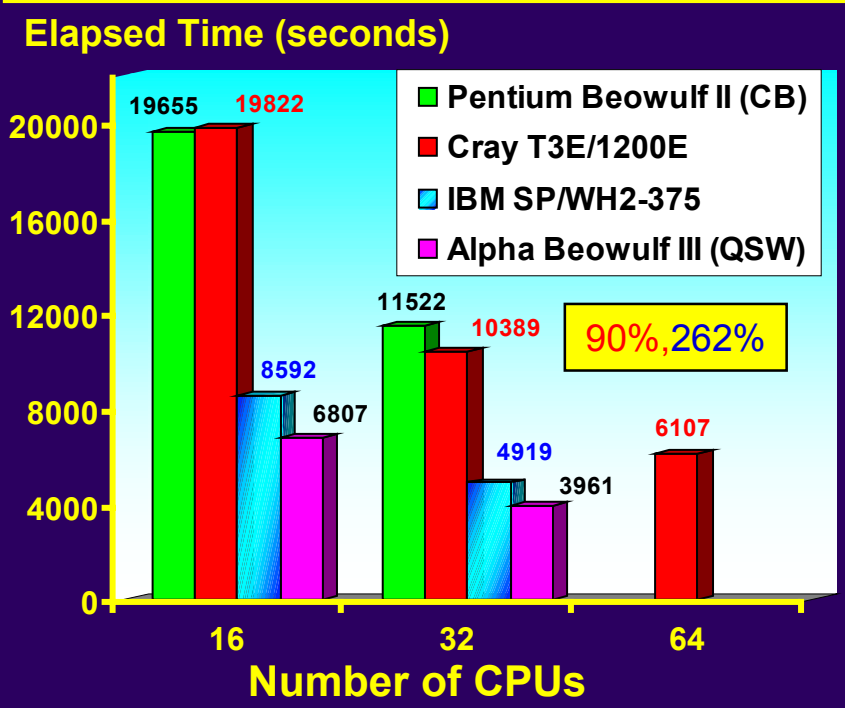
DFT BLYP Gradient:

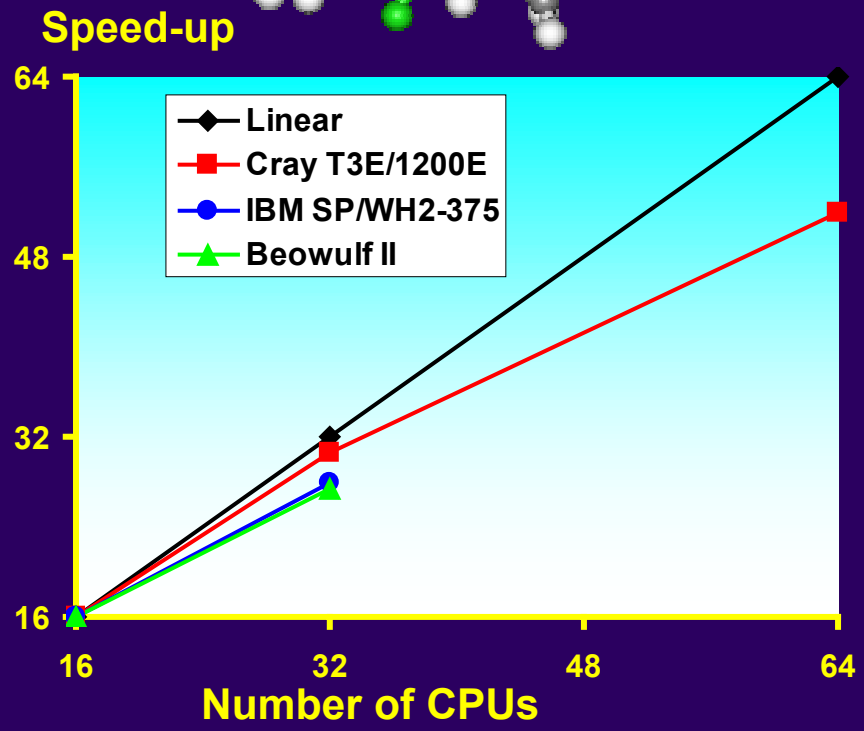
Cray T3E/1200, IBM SP/WH2 and Beowulf Systems

Geometry optimisation of polymerisation catalyst $Cl(C_3H_5O).Pd[(P(CMe_3)_2)_2.C_6H_4]$

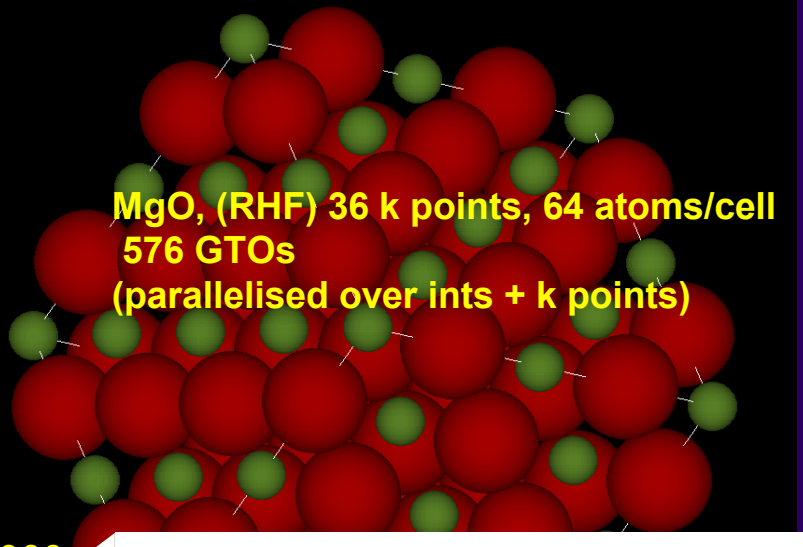
Basis 3-21G* (446 GTOs)
10 energy + gradient evaluations



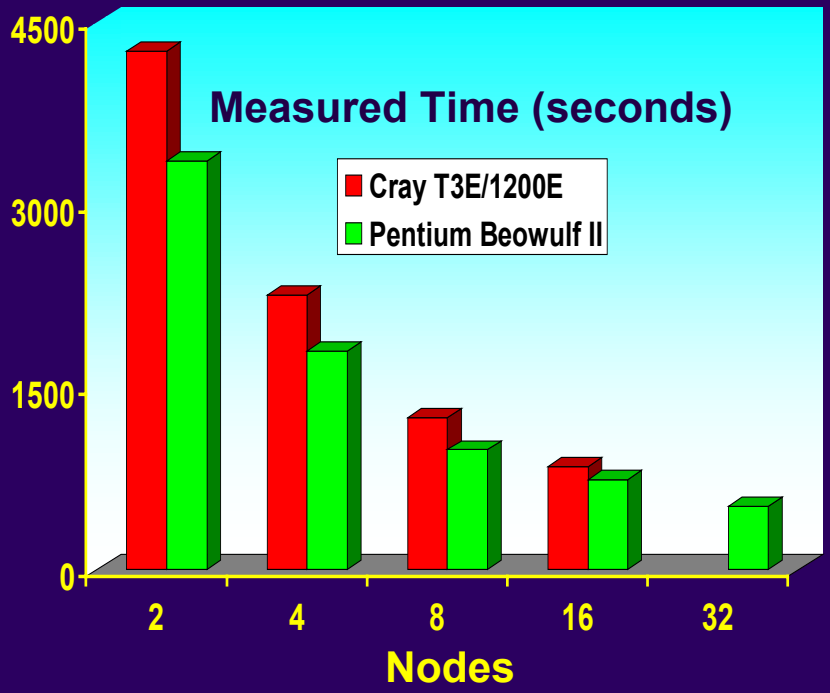


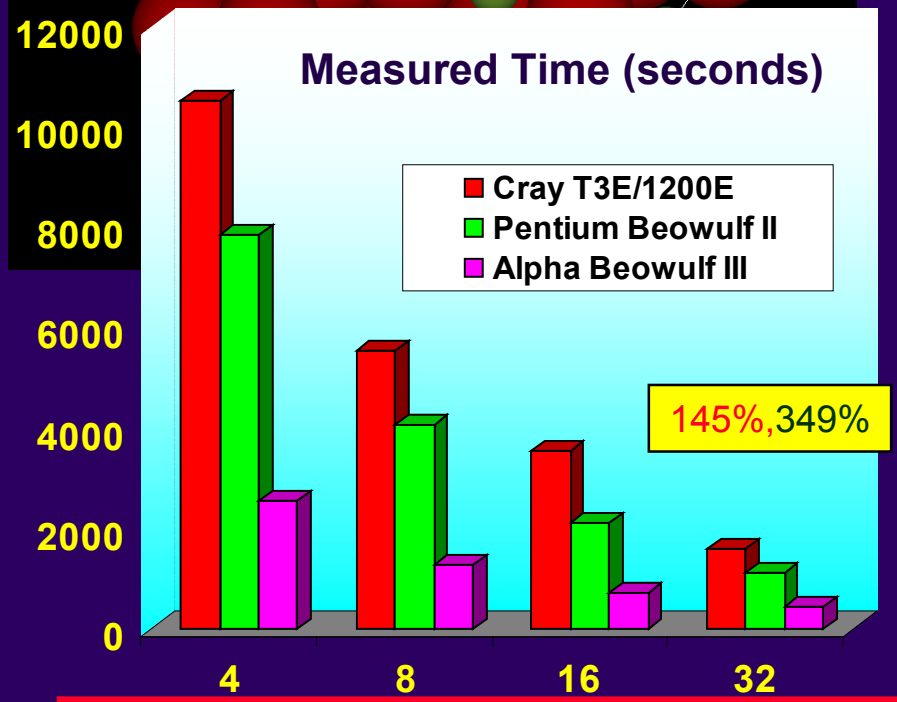


CRYSTAL98: Periodic SCF for MgO and TiO_2

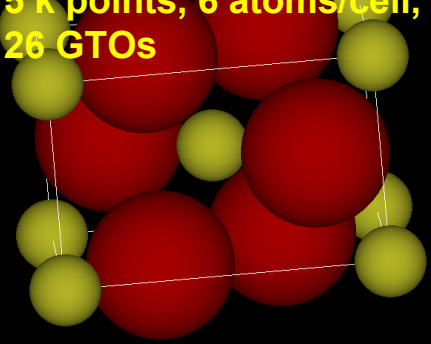


MgO, (RHF) 36 k points, 64 atoms/cell
576 GTOs
(parallelised over ints + k points)





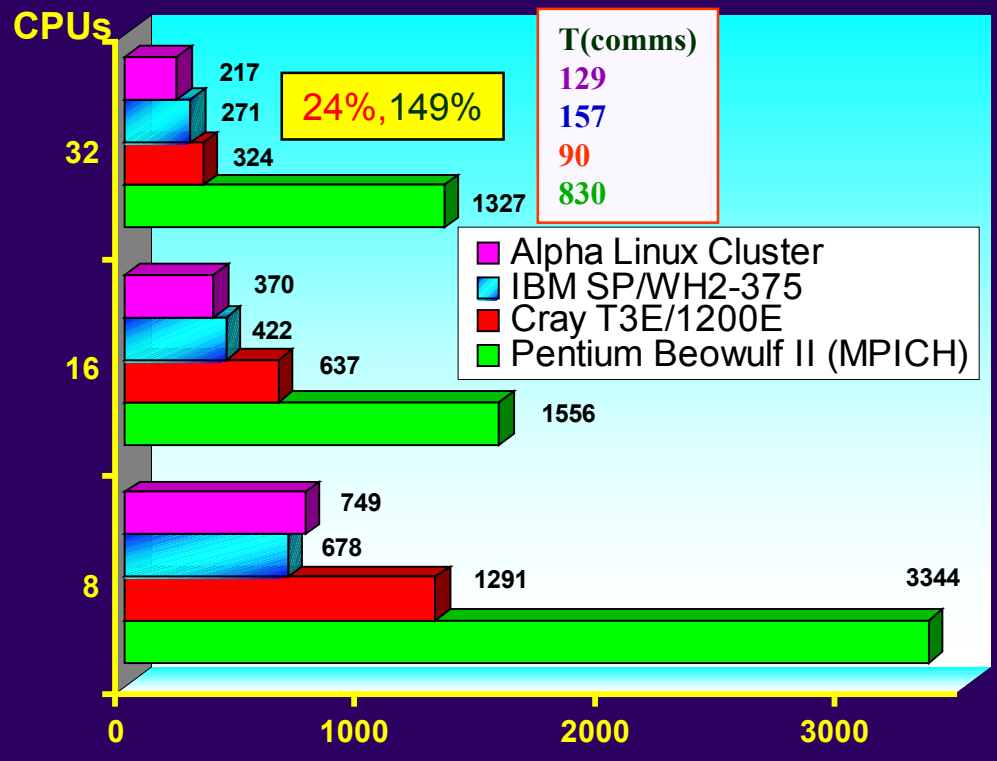
TiO_2 bulk crystal,
75 k points, 6 atoms/cell,
126 GTOs



CASTEP - The UK Car-Parrinello Consortium

UK Car-Parrinello Consortium

- The Cambridge Serial Total Energy Package CASTEP (M. Payne et al.) calculates the total energy, forces and stresses in a 3D-periodic system. (Rev. Mod.Phys. 64 (1992) 1045)
- DFT, plane-waves, pseudo-potentials & FFT's



CASTEP 4.2β Key Features

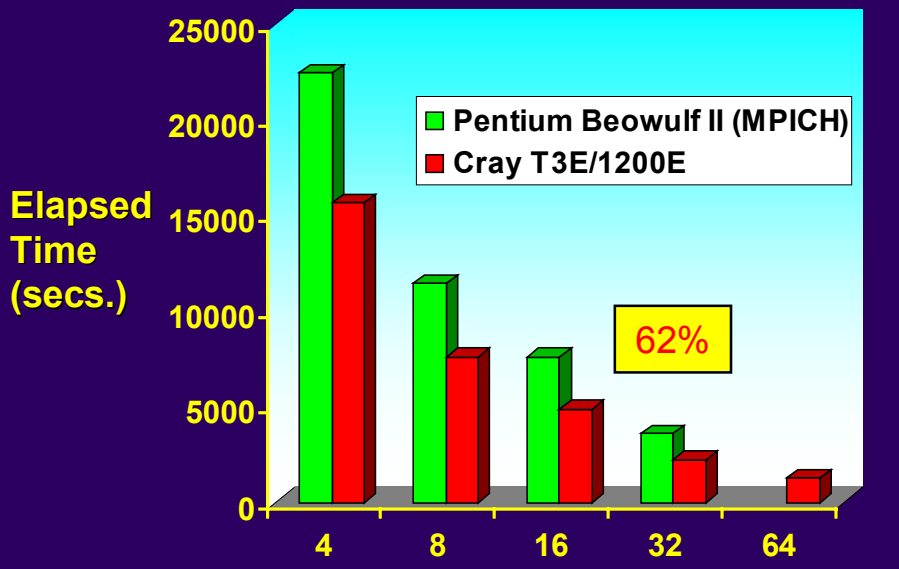
- Ultrasoft pseudo-potentials with non-linear core corrections
- Range of minimisation methods: Density Mixing, RM-DIIS, Conjugate Gradients band-by-band & all-bands. Full structural relaxation and MD
- LD and GGAs, spin-polarisation

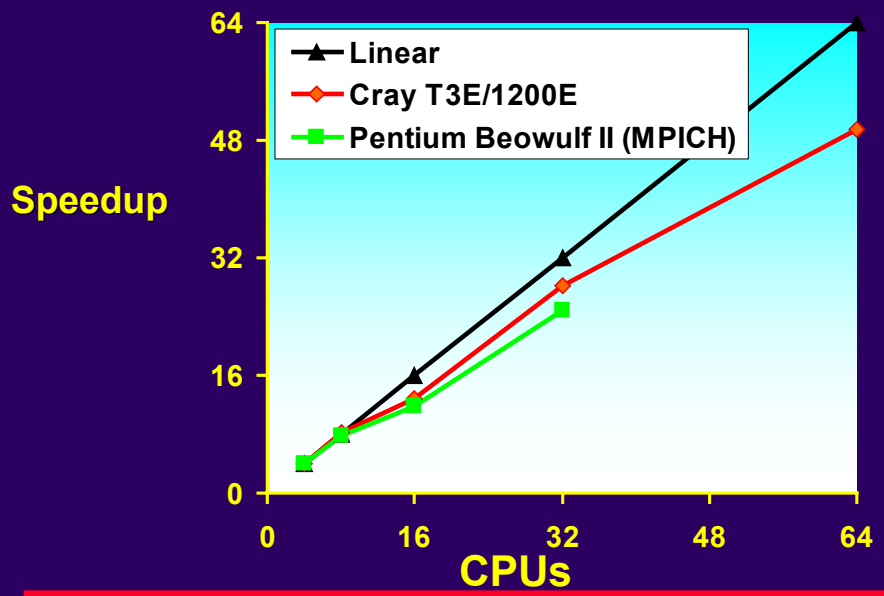
Benchmark Example

Chabazite

- Acid sites in a zeolite. (Si_{11} O_{24} Al H)
- Vanderbilt ulatrasoft pseudo-potential
- Pulay density mixing minimiser scheme
- single k point total energy, 96 bands
- 15045 plane waves on 3D FFT grid size = 54x54x54; convergence in 17 SCF cycles

CPMD - Car-Parrinello Molecular Dynamics





CPMD

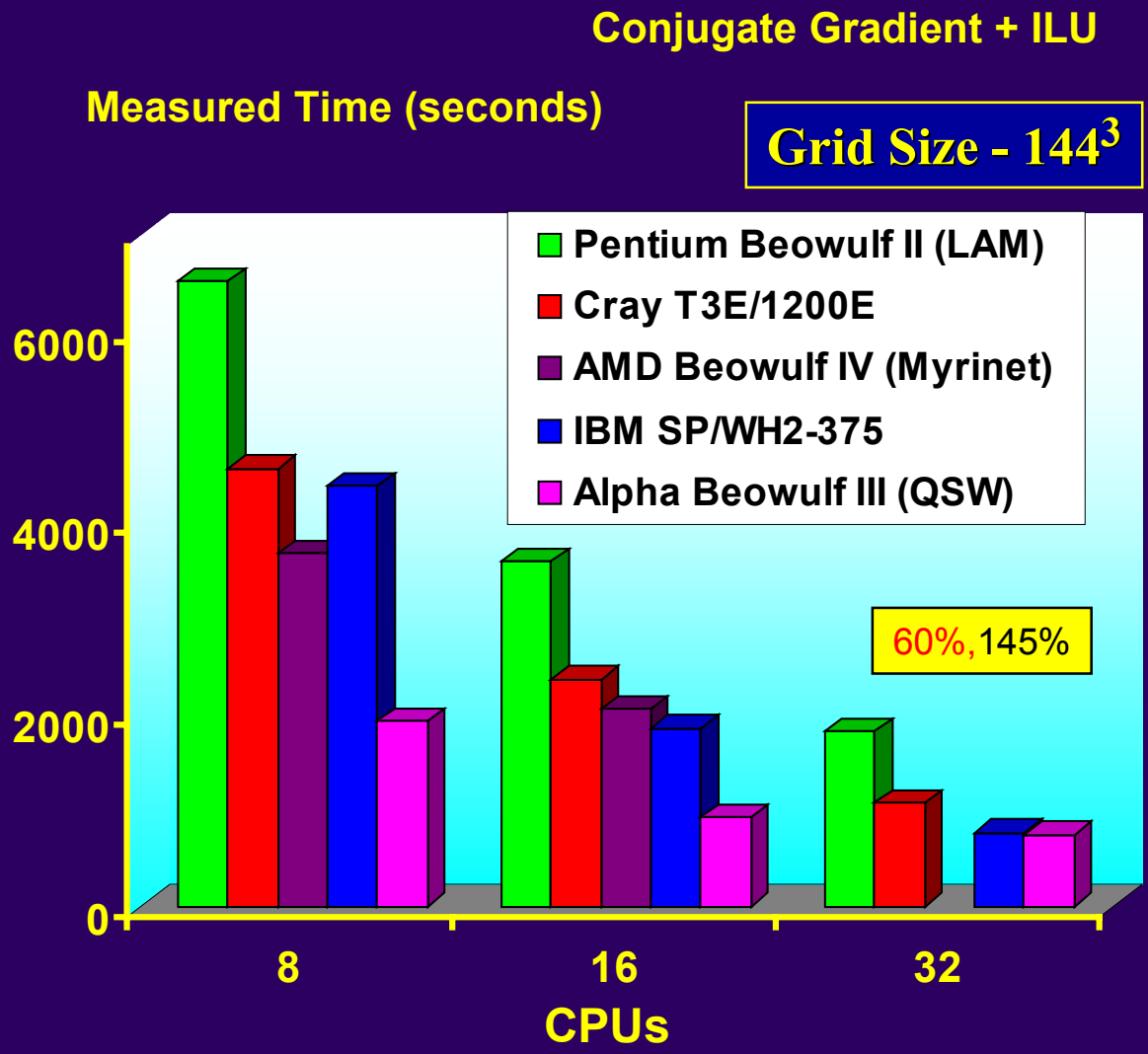
- Version 3.3: Hutter, Alavi, Deutsh, Bernasconi, St. Goedecker, Marx, Tuckerman and Parrinello (1995-1999)
- DFT, plane-waves, pseudo-potentials and FFT's

Benchmark Example: Liquid Water

- Physical Specifications:
  32 molecules, Simple cubic periodic box of length 9.86 A, Temperature 300K
- MD parameters;
  Time step 7 au = 0.169 fs; Length test run 200 steps = 34 fs
- Electronic Structure;
  BLYP functional, Trouillier Martins pseudopotential, Recriprocal space cutoff 70 Ry = 952 eV

Sprik and Vuilleumier (Cambridge)

ANGUS: Combustion modelling (regular grid)
The Cray T3E/1200, IBM SP/WH2 and Beowulf Systems



Direct numerical simulations (DNS) of turbulent pre-mixed combustion solving the augmented Navier-Stokes equations for fluid flow.

Discretisation of equations is performed using standard 2nd order central differences on a 3D-grid.

Pressure solver utilises either a conjugate gradient method with modified incomplete LU pre-conditioner or a multi-grid solver (both make extensive use of Level 1 BLAS) or fast Fourier transform.

ANGUS: Combustion modelling (regular grid)

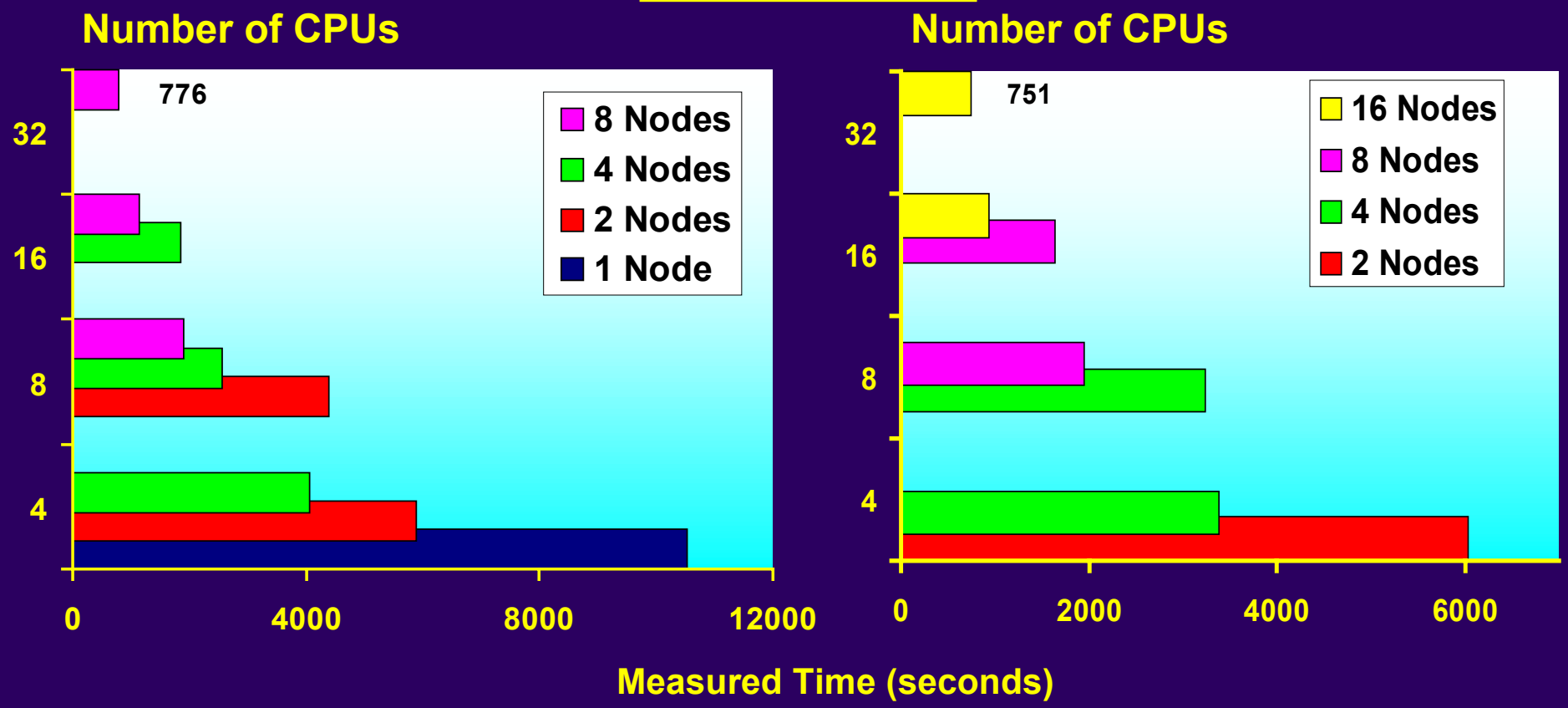
Memory Bandwidth Effects: The IBM SP and Alpha Beowulf Systems

Conjugate Gradient + ILU

IBM SP/WH2-375

Grid Size - 144^3

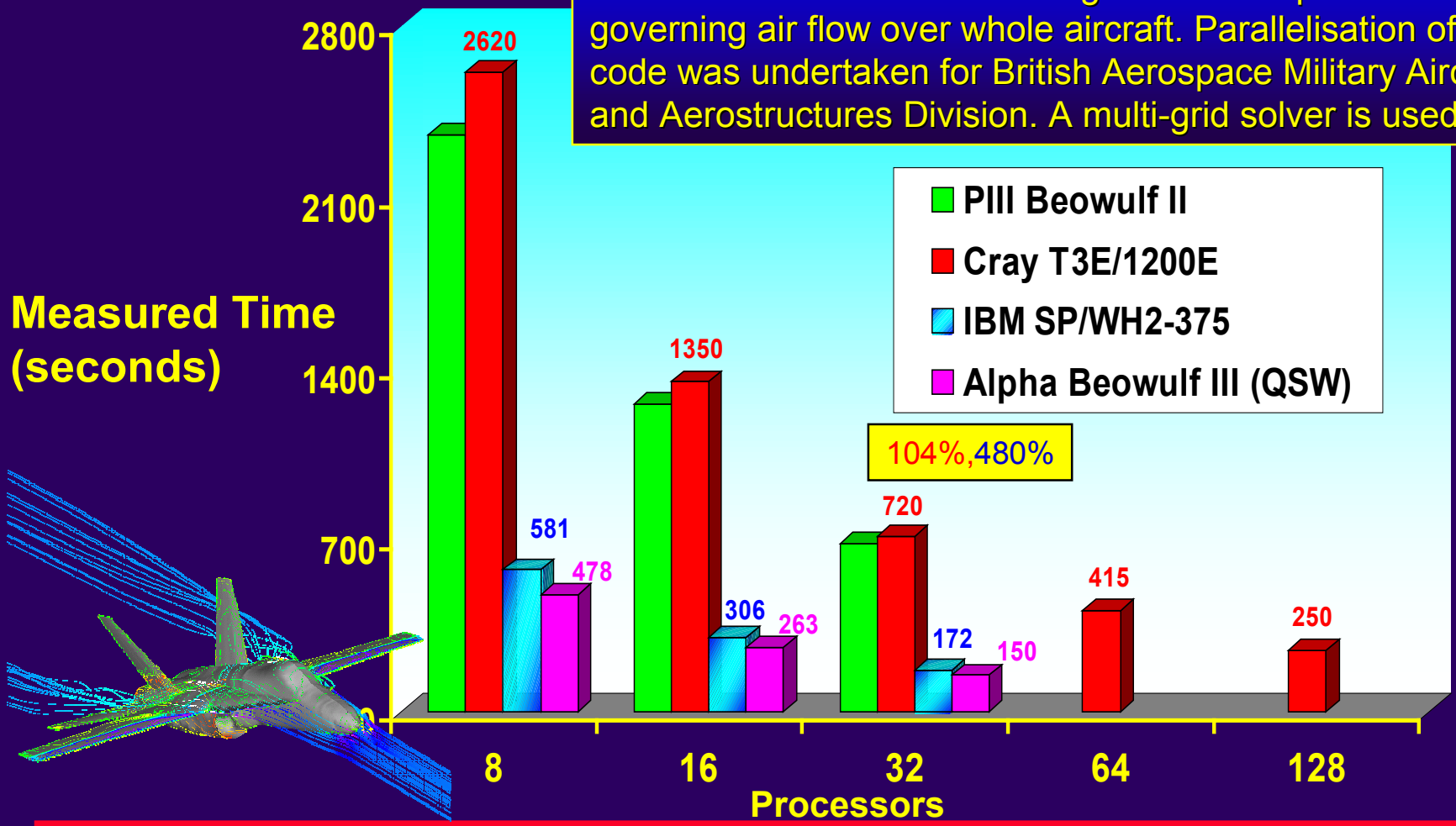
Alpha Beowulf III



FLITE3D: An Industrial Aerospace Code

2. F18 test, 3444350 elements

A finite-element code for solving the Euler equations governing air flow over whole aircraft. Parallelisation of the code was undertaken for British Aerospace Military Aircraft and Aerostructures Division. A multi-grid solver is used.

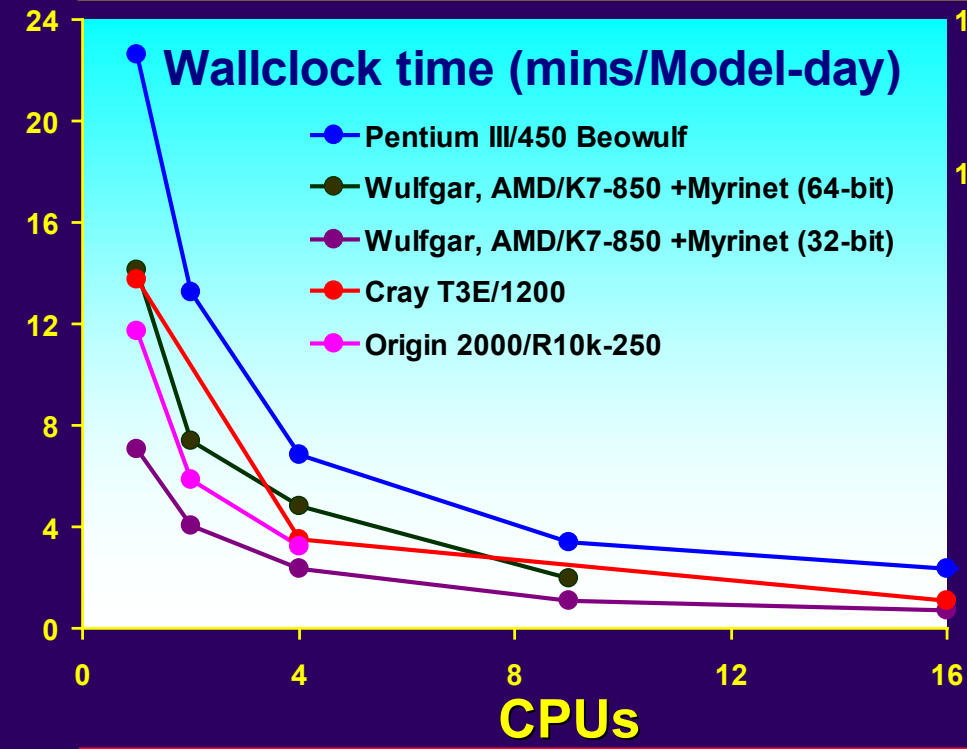


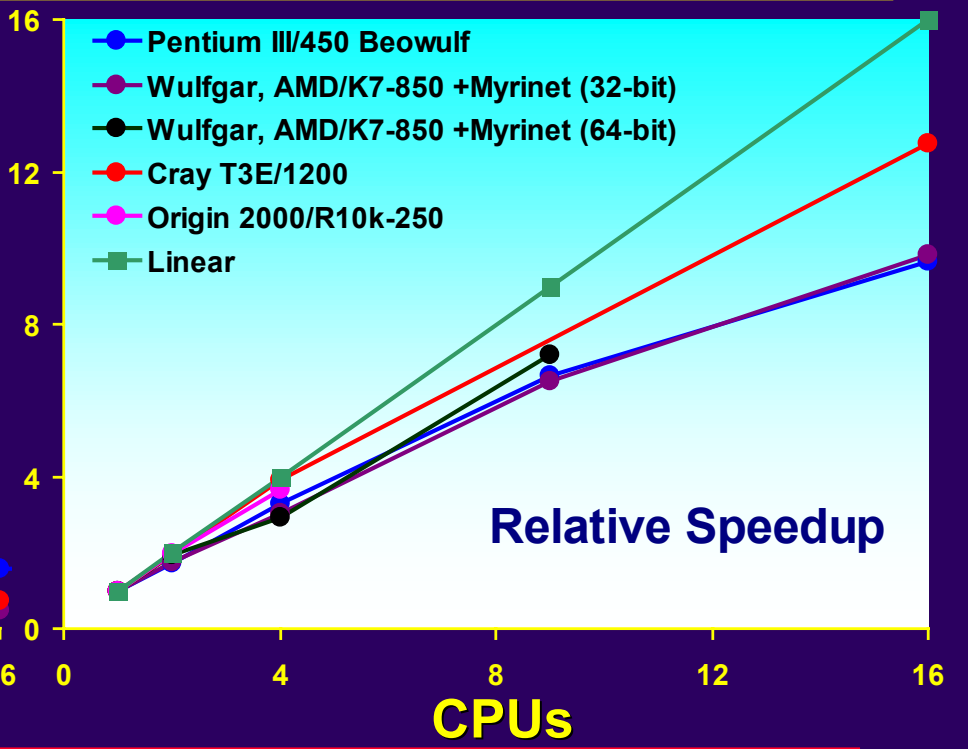
Atmospheric Climate Modelling

University of Reading and UK Meteorological Office

The Unified Model system consists of the full UKMO operational forecasting suite & the coupled Ocean-Atmosphere model used for climate prediction.

For a 5-day model period in atmosphere-only configuration, the T3E & Origin 2000 produce good speedups and the Beowulf systems scale well to 8 processors. This version of the model has many short messages and barriers & so is not particularly well coded.





Beowulf Comparisons with the Cray T3E/1200E

Pentium III Cluster [Beowulf II] % of 32-node Cray T3E/1200E	
GAMESS-UK	
SCF	53-69%
DFT	65-85%
DFT (Jfit)	43-77%
DFT Gradient	90%
MP2 Gradient	44%
SCF Forces	80%
NWChem (DFT Jfit)	40-51%
REALC	67%
CRYSTAL	145%
DLPOLY	
Ewald-based	95-107%
bond constraints	34-56%
CHARMM	96%
CASTEP	24%
CPMD	62%
ANGUS	60%
FLITE3D	104%

Alpha Linux Cluster [Beowulf III] % of 32-node Cray T3E/1200E	
GAMESS-UK	
SCF	202%
DFT †	255-326%
DFT (Jfit)	174-226%
DFT Gradient †	262%
MP2 Gradient	228%
SCF Forces	160%
NWChem (DFT Jfit) †	215-401%
CRYSTAL †	349%
DLPOLY	
Ewald-based †	352-447%
bond constraints	143-260%
CHARMM †	318%
CASTEP	149%
ANGUS	145%
FLITE3D †	480%

UKHEC

Summary

- Background - Issues of cost-effectiveness:
  - Beowulf systems within UKHEC
- High-End Computational Chemistry codes
  - Distributed data structures and GAs -NWChem & GAMESS-UK; Parallel Linear Algebra (PeIGS)
- Application Performance of 32 node Pentium Beowulf II and Alpha Linux Cluster
  - Comparison with CSAR Cray T3E/1200E, IBM SP/WH2-375 and Compaq AlphaServer SC
  - Pentium Beowulf delivers >50% T3E/1200E
  - Linux Alpha Cluster delivers between 150-400% of T3E/1200E, 80% of Compaq AlphaServer SC
- Materials **& Engineering**
  - CRYSTAL, CASTEP, CPMD
  - **ANGUS and FLITE3D**
- Limited by CPU counts (access to 128 CPUs)

PIII Beowulf II % 32-node Cray T3E	
GAMESS-UK	
SCF	53-69%
DFT	65-85%
DFT (Jfit)	43-77%
DFT Gradient	90%
MP2 Gradient	44%
SCF Forces	80%
NWChem	
DFT (Jfit)	40-51%
CRYSTAL	145%
DLPOLY	
Ewald-based	100%
bond constraints	45%
CHARMM	96%
REALC	67%
CASTEP	24%
CPMD	62%
ANGUS	60%
FLITE3D	104%