High Performance Enterprise Computing

Hardware Design & Performance

Application Optimization Guide

Performance Evaluation

Lynn Lewis
Director WW FAE MSS
lynn.lewis@amd.com

AMD Opteron™ Agenda

AMD

- ✓ ***Architectural Comparisons***
  - ✓***Feed the Beast: data flow***
  - ✓***Fit-in, Stand-out: x86, x86-64***
  - ✓***KISS: Keep it simple***
- ✓ ***AMD Core Math Library***
- ✓***Performance Benchmarks***

© 2005 Copyright by DYNA*more* GmbH

AMD

Architectural Comparisons

Architectural Comparisons

AMD

Architecture Agenda
The 3 Caveats to Processor Performance

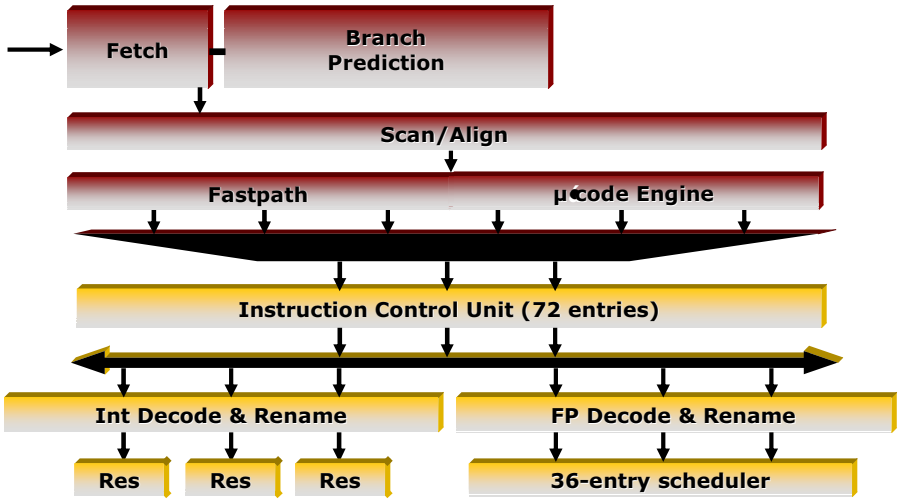
- **Memory Latency and SMP (*feed the beast*)**
  - Integrated IO controller decreases latency of memory access
  - Hyper-transport™ prevents latency increase in SMP configurations
  - Memory bandwidth/addressability increases with increase of CPUS
- **CISC → RISC Instruction Decoding**
  - Evolutionary 64-bit computing ***(kiss)***
  - Instruction compatibility on x86 architectures ***(fit-in)***
  - improved performance without recoding ***(fit-in, stand-out)***
- **RISC Engines dictate CPU Performance**
  - Choice of instruction set dictates number of RISC operations
  - Design of engine dictates throughput
  - Compiler dependency

© 2005 Copyright by DYNA*more* GmbH

Architectural Comparisons

AMD

Customer Centric 64-bit Computing
x86 Architecture: Evolution rather than Revolution



Architectural Comparisons

AMD

Customer Centric 64-bit Computing
x86 Architecture: Evolution rather than Revolution

- **Evolutionary 64-bit Architecture (KISS)**
  - Prefix-byte allows reuse of x86 instructions and decoders
  - Re-leveraging of x86 compiler IT to 64-bit
  - Reliable compilers build reliable ports
- **Artificially Intelligent CPUS (stand-out)**
  - manage their resources rather than compiler managed
  - strong OOE lessens compiler reliance
  - IO controller reduces cache coherency maintenance
- **RISC design masked from compiler and user**
  - increase performance with future design without investment
  - legacy compatibility without emulation

Evolution (AMD64) not Revolution (IA64)

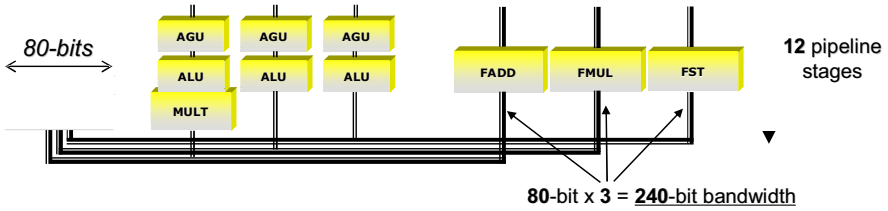
© 2005 Copyright by DYNA*more* GmbH

Architectural Comparisons

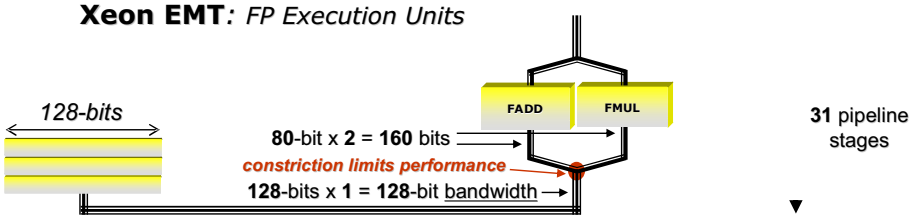
AMD

Customer Centric 64-bit Computing
x86 RISC Cores are not created equal

Opteron: *INT and FP Execution Units*



Xeon EMT: *FP Execution Units*



Architectural Comparisons

AMD

Customer Centric 64-bit Computing
x86 RISC Cores are not created equal

- **Number of pipes, FPU Design and RISC pipelining**
  - $^3/_2$**x** greater data throughput per clock
  - vectorization not required to achieve high levels of performance
  - high GHz without compromising instruction latency
- **Different widths**
  - CISC instructions decode into different number of RISC ops
  - instruction selection preferences different on x86 CPUS
- **Power Output and Pipelining**
  - length of pipeline necessitates high clock freq and increased voltage
  - power output does not scale linearly with voltage

CISC → RISC Decoding = compatibility, RISC design = power output (✓), throughput (✓), scalar (✓) and vector (✓)

© 2005 Copyright by DYNA*more* GmbH

Architectural Comparisons

AMD

Customer Centric 64-bit Computing
On die IO Controller: Feed the BEAST

- **On Die IO Controller at 2600 MHz (not 400 MHz)**
  - CPUs maintain cache coherency autonomously
  - faster clock - shorter clock cycles - less time waiting
  - dual-core maintains cache coherency without **HT**
- **Crossbar provide 4X more ports from each CPU**
  - more plumbing for dedicated IO purposes
  - up to ***20.8*** **GB/s** (**2S**) or ***28.8*** **GB/s** (**4S**) *per dual-core*
  - bandwidth / core does not diminish with # of sockets
- **Hyper-transport™ at 1000 GHz (not 400 MHz)**
  - SMP memory latency decreased over any North Bridge based solution

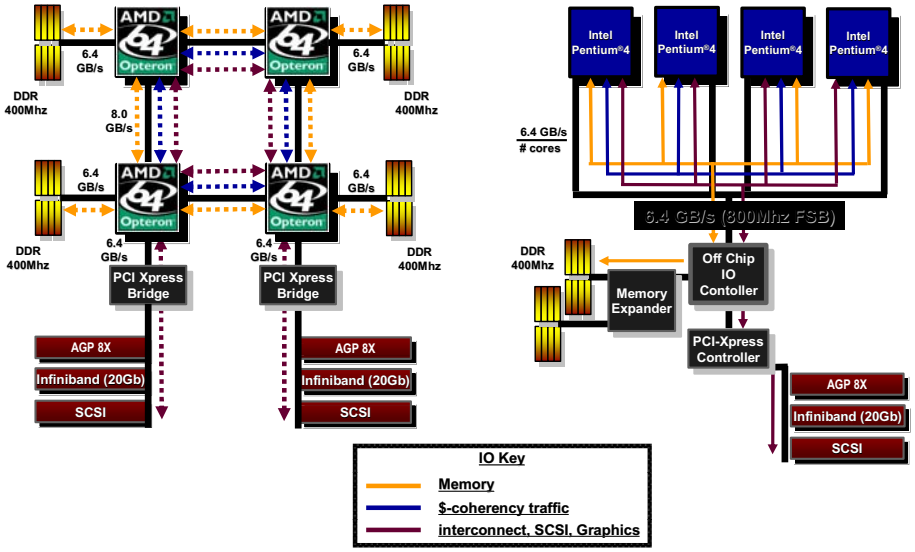
SMP/MPI and Workstations require IO for cache coherency, Interconnect, HDs and Graphics: What are the requirements?

September 28, 2005 | Computation Products Group | 9

Architectural Comparisons

AMD

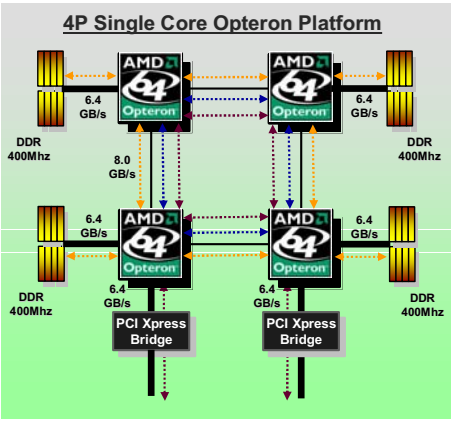
Customer Centric 64-bit Computing
IO Platform Comparison

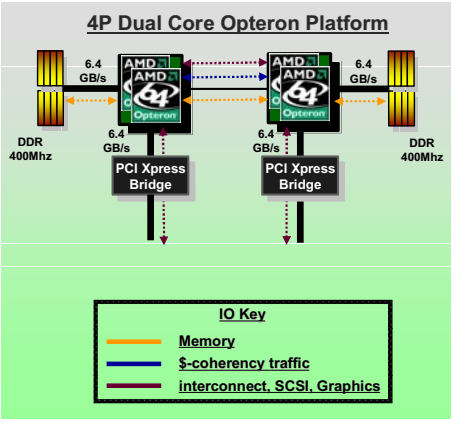


September 28, 2005 | Computation Products Group | 10

Architectural Comparisons

IO Advantages of AMD Dual-Core
Memory Latency





Dual Core Memory Latency is better than the Comparable Single Core Equivalent system

Architectural Comparisons

Customer Centric 64-bit Computing
IO Platform Comparison

- **Interconnect Bandwidth Trends**
  - Myrinet : ***500*** **MB/s**
  - Infiniband : ***1250*** **MB/s** (10Gb/s), ***2500*** **MB/s** (20Gb/s), ***5000*** **MB/s** (40Gb/s)
- **Graphics Card Bandwidth Trends**
  - AGP 8x : ***2000*** **MB/s**
  - PCI Express : ***6400*** **MB/s**
- **Dual Core Comparison**
  - On Opteron, bandwidth per core is fixed regardless of # of sockets
  - On competitor platforms, bandwidth per core degrades linearly

Only AMD OpteronTM Platforms provide scalable solutions to customers with enough IO to scale upon Dual Core

© 2005 Copyright by DYNA*more* GmbH

Architectural Comparisons

Scalable Memory Bandwidth and IO
Bandwidth/Latency Comparison

AMD

- **4 Separate IO Channels per CPU – *Scalable SMP Bandwidth***

Architecture	1P	2P	4P
Opteron	12.8 **GB/s**	41.6 **GB/s**	115.2 **GB/s**
XeonEMT	6.4 **GB/s**	6.4 **GB/s**	6.4 **GB/s**

- **Hyptertransport™ Interconnect – *low SMP memory latency***

Architecture	1P	2P	4P
Opteron	50 **ns**	75 **ns**	110 **ns**
XeonEMT	80 **ns**	~200 **ns**	> 200 **ns**

- **Commodity/High Performance SMP Solution**
  - presently dual core ready – SRQ controller has port for 2nd core
  - fewer # of chips required for MP chipsets – lowering cost of SMP systems

September 28, 2005 | Computation Products Group | 13

AMD

Develop with AMD

September 28, 2005 | Computation Products Group | 14

Develop with AMD

Doing More with Dual

Utilize AMD Optimized / Multi-threaded Math Libraries

AMD

- **ACML**
  - **BLAS**, **LAPACK**, **FFT**s, **SCALAPACK** and vectorized **MLIB**
  - +1.5 million lines of **AMD** invested assembly optimization
  - parallelized with Open-MP (OMP) and MPI
  - **PGI OMP** support for **Win**$_{64}$ *BETA*, **Win**$_{32}$, **Linux**$_{64}$ and **Linux**$_{32}$
- **Multi-threaded/OMP PGI/Pathscale Libraries**
  - matrix-matrix and multi-dimensional **FFT**s scale well on dual-core
  - scaling dictated by % of time spent in parallelized code regions
- **Increase your productivity**
  - Equivalent user developed functions will not match **ACML** performance
  - Leverage libraries **AMD** has invested in rather than developing your own

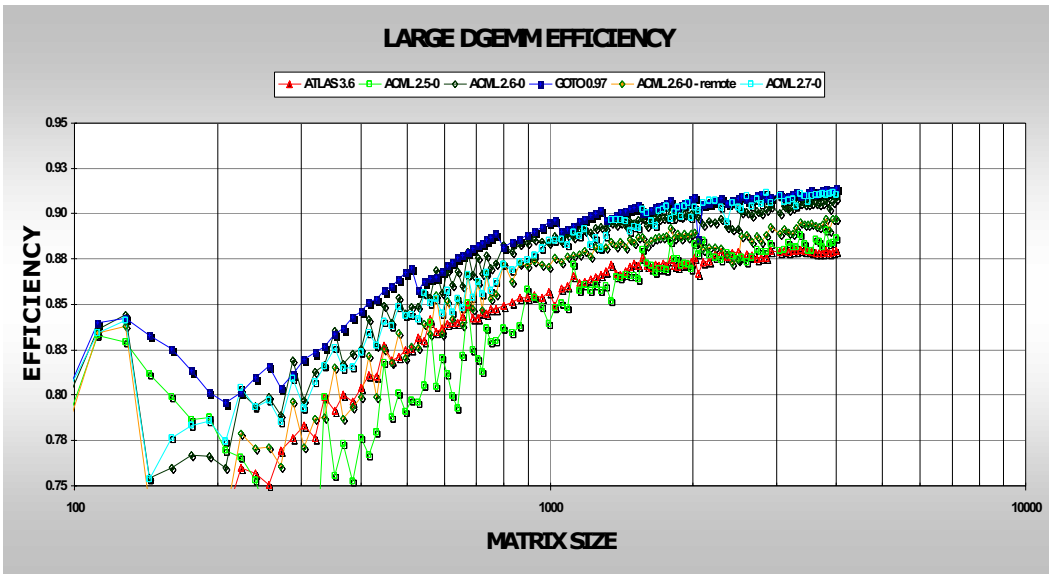
September 28, 2005 | Computation Products Group | 15

Develop with AMD

64-bit DGEMM Performance

Large Argument Regime

AMD



September 28, 2005 | Computation Products Group | 16

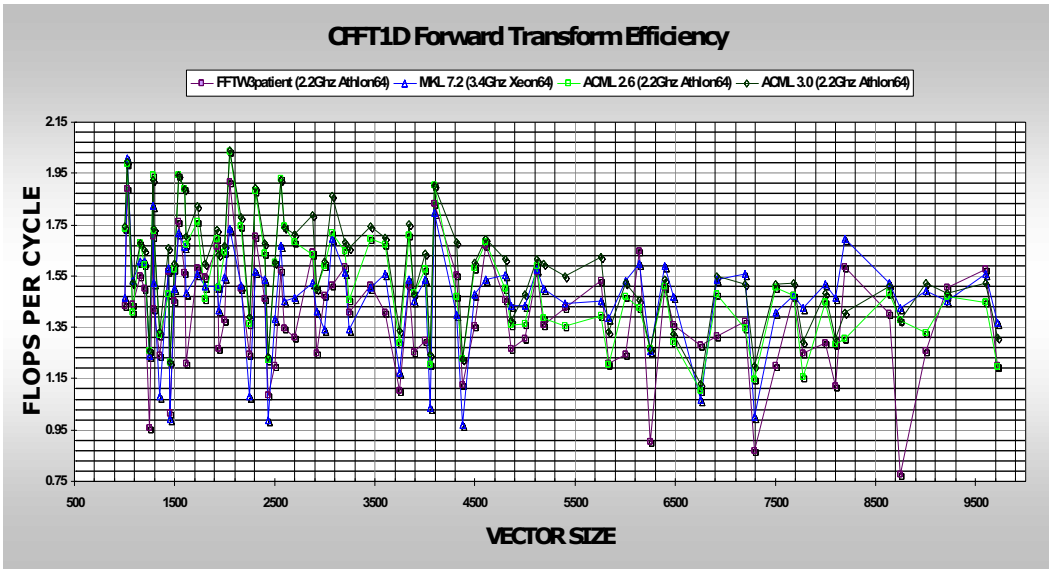
© 2005 Copyright by DYNA*more* GmbH

Develop with AMD

AMD

64-bit FFT Performance (non-power of 2)

ACML 2.6, MKL 7.2 and FFTW 3

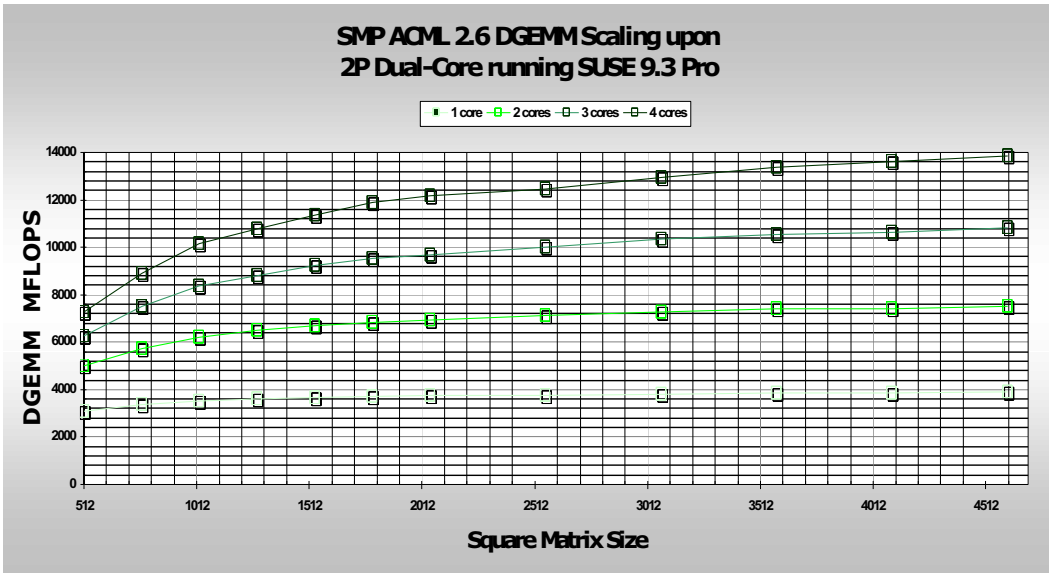


Develop with AMD

AMD

Doing More with Dual

Utilize AMD Optimized / Multi-threaded Math Libraries



© 2005 Copyright by DYNA*more* GmbH

AMD

Compiler Ecosystem

PGI , Pathscale , GNU , Absoft
Intel, Microsoft and SUN

Develop
with AMD
Compiler Truth Table

AMD

	Vector SIMD Support	Peels Vector Loops	Global IPA	Open MP	Links ACML Libraries	Profile Guided Feedback	Aligns Vector Loops	Parallel Debuggers	Large Array Support	Medium Memory Model
PGI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
GNU	✗	✗	✗	✗	✓	✓	✗	✗	✗	✓
Intel	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗
Pathscale	✓	✗	✓	✗	✓	✓	✗	✗	✗	✓
Absoft	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SUN	✓	✓	✓	✓	✗	✓	✓	✓		
Microsoft	✗	✗	✓	✓	✓	✓	✗	✓	✗	✗

© 2005 Copyright by DYNA*more* GmbH

Develop with AMD

Intel Compiler and MKL upon Opteron

Threat Assessment to Opteron Performance

AMD

- **The compiler is a weapon – maker can control the code generated and run upon their chip and their competitor**
  - working with PGI and NAG we can address the performance and functionality issues of a customer by modifying the compiler or ACML
- **CPUID checks – instruction compatibility not checked but rather the Vendor ID**
  - AMD platform not supported unless reproducible on Intel platforms
  - CPUID checks placed into code because Intel doesn't trust users intellect

http://support.intel.com/support/performancetools/c/sb/cs-009787.htm

Issues on AMD platforms can not be addressed and will not be reproducible since we do not issue the same VENDOR ID in the CPUID instruction → ISVs and customers draw the conclusion AMD Platforms aren't dependable

September 28, 2005 | Computation Products Group | 21

Develop with AMD

Tuning Memory IO Bandwidth

Optimizing Large Streaming Operations

AMD

- **2 Methods of writing to memory in x86/x86-64:**
  - traditional memory stores cause write allocates to cache

```
Mov %rax,[%rdi]  movsd %xmm0,[%rdi]  movapd %xmm0,[%rdi]
```

    1. page to be modified is read into cache
    2. cache is modified, written to memory when new memory page loaded
    3. **to write N bytes, 2N bytes of bandwidth generated**
  - non-temporal stores bypass cache and write directly to memory
    1. no write allocate to cache, to write N bytes, **N bytes of bandwidth generated**
    2. data is not backed up into cache, do not use with often reused data
- **Use only on functions which write L2/2 > bytes of data or more, normally would assure little cache reuse value**

Group all eligible routines into a common file to as to simplify the compilation procedure. Enable non-temporal stores in PGI compiler with the –Mnontemporal compiler option

September 28, 2005 | Computation Products Group | 22

Develop with AMD

Tips to Doing More with Dual

Controlling CPU Core/Memory Affinity

AMD

- **"numactl" command allows control of core/memory affinity for a process "test.exe":**

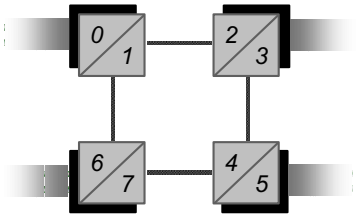
  EXAMPLE: HP DL585 with 4 DC Opterons (8 cores)

  - run upon 4 sockets using only 1 of every 2 cores

    ```
    numactl -cpubind=0,2,4,8 -membind=0,2,4,8 test.exe
    ```

  - run upon 4 sockets but interleave memory for process across all 4 IO controllers

    ```
    numactl -interleave=0,2,4,8 test.exe
    ```



September 28, 2005 | Computation Products Group | 23

Develop with AMD

Tips to Doing More with Dual

Controlling CPU Core/Memory Affinity

AMD

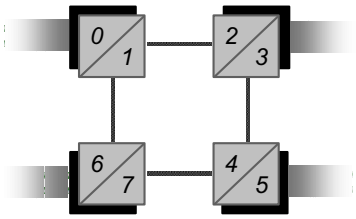
- **PGI 6.0-5 allows users to control process/core affinity in OMP applications when compiled with –mp=numa[,align]**

  EXAMPLE: HP DL585 with 4 DC Opterons (8 cores)

  - first set environment variables required for OMP and NUMA

    ```
    Setenv OMP_NUM_THREADS 4
    Setenv MP_BIND yes
    Setenv MP_BLIST 0,2,4,6,1,3,5,7
    ```

  - run however many jobs you wish to run 1-8 way knowing they are locked to a core, do not wander and maximally utilize memory IO



September 28, 2005 | Computation Products Group | 24

© 2005 Copyright by DYNA*more* GmbH

Develop with AMD

Tips to Doing More with Dual

Controlling CPU Core/Memory Affinity

AMD

- **Schedule Utilities package "taskset" allows users to dictate/change placement of processes:**

  http://www.novell.com/products/linuxpackages/professional/schedutils.html

  - Allows user to monitor cores upon which a process is run

    ```
    taskset -p <PROCESS ID>
    ```

  - Change the cores currently processing <PROCESS ID>

    ```
    taskset -cp 0,2 <PROCESS ID>
    ```

  - Run an executable upon nodes 0 and 2

    ```
    taskset -c 0,2 test.exe
    ```

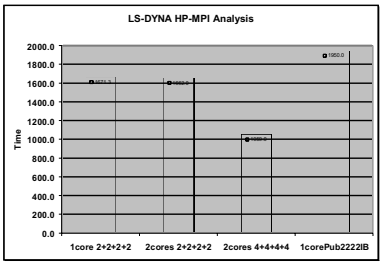

Develop with AMD

Tips to Doing More with Dual

Dual Core scales seamlessly upon OpenMP and MPI

AMD

- **Utilize OMP and MPI upon Dual-core AMD systems**
  - no software investment required to tune for dual core
  - to OS dual core is just doubling the # of "real" cores
  - upon AMD dual-core technology in MCAE: +80% scaling
- **Utilize HP-MPI**
  - Allows user to specify core and memory affinity for MPI processes



© 2005 Copyright by DYNA*more* GmbH

Develop with AMD

64-bit Operating Systems
Status and Recommendations

AMD

- **SUSE SLES 9 with latest Service Pack available**
  - Has technology for supporting latest AMD processor features
  - Widest breadth of NUMA support and enabled by default
  - Oprofile system profiler installable as an RPM and modularized
  - complete support for static & dynamically linked 32-bit binaries
- **Red Hat Enterprise Server 3.0 Service Pack 2 or later**
  - NUMA features support not as complete as that of **SUSE SLES 9**
  - Oprofile installable as an RPM but installation is not modularized and may require a kernel rebuild if RPM version isn't satisfactory
  - only SP 2 or later has complete 32-bit shared object library support (a requirement to run all 32-bit binaries in 64-bit)
  - Posix-threading library changed between 2.1 and 3.0, may require users to rebuild applications

AMD

Performance Benchmarks

© 2005 Copyright by DYNA*more* GmbH

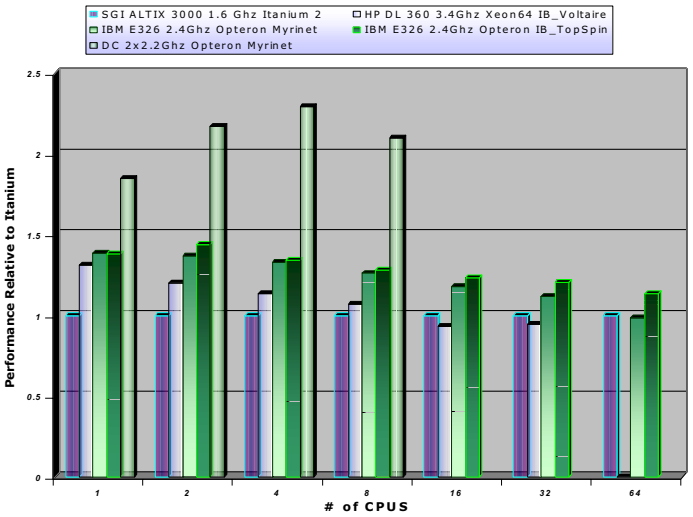
Performance Benchmarks

64-bit Fluent v6.2

FL5M2 Performance

AMD

Fluent 6.2 Timing upon FL5M2 Model



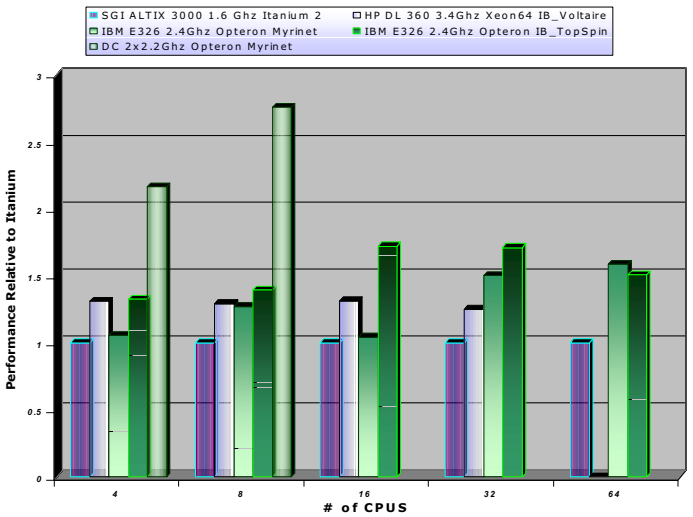
Performance Benchmarks

64-bit Fluent v6.2

FL5L3 Performance

AMD

Fluent 6.2 Timing upon FL5L3 Model



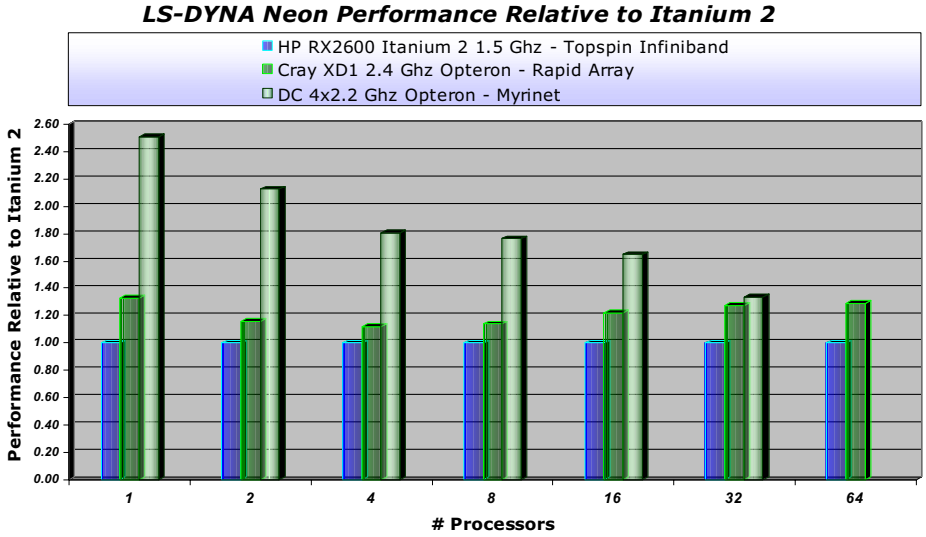
© 2005 Copyright by DYNA*more* GmbH

Performance Benchmarks

64-bit LS-DYNA 5434a

Neon Benchmark Performance

AMD



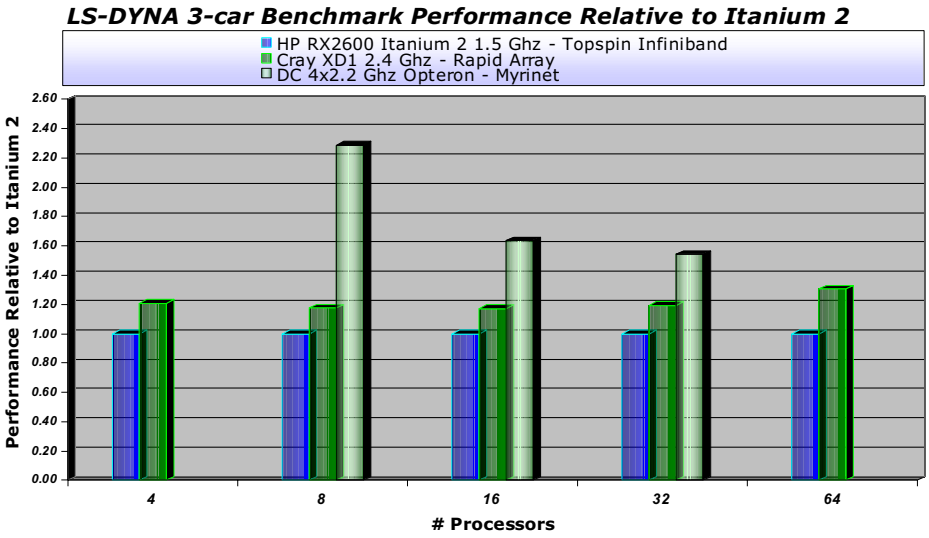
September 28, 2005 | Computation Products Group | 31

Performance Benchmarks

64-bit LS-DYNA 5434a

3-car Benchmark Performance

AMD



September 28, 2005 | Computation Products Group | 32

© 2005 Copyright by DYNA*more* GmbH

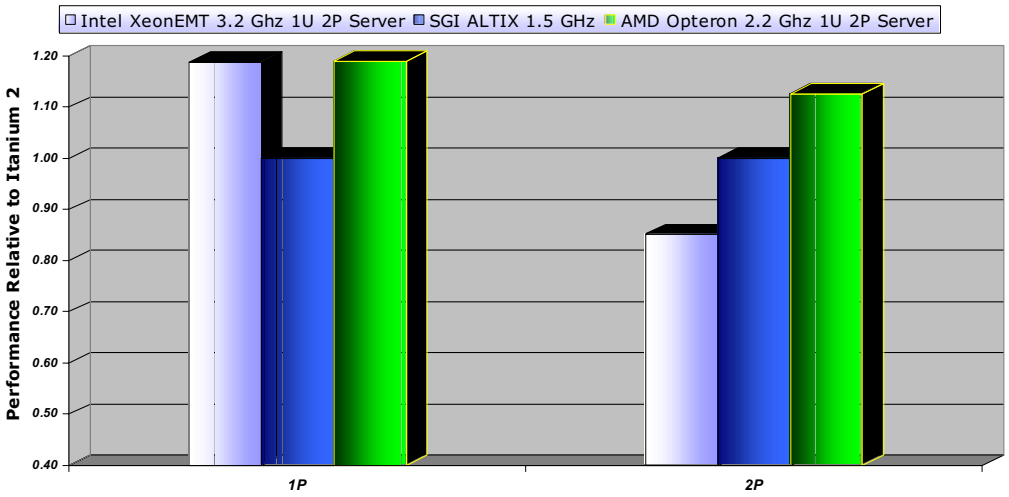
Performance Benchmarks

64-bit STAR-CD

Engine Benchmark performance

AMD

Star-CD Engine Benchmark Performance Relative to Itanium 2



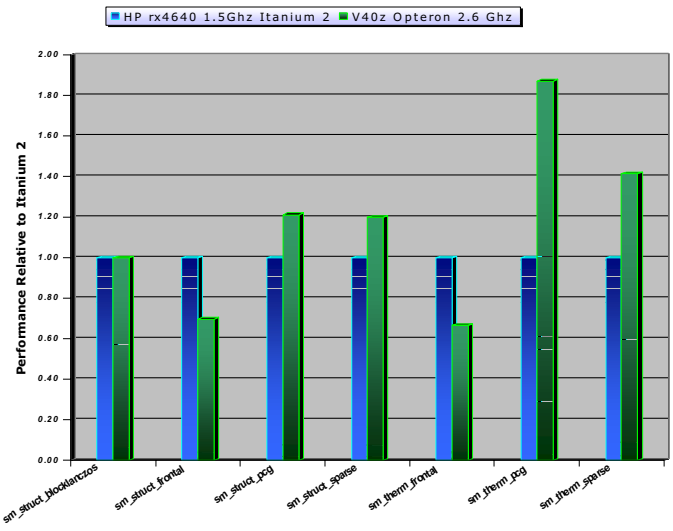
Performance Benchmarks

64-bit ANSYS 9.0

Brake Automotive Performance (4P)

AMD

Ansys 9.0 Automotive Brake Rotor Benchmarks



© 2005 Copyright by DYNA*more* GmbH

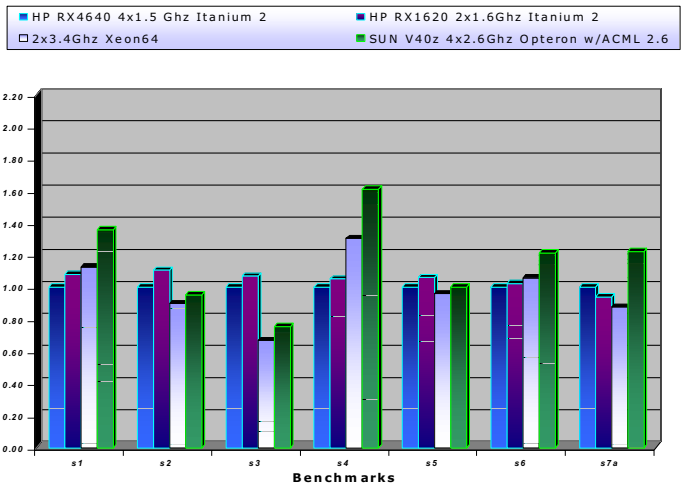
Performance Benchmarks

64-bit ABAQUS 6.5

Standard Benchmark Performance (1P)

AMD

1P ABAQUS 6.5 Performance Relative to RX4640 Itanium 2 1.5Ghz



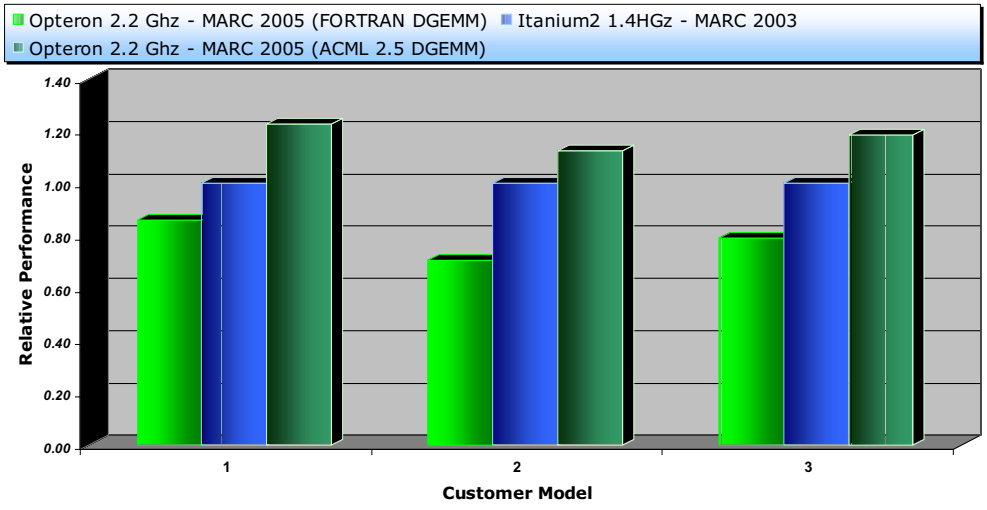
Performance Benchmarks

64-bit MARC 2005

Customer Model Performance

AMD

MSC MARC *Performance Relative to Itanium 2*



© 2005 Copyright by DYNA*more* GmbH

Trademark Attribution

AMD, the AMD Arrow Logo, AMD Opteron and combinations thereof are trademarks of Advanced Micro Devices, Inc. HyperTransport is a licensed trademark of the HyperTransport Technology Consortium. Other product names used in this presentation are for identification purposes only and may be trademarks of their respective companies.

 © 2005 Copyright by DYNA*more* GmbH