Tera Tera Tera Tera Tera

Ed Davis – Chief Architect
AGENDA

- Supercomputing Roadmap
- Quad-core, Eight-core, 10-100 core
- Compute / Networking / Storage / Software
- Tera-Scale Computing
- Software Solutions & Challenges
- Ongoing Research
Pat Gelsinger laying out the future roadmap hurdles

- Xonabyte = $10^{27}$
- Wekabyte = $10^{30}$
- Vundabyte = $10^{33}$
- Udabyte = $10^{36}$
- Tredabyte = $10^{39}$

Currently Google uses 2 PetaBytes of disk space
We’ve already achieved a PetaFLOP in aggregate now the race is on for World’s first PetaFLOP system
Quadcore Launch at CERN

- Pat Gelsinger Sr. V.P. and GM Digital Enterprise Group Launches Quad-core “Clovertown” at CERN
Intel’s Quad-Core solution is a careful choice of performance, schedule, and cost.

Clovertown (Server)
Kentsfield (Desktop)

+ Faster TTM (Out ship AMD 20:1 in the first 3 Qtrs)
+ Simple Coherency Model
+ Flexible Die Selection
+ Utilizes Capacity

Up to 70% better perf. over Core2 Duo*

Gigabyte GA-965P-S3

1066/1333 MHz

*SPECint_rate
“Intel Enhances Its Best Processors”
- Computing News, Nov 2006

“If you want quad-core today, Intel is the only game in town”

“A total of eight cores offering unmatched server performance”
- Koen Crijns, Hardware.Info, Nov 13, 2006

“With no price premium, Intel has laid the gauntlet down for AMD.”
- Charlie Demerjian, Inquirer Nov 14, 2006

“Woodcrest offers better performance than [Opteron] “Rev F”, but Clovertown takes performance to a new level”
- Thomas Weisel Partners, EQUITY RESEARCH, Nov 14, 2006

“These new Xeon chips fit into the same power envelope as their dual-core predecessors, but pack twice the processor cores”
Intel will ship greater than 60 million multi-core CPUs by EOY 2006.

Source – Intel Corporation
Advancing to Terascale Applications

- **Tera-scale**
- **Dataset Size**
  - Kilobytes
  - Megabytes
  - Gigabytes
  - Terabytes

- **Performance**
  - TIPS
  - GIPS
  - MIPS
  - KIPS

- **RMS**
  - 3D & Video
  - Multi-Media
  - Single-core
  - Multi-core

- **Applications**
  - Entertainment
  - Learning & Travel
  - Personal Media Creation and Management
  - Next Gen. Particle Accelerators LHC, ILC, ...
  - Health
  - Source: electronic visualization lab University of Illinois

- **Multimedia**
- **Text**

- **Intel** Leap ahead
Compute, Storage, Networking & Software

Intel® Blade Server Family
Consolidate – Compute, Networking, Storage, & Management

Compute
- Dual and quad processor compute blade options

Networking
- Ethernet switching & other fabric options

Storage
- SAN & NAS connections via Fibre Channel & Ethernet

Integration of resources in a single package for improved utilization

Management
- Complementary hardware and software simplifies daily tasks
HPC Application Spectrum

Bandwidth and processor compute capability assessed
Applications span the spectrum
No single industry accepted metric exists

Bandwidth Limited

Core Limited

NWChem
Fluid Dynamics
Ocean Models

Petro Reservoir
Auto NVH

Auto Crash
Weather

Seismic
GAMESS

Stream
DAXPY
SparseMV

~1 Byte per
Flop

SPECfp2000

Linpack
DGEMM

http://www.linuxclustersinstitute.org/Linux-HPC-Revolution/Archive/PDF05/presentations/Pase_D.pdf
Compute: TERAFLOP OF PERFORMANCE

- **80 Cores**
- **98 Watts**
- **256 GB/s bisection**

Source: Intel
Storage: TERABYTES OF BANDWIDTH

256 KB SRAM per core
4X C4 bump density
8490 thru-silicon vias

Polaris with Cu bump

Source: Intel
Networking/Comms: TERABITS OF I/O

TERABITS OF I/O
Target: 1 Tbps chip-to-chip link

Optical Fiber

30 Gb/s modulator

Multiplexer

Modulators

Hybrid lasers

Source: Intel
PARALLEL PROGRAMMING

CHALLENGES

- Extracting concurrency
  - Tools
- Expressing concurrency
  - Programming abstractions & languages
- Exploiting concurrency
  - Compilers, runtimes & hardware support

KEY IS TO IDENTIFY (algorithmic, manual) & MANAGE (locks, conditionals) PARALLELISM
Chip and System Interconnects

Optical
- Metro & Long Haul: 0.1 - 80 km
- Intel Optical Products
- Rack to Rack: 1 to 100 m

Copper
- Chip to Chip: 1 - 50 cm
- Board to Board: 50 - 100 cm

Decreasing Distances →

Driving optical to higher volume and lower cost
Transactional memory is a technique for coordinating how multiple threads access the same memory.

- Ensures concurrent access with no errors
- Greater performance due to concurrent execution
- Eases the writing of parallel programs that work and scale
- Eliminates deadlocks
- Provides a failure recovery mechanism

Java based proof-of-concept shows 3-4x performance gain on a 16-way Xeon® Processor-based system.
Robson Technology

- Over the past 10 years alone, processor performance has increased by over 30X while hard-drive performance has increased by only 1.3X. And, the gap will continue to grow.
- A factor of 100,000 difference between DRAM and HDD (random read latency of 150 nanoseconds vs. 15 milliseconds).
- About two orders of magnitude in cost per bit for equivalent capacity.
- NAND Flash could go between L2 and Disk or in some cases to replace disk altogether.
- Robson’s smart controller uses clever write-leveling algorithms to spread the block erasures evenly across the array giving the NAND flash memory a service life consistent with the rest of the platform.

http://blogs.zdnet.com/OverTheHorizon/?p=11
UNLOCKING PARALLELISM

Must carefully control how multiple threads access common memory

Today we “lock” memory for one thread at a time

- Other threads must wait, reducing multi-core benefit
- Locking code scales poorly, must re-do for more threads
- Can cause critical software deadlocks and errors
Virtual Container VMM:

OS | OS | OS

Operating System:

App | App | App

Application:

Thread | Thread | Thread

Thread:

Code Fragment | Code Segment | Code Fragment

- New Parallel Languages
- New Threading tools
- Thread Management & Abstraction layers
- Transactional memory
- Auto-threading compilers
- Auto-threading hardware
## Media and Music
- Real Networks, MusicMatch, Sony, Thompson, Windows Media

## Digital Home
- Roxio, Microsoft, Pinnacle, Cakewalk, Steinberg

## Digital Office
- Microsoft Office, UGS

## Games
- EPIC, Id, Square Enix, Activision

## Content Creation
- Macromedia, Adobe, Discreet, Pixar, Alias, Canopus

## Server
- Oracle, SQL, DB2, Exchange, SAP, Lotus, BEA, SAS
TERA-SCALE RESEARCH

Microprocessor

Essential
- Scalable memory
- Multi-core architectures
- Specialized cores
- Scalable fabrics
- Energy efficient circuits

Complementary
- Si Process Technology
- Resiliency

Platform

Essential
- 3D Stacked Memory
- Cache Hierarchy
- Virtualization/Partitioning
- Scalable OS’s
- I/O & Networking

Complementary
- CMOS Radio
- Photonics

Programming

Essential
- Speculative Multithreading
- Transactional memory
- Data parallel operations
- Compilers & Libraries
- Tools

Complementary
- Domain specific languages
- Usage Models

100+ Projects Worldwide
PUTTING IT ALL TOGETHER

- New instructions
- Cache improvements
- HW thread scheduling
- Baseline

Value of Tera-scale Research
Just Adding Cores

Speedup
Number of cores

- 1
- 2
- 4
- 8
- 16
- 32
- 64

0
5
10
15
20
25
30
Figure 1. Scaling transistors. The number of transistors is expected to continue to double about every two years, in accordance with Moore’s Law. Over time, the number of additional transistors will allow designers to increase the number of cores per chip.

A Multiplier is about 150K gates, roughly 7 per million gates.
About 1200 will fit, running at 2 GHz is 2.4 TeraFLOPS!

AND IT GETS BETTER!!! >>>>
**Intel – New technology generation every 2 years**

### 45 nm Logic Process on Track for Delivery in 2007

<table>
<thead>
<tr>
<th>Process Name</th>
<th>P1262</th>
<th>P1264</th>
<th>P1266</th>
<th>P1268</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithography</td>
<td>90 nm</td>
<td>65 nm</td>
<td>45 nm</td>
<td>32 nm</td>
</tr>
<tr>
<td>1st Production</td>
<td>2003</td>
<td>2005</td>
<td>2007</td>
<td>2009</td>
</tr>
</tbody>
</table>

*Moore’s Law continues!*

Intel continues to develop a new technology generation every 2 years.
Shrinking geometries rapidly is a huge win, but it’s not the only way to win! More efficient use of the silicon is equally important.

Intel will continue to Lead the Way to Exa, Yotta, Zetta,...!!!
Comparison AMD, Xeon, Itanium Memory Speed, etc.
http://www.amd.com/us-en/Processors/ProductInformation/0,,30_118_8796_8799,00.html
MULTI-CORE ENABLING INITIATIVES

**USE INNOVATION**
Joint effort with SSG, CTG and DEG focused on driving MC uses into BUs

**DEV. OUTREACH**
64 classes, 1587 Students, Webcasts and Tutorials: 14,239, Downloads: 19,356

**ENGR. METHODS**
Tools: Compiler, Libraries, Thread Checking, performance analyzers;

**RESEARCH AND EDUCATION**
30 3-YR grants awarded, Apps, Algo., Tools, 20 Unvvs. offering courses

**GTM PROGRAMS**
GTM Program with MC-enabled ISVs Reseller Matchmaking program for ISVs
SYSTEM CONFIGURATION

1. SPECjbb™ 2005 Configuration Details: Published/measured results as of Oct 09, 2006. System power was measured during the steady-state of window of the performance run.


   - Dual-Core Intel Xeon Processor 5160 based platform details: Intel Server preproduction platform with two Dual-Core Intel Xeon Processor 5160, 3.00 GHz with 4M L2 Cache, 1333 MHz system bus, 8GB (8x1GB) 667MHz FB-DIMM memory, Windows 2003 Enterprise Edition. BEA JRockit(R) 5.0 P26.4.0. Run with two jvm instances

   - Quad-Core Intel Xeon Processor based platform details: Intel Server preproduction platform with two Quad-Core Intel Xeon Processor X5345, 2.33 GHz with 2x4M L2 Cache, 1333 MHz system bus, 8GB (8x1GB) 667MHz FB-DIMM memory, Windows 2003 Enterprise Edition. BEA JRockit(R) 5.0 P26.4.0. Run with two jvm instances

2. Performance tests and ratings are measured using specific systems and/or components and reflect approximate performance of Intel products as measured by those tests. Any difference in system hardware, software, or configuration may affect actual performance. Buyers should consult other sources of information to evaluate system or component performance they are considering purchasing. For information on performance tests and performance of Intel products, visit http://www.intel.com/performance/resources/limits.htm
Intel® Xeon® 7100 Series 4-Way Servers

- **Configurations Published:**
  - Database Performance on TPC-C*: represents the transaction throughput of a database server in an OLTP client/server environment.
  - **Intel Xeon 7140M platform** IBM® System x3950 Server System using 4x Dual-Core Intel® Xeon® processor 7140N (3.33 GHz, 16 MB L3 cache, 4 Processors/8 Cores/16 Threads), 32x 4 GB (128 GB) memory, DB2® 9 database, SuSe® LINUX Enterprise 10. Results at http://www.tpc.org/results/individual_results/IBM/ibm_x3950_DB2_Linux_es.pdf Referenced as published: 314,468 tpmC; $4.75/tpmC; Availability Date Nov 30, 2006.
MULTI-CORE MOTIVATION

Over-clocked (+20%)

- Relative single-core frequency:
  - Dual-Core: 1.73x
  - Performance: 1.13x
  - Power: 1.00x

Design Frequency

- Dual-Core: 1.73x
- Design: 1.00x
- Frequency: 22%

Under-clocked Dual-core (-20%)

- Dual-Core: 1.73x
- Under-clocked: 0.87x
- Design: 1.02x
- Dual-core: 51%

Relative single-core frequency and Vcc
MULTI-CORE POWER EFFICIENCY

System Power (measured system Watts)
- Lower is better

<table>
<thead>
<tr>
<th>Single 3.6 GHZ</th>
<th>Dual 5160 3.0 GHZ</th>
<th>Quad E5345 2.33 GHZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>323</td>
<td>322</td>
<td>335</td>
</tr>
</tbody>
</table>

Performance (SPECjbb2005*)
- Higher is better

<table>
<thead>
<tr>
<th>Single 3.6 GHZ</th>
<th>Dual 5160 3.0 GHZ</th>
<th>Quad E5345 2.33 GHZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>33114</td>
<td>102313</td>
<td>153500</td>
</tr>
</tbody>
</table>

Exceptional Software Scaling**
Same Power Envelope
Outstanding Performance per Watt

* [http://www.principledtechnologies.com/](http://www.principledtechnologies.com/)
** For most software applications