Intel® Pentium® 4 Processor on 90nm Technology

Ronak Singhal

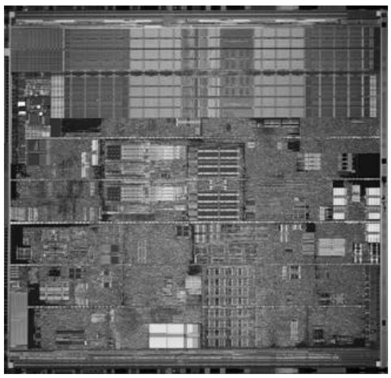
August 24, 2004

Hot Chips 16

Agenda

- Netburst® Microarchitecture Review
- Microarchitecture Features
- Hyper-Threading Technology
- SSE3
- Intel® Extended Memory 64 Technology
- Performance Results

Vital Statistics

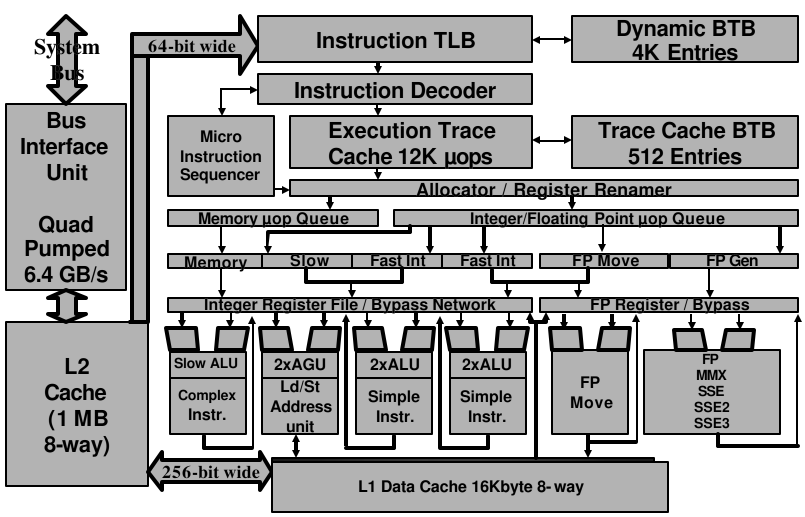


- 125 million transistors
- 112 mm^2 die size
- 90nm manufacturing process
- Introduced in Feb. 2004 @ >3GHz

3

3

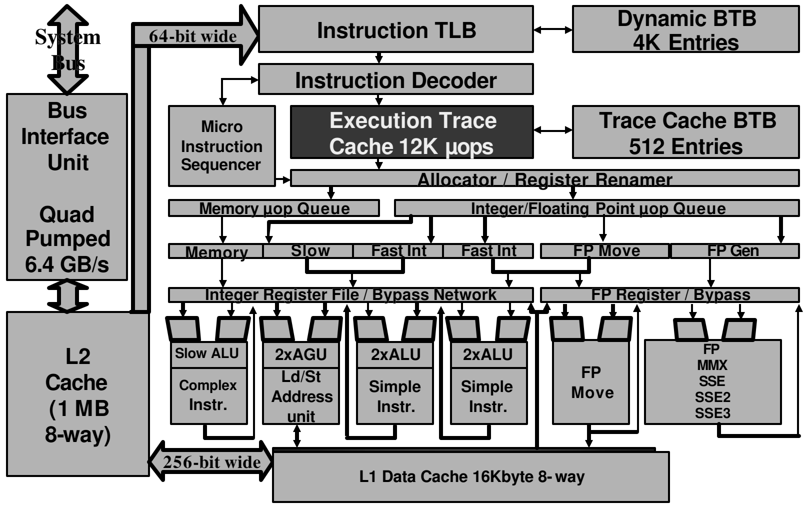
Intel® Pentium® 4 Processor Block Diagram



4

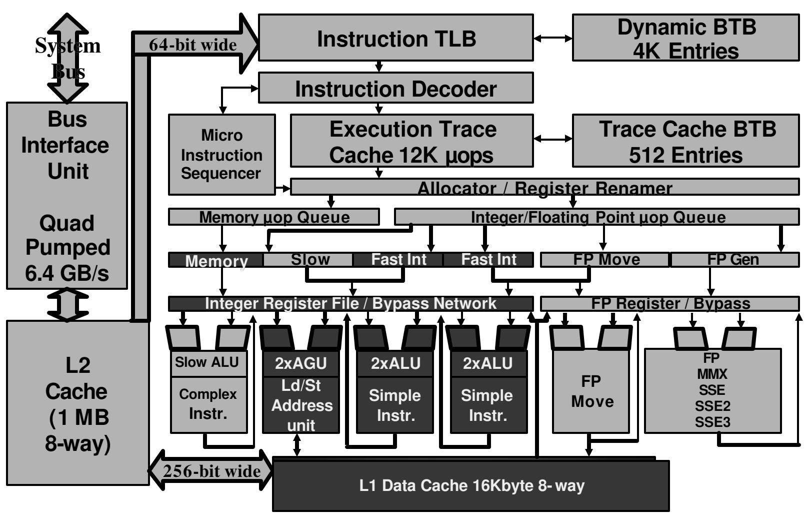
4

Key Characteristics



Trace Cache instead of conventional I-Cache

Key Characteristics



2x Frequency Execution Core for High Throughput

New Microarchitecture Features

- Larger Caches
- Deeper Buffers
- Faster Execution Units
- Algorithmic Enhancements

Cache Comparison

	130nm	90nm
1st level data cache	8KB, 4-ways, Write-through	16KB, 8-ways, Write-through
2nd level data cache	512KB, 8-ways, Write-back	1MB, 8-ways, Write-back
Trace Cache	12k uops	12k uops

Larger Buffers

	130nm	**90nm**
ROB Size	126	126
Load Buffers	48	48
Store Buffers	24	32
Write Combining Buffers	6	8
Outstanding 1st level Data Cache Misses	4	8
FP Schedulers	10/12	14/16

9

9

Faster Execution Units

- Shifts
  - Typical shifts now handled inside of fast execution core w/ single cycle latency
  - Previously handled in complex integer unit with 6 cycle latency
- Integer Multiply
  - Adds dedicated integer multiplier
  - Previously handled by the FP multiplier

10

10

Algorithmic Enhancements

- Branch Prediction
- Hardware Prefetching

11

11

Branch Prediction

- Continued improvement of existing algorithms
- Improved static prediction algorithm
  - Displacement check
  - Condition check
- Added indirect branch predictor
  - Idea first introduced on Intel® Pentium® M processor

12

12

Branch Predictor Comparison

	130nm	90nm
164.gzip	1.03	1.01
175.vpr	1.32	1.21
176.gcc	0.85	0.70
181.mcf	1.35	1.22
186.crafty	0.72	0.68
197.parser	1.06	0.87
252.eon	0.44	0.39
253.perlbmk	0.62	0.28
254.gap	0.33	0.24
255.vortex	0.08	0.09
256.bzip2	1.19	1.12
300.twolf	1.32	1.23

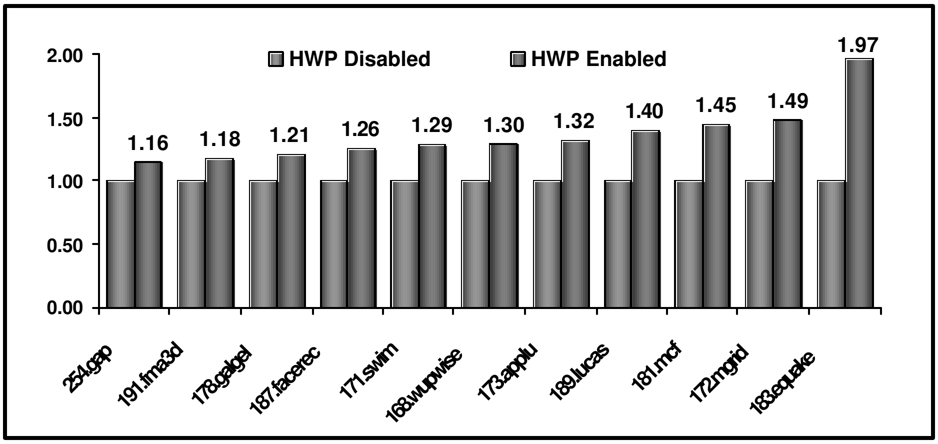
# of Branch Mispredicts Per 100 Instructions on SPECint*_base2000

* Other names and brands are the property of their respective owners

Hardware Prefetching

- Primary mechanism to hide DRAM latency
- Processor predicts what data will be needed in the future and proactively fetches it from DRAM
- Exists on all Intel® Pentium® 4 Processor implementations
- 90nm version improves on what data to get and when to get it

Hardware Prefetching



Impact of HW prefetcher on most sensitive benchmarks in SPEC CPU2000

Hyper-Threading Technology†

- Makes a single processor look like two processors to software
- Takes advantage of underutilized resources when running a single thread through the processor

† Hyper-Threading Technology requires a computer system with an Intel® Pentium® 4 processor supporting HT Technology and a Hyper-Threading Technology enabled chipset, BIOS and operating system. Performance will vary depending on the specific hardware and software you use. See http://www.intel.com/info/hyperthreading/ for more information including details on which processors support HT Technology.

Hyper-Threading Technology Improvements

- 1st level data cache
  - Uses partial virtual address index
  - Aliasing can occur due to stacks of two threads being offset by a fixed amount
  - Use context identifier to differentiate between data from different threads. Better than thread identifier to allow data sharing between threads.
  - Introduced on later steppings of 130nm version.
- Parallel Operations
  - Allow page walks and split memory access handling in parallel
  - Allow multiple page walks if one goes to DRAM
- Buffer sizes
  - Motivated increase in # of outstanding 1st level cache misses

17

17

SSE3

- 13 new instructions
  - x87 to integer conversion
  - Graphics (Horizontal Add/Subtract)
  - Complex arithmetic
  - Video Encoding
  - Thread Synchronization

18

18

Complex Arithmetic

- MOVDDUP, MOVSHDUP, MOVSLDUP
  - Instructions to load and duplicate data implicitly
- ADDSUBPS, ADDSUBPD
  - Perform a mix of addition and subtraction simultaneously
- 10-20% gain on 168.wupwise from these instructions (complex matrix multiply)

19

Video Encoding

- Motion Estimation compares previous frame to current frame
  - Loads from the previous frame are unaligned
  - Leads to costly cache line split memory accesses
- LDDQU instruction loads 128-bits at an arbitrary alignment with no cache line split
- Speedups of > 10% on MPEG-4 encoders

20

Thread Synchronization

- Used to indicate that a thread is spinning and waiting for work
- Allows processor to go into an optimized state
- MONITOR – Sets up address monitoring hardware
- MWAIT – Sets processor into optimized state. Will wake up when monitored address in written to

21

21

Intel® Extended Memory 64 Technology

- Additional capability in today's Intel® Xeon™ processors on top of:
  - Netburst® Microarchitecture
  - Hyper-Threading Technology
  - SSE3
- Provides 8 more integer and SSE registers
- Larger addressing capability
  - 48 bits of virtual address on this implementation
  - 36-40 bits of physical address on this implementation
- Full 64-bit support carefully engineered into the 90nm design
  - Limited differences between 32-bit and 64-bit operations
  - Similar optimizations for 32-bit and 64-bit code

22

22

32-bit vs. 64-bit Comparison

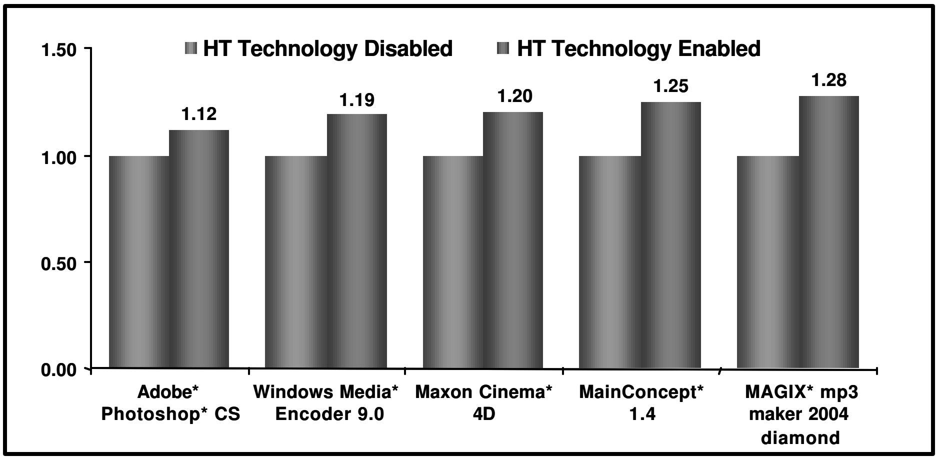
	32-bit	64-bit
ALU Latency	1 cycle	
ALU Throughput	4 operations/cycle	
Memory Throughput	1 load + 1 store per cycle	
Page walks	2 levels -- use PDE cache to reduce to 1 level in common case	4 levels – use PDE cache to reduce to 1 level in common case

Enable strong 64-bit performance without compromising 32-bit performance

Optimizing 64-bit code

- Rule #1: Follow 32-bit optimizations
- Rule #2: Compile with Pentium 4 specific optimizations enabled
- Few additional new rules:
  - When data size is 32 bits, typically use 32-bit instructions
    - Example: XOR EAX, EAX instead of XOR RAX, RAX
  - But sign extend to full 64-bits instead of only 32 bits (even for 32-bit data size)
    - Example: Load 16 bits, sign extend to RAX not EAX

Performance Results

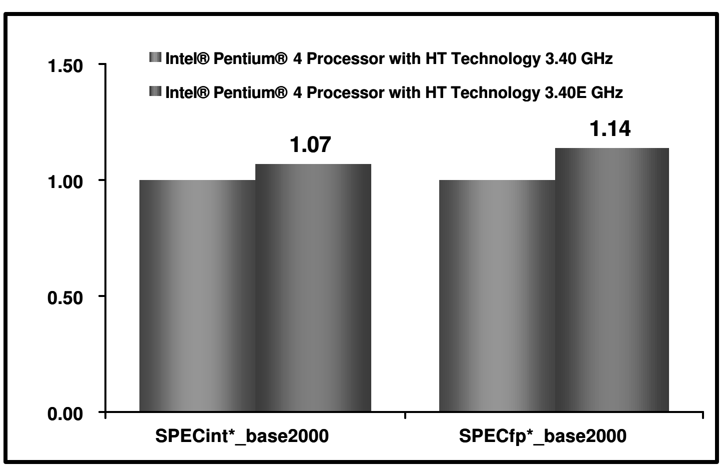


Source: Intel Configuration: **Intel® Pentium® 4 processor with HT Technology 3.40E GHz** – Intel® D875PBZ Desktop Board (AA-301); **All Platforms** – 1GB DDR400 CL3-3-3, ATI* Radeon* 9800 Pro AGP graphics, ATI* Catalyst* 3.5 Driver Suite: display driver 6.14.10.6360, Intel® Application Accelerator RAID Edition 3.5 with RAID ready, Intel® Chipset Software Installation Utility 5.01.1015, Seagate* Barracuda* 7200 Serial ATA 160GB Hard Drive - ST3160023AS, Intel C & Fortran compilers 8.0, DirectX* 9.0b, Windows* XP Build 2600 SP1, Intel® PRO/1000 MT Desktop Adapter. *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.*

25

25

Performance Results

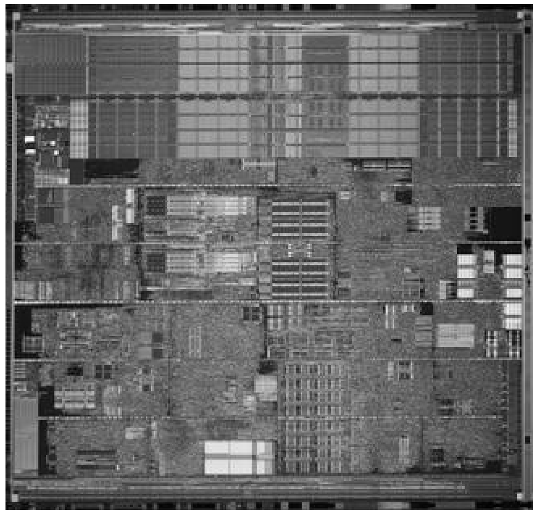


Source: Intel Configuration: **Intel® Pentium® 4 processor with HT Technology 3.40 GHz** – Intel® D875PBZ Desktop Board (AA-204); **Intel® Pentium® 4 processor with HT Technology 3.40E GHz** – Intel® D875PBZ Desktop Board (AA-301); **All Platforms** – 1GB DDR400 CL3-3-3, ATI* Radeon* 9800 Pro AGP graphics, ATI* Catalyst* 3.5 Driver Suite: display driver 6.14.10.6360, Intel® Application Accelerator RAID Edition 3.5 with RAID ready, Intel® Chipset Software Installation Utility 5.01.1015, Seagate* Barracuda* 7200 Serial ATA 160GB Hard Drive - ST3160023AS, Intel C & Fortran compilers 8.0, DirectX* 9.0b, Windows* XP Build 2600 SP1, Intel® PRO/1000 MT Desktop Adapter. *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.*

26

26

The End



27