Design Objectives of the 0.35\(\mu\)m Alpha 21164 Microprocessor
(A 500MHz Quad Issue RISC Microprocessor)

Gregg Bouchard
Digital Semiconductor
Digital Equipment Corporation
Hudson, MA
Outline

- 0.35µm Alpha 21164 Overview
- Design Goals
- Technology Issues
- 21164 Internal Architecture Review
- Architectural Enhancements
- System Performance Enhancements
- Alpha Processor RoadMap
- Summary
0.35µm ALPHA 21164

- Technology shrink of 0.5µm design
  + Architectural Enhancements
  + System Performance Enhancements

<table>
<thead>
<tr>
<th></th>
<th>0.5 µm process</th>
<th>0.35 µm process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transistor Count</td>
<td>9.3 Million</td>
<td>9.66 Million</td>
</tr>
<tr>
<td>Die Size</td>
<td>16.5 mm x 18.1 mm</td>
<td>14.4 mm x 14.5 mm</td>
</tr>
<tr>
<td>Power Supply</td>
<td>3.3V external</td>
<td>3.3V external</td>
</tr>
<tr>
<td></td>
<td>3.3V internal</td>
<td>2.5V internal</td>
</tr>
<tr>
<td>WC Power Dissipation</td>
<td>50W @ 300 MHz</td>
<td>37W @ 433 MHz</td>
</tr>
<tr>
<td>Target Cycle Time</td>
<td>300 MHz</td>
<td>433 MHz</td>
</tr>
</tbody>
</table>
Design Goals

- **Reduced Cost**
  - Die size reduction
  - Remove processing steps

- **Higher Performance**
  - 433 MHz design target
  - Architectural enhancements

- **Reduced Power**
  - Lower Core Operating Voltage (2.0v-2.5v)

- **Time to Market**
  - Significant leverage from previous design
# Die Size Analysis

<table>
<thead>
<tr>
<th></th>
<th>X-dim</th>
<th>Y-dim</th>
<th>NormA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>16.5</td>
<td>18.1</td>
<td>1.00</td>
</tr>
<tr>
<td>25% shrink</td>
<td>- 4.1</td>
<td>- 4.5</td>
<td>0.56</td>
</tr>
<tr>
<td>Conversion</td>
<td>+0.0</td>
<td>+0.8</td>
<td>0.62</td>
</tr>
<tr>
<td>LI Removal</td>
<td>+1.8</td>
<td>+0.0</td>
<td>0.70</td>
</tr>
<tr>
<td>Actual</td>
<td>14.4</td>
<td>14.5</td>
<td>0.70</td>
</tr>
<tr>
<td>Ideal 30%</td>
<td>11.5</td>
<td>12.7</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Original Die: 16.5mm x 18.1mm, 298 mm²

Actual shrink: 14.4mm x 14.5mm, 209 mm²

Ideal 30% shrink: 11.5mm x 12.7mm, 146 mm²
Layout Conversion Strategy

• **Full 30% linear shrink was not possible**

• **Solution:**
  – 25% linear shrink
  – Semi-automated conversion of design rules
  – Polysilicon mask layer pushed to full 30% shrink dimensions for performance

• **Redesign of caches**
  – Local Interconnect removed
Layout Conversion Example

Original  After Script  Final

- Well
- Diffusion
- Well Plug
- Poly
- M1C
- Metal 1
- M2C
- Metal 2
- DRC Errors
Speed

- Single-wire, two phase clocking scheme
- 14 gates per cycle including latches
- Single global clock grid
  - Global clock skew < 90ps
  - Local clock skew < 25ps
- Clock statistics (0.5 μm design)
  - Clock load = 3.75 nF
  - Size of final clock inverter = 58 cm
  - Edge rate = 0.5 ns
  - Clocking consumes 40% of chip power
  - $\frac{di}{dt} = 50$ A
Speed Verification

- Test circuits were used to determine the speed scaling of different circuit configurations
  - Predicted average process speed up
  - Identified “slow” circuit types
- New circuits were evaluated in SPICE
- Chip sections with major modifications were completely re-verified
Power

- Significant Power Reduction
- $V_{ddi} = 2.2\text{v}$ (to 2.5v)
- 3.3V only interface to external devices
Alpha 21164 Features

Key Attributes

- **4-way issue superscalar**
  - Up to 2 Integer AND 2 Floating Point instructions issued per CPU cycle.

- **Large on-chip L2 cache**
  - 96KB, writeback, 3-way set associative

- **Fully Pipelined**
  - 7-stage integer pipeline
  - 9-stage floating point pipeline

- **Emphasis on low latency at high clock rate**
- **High-throughput memory subsystem**
Alpha 21164 Block Diagram

Instruction Unit

- Instruction Cache (8KB)

Execution Units

- Four-way Issue Unit
- Integer Unit
- FP Adder
- FP Mult.

Memory Unit

- Merge Logic
- Write-Through Data Cache (8KB)
- Write-back L2 Cache (96KB)

Bus Interface Unit

ITB: 48 entry

DTB: 64 entry

128-bit internal data bus

40b Address

L3 Cache

128b Data
Instruction Issue Pipeline Review

Instruction Cache (8KB) → Instruction Buffer → 2kx2 branch prediction

Next PC

0

1

Prefetch Buffer

S0 S1

Instruction Slot

S2

Instruction Issue

S3

to floating point multiply pipeline

to floating point add pipeline

to integer pipeline 0

to integer pipeline 1

Issue Conflict Checker

X
Execution Pipeline Review

Integer Pipeline 0: arith, logical, ld/st, shift

Integer Pipeline 1: arith, logical, ld, br/jmp

FP Pipeline 0: add, subtract, compare, FP br

FP Pipeline 1: multiply

Int mul

FP div
On-chip Cache Resources Review

- 2 System Commands
- 6 Data Reads
- 6 Data Writes
- 3 Instr. Reads
- 96K, 3 Set, Writeback L2 Cache
- System data
- Victim 0 data
- Victim 1 data

- Miss Addr 0
- Victim Addr 0
- Miss Addr 1
- Victim Addr 1

- 40 bit PA

- 128 bits/cycle
L3 Cache (off-chip)

- L3 cache is a direct-mapped writeback superset of on-chip L2 cache
- Up to 2 reads (or outstanding read commands) in L3 cache
- Programmable wave pipelining for L3 cache
- Support for Synchronous Flow-Thru SRAMs
- L3 cache is optional
Off-Chip L3 Cache Options

Selectable via on-chip programmable registers

- **Cache Size**: 1 to 64M Byte
- **Cache Read/Write Speed**: 4 to 15 cpu cycles
- **Read to Write Spacing**: 1 to 7 cpu cycles
- **Write Pulse (Bit Mask)**: Up to 9 cpu cycles
- **Wave pipelining**: 0 to 3 cpu cycles
- **Support for Synchronous SRAM’s**
Architectural Enhancements

New Instructions

• Scalar support for Byte and Short data types
  – LDBU, LDWU load an unsigned byte or short
  – STB, STW, store an byte or short
  – SEXTB, SEXTW sign extend a byte or short

• Eases porting of device drivers to Alpha

• Improves emulation of Intel code on Alpha

• Implemented in this and all future Alpha microprocessors
Architectural Enhancements

New Instructions (continued)

• IMPLVER
  – returns a small integer indicating the core design
  – used for code scheduling decisions
  – implemented in all Alpha microprocessors

• AMASK
  – clears bits to indicate which features are present
  – implemented in all Alpha microprocessors

Example:
  AMASK     #1, R0 ;byte/word present
  BNE       R0, emulate ;if not emulate
  ...

20
System Performance Enhancements

1. oe_we_active_low (eliminates off-chip buffering)
2. big_drv_en (50% more drive)
3. st_clk1,2 (additional pin)
4. clk_mode<2> (1x clock mode)
5. Bus utilization improvements (Auto Dack)

Bus utilization improvements (Auto Dack)
- cmd: Write
- addr: A0
- data: D00, D01, D02, D03
- dack

Cache Coherence
Pre-Silicon Logic Verification

- Used extensive test suite developed for original chip as testing baseline
- Random and focused testing
- Coverage analysis to ensure excellent test coverage
- Three simulation systems used:
  - RTL
  - Transistor-level
  - Gate-level
Alpha Microprocessor Road Map

SpecInt95

21064 -150MHz
21064 -200MHz
21064A -275MHz
21164 -300MHz
21164 -333MHz
21164 -366MHz
21164 -400MHz
21164 -433MHz
21164 -500MHz
Summary

- FIRST PASS SILICON - October 1995
- Booted first operating system in 2 days!
- Continued Performance Leadership (4+ years)
- Look for next generation 30+ SpecInt95 by Q3’97
Acknowledgments