A 200MFLOP
Precision Architecture Processor

William Jaffe, Bob Miller, Jeff Yetter
Hewlett-Packard
Fort Collins, Colorado

HOT CHIPS IV

"Talk with Jaffe:
- Probably will have to overcome noise in subsequent guesses
- Much longer with off-chip execution counts
- Will use chip with on-chip startup for each
  data than on-chip Dk: size = 100 I/S up for
  much larger blocks
- John Still from Apple left for startup; with them
  came Andrew Fielding.
  "Almost done with large on-chip cache," especially
  for communications gains.

HP Processor Family

<table>
<thead>
<tr>
<th>Year</th>
<th>Technology</th>
<th>Series</th>
<th>800</th>
<th>900</th>
<th>700</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>TTL</td>
<td>series</td>
<td>800</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>NMOSIII</td>
<td>series</td>
<td>800</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>NMOSIIIB</td>
<td>series</td>
<td>800</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>CMOS26</td>
<td>series</td>
<td>800</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>CMOS26</td>
<td>series</td>
<td>800</td>
<td>900</td>
<td>700</td>
</tr>
</tbody>
</table>
Quest for Performance Improvement

Relative Performance

<table>
<thead>
<tr>
<th>Year</th>
<th>86</th>
<th>87</th>
<th>88</th>
<th>89</th>
<th>90</th>
<th>91</th>
<th>92</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TTL --- NMOS --- CMOS

\[ 20 \% \text{ Improvement} \]

Design Goals

- Price/Performance Improvement
- Scalability (High and low, universal)
- Compatibility (With previous generations, many subsets)
- Time to Market
System Block Diagram

Chip Vital Statistics

Architecture
Frequency
I-Cache
D-Cache
Physical Memory Address
Virtual Memory Address
TLB

PA-Risc v1.2
0–100MHz
4k–1M, hashed virtual indexed
64 bits wide, 32 byte line
4k–2M, hashed virtual indexed
64 bits wide, 32 byte line
32 bits
48 bits segmented
Fully associative, unified
120 page entries, 16 block entries

128–16k Pages
Technology Overview

Technology: CMOS26B 0.8 micron drawn
Transistor count: 850K
Die Size: 14mm x 14mm
Interconnect: 1 Silicide Poly
Interconnect: 3 Levels Metal Aluminum
Power: 5v internal
         3.3v I/O
         20W at 100MHz
Pad Count: 520
Package: 504 pin PGA

CPU Block Diagram

I-Cache Interface

Math Megacells
Floating Point Control
Floating Point Datapath

Integer Datapath

I-Cache Control
Main Control
D-Cache Control

System/MP Control

System Bus

D-Cache Interface
Design Methodology

- Leverage 66+ MHz Design
- Scale to 100MHz with Process shrink
- Incorporate Characterization Data
- Redesign Critical Circuits
- Add Enhanced Features
- Design on Chip Floating Point
- Faster Rams
Starting Feature Set

Integer Unit
- Branch Prediction
- Extensive Bypassing
- Shadow Registers for TLB Miss Traps
- Cache Streaming

Cache and TLB
- Large off Chip I and D cache
- Virtual Index Hashing
- Fully Associative Split I/D
- 96 Page and 4 Block Entries

Feature Set Enhancement

- On Chip Floating Point Unit
- Dual I Fetch for Superscalar
- Cache Miss Optimizations (Stall-on-Use)
- Block Copy Hint
- Fully Associative Unified TLB
- 120 Page and 16 Block entries
- Hardware TLB miss handler
- Low cost 2-way MP functionality
**FPU FUNCTIONS**

**ALU Functions:**
- Sgl/Dbl
- Sgl/Dbl
- Sgl/Dbl
  Add/Subtract
  Compare/Complement
  Convert
  (float -> int)
  (int -> float)
  (float -> float)

**MPY Functions:**
- Sgl/Dbl 32-bit
  Multiply
  Integer Multiply

**Div Functions:**
- Sgl/Dbl
- Sg./Dbl
  Divide
  Square Root

**FLOATING POINT BLOCK DIAGRAM**

![Floating Point Block Diagram]
Floating Point Instruction Timing

Latency/Dispatch (Cycles)

<table>
<thead>
<tr>
<th></th>
<th>Single Precision</th>
<th>Double Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALU</td>
<td>2/1</td>
<td>2/1</td>
</tr>
<tr>
<td>Multiply</td>
<td>2/1</td>
<td>2/1</td>
</tr>
<tr>
<td>MPY/ALU</td>
<td>2/1</td>
<td>2/1</td>
</tr>
<tr>
<td>Divide</td>
<td>8/8</td>
<td>15/15</td>
</tr>
<tr>
<td>Square Root</td>
<td>8/8</td>
<td>15/15</td>
</tr>
</tbody>
</table>

200 Mflop Peak Performance at 100MHz (MPY/ALU)
Superscalar Execution with an Integer Instruction

MPY ORGANIZATION

OpA OpB Mpy ALU
Op Select/Align
Booth Encode
14:2 CARRY SAVE
PS PC
Stage 0

15:2 CARRY SAVE
PS PC
Stage 1

SUM ROUND
PS PC I/O
Stage 2

F₀
Stage 3
Superscalar Execution

Instruction Pairs:  A – E
     E – A
     E – B

A: Integer Operation
   Integer Load or Store
   Floating Point Load or Store

B: Branch

E: Floating Point Operation
   – No Alignment Constraints

Cache Miss Optimizations

○ Stall-on-use (Hit Under Miss)

○ Hides Miss Penalty

○ Implemented for loads and stores

○ Will Stall only for register dependency or another Miss

○ Implemented "don't fill" cache hint
Hardware TLB Miss Handler

○ Invoked on I/D translations that miss the on chip TLB

○ Hardware Computes the PDIR entry address

○ If the entry is present and valid, it is inserted on chip

○ If it is not present or valid a software trap is taken

○ Page entries can be updated in 11 states

Multiprocessor Configurations

○ Full compatibility will all previous MP implementations

○ Scalable High End

○ Low cost 2-way Implementation
Scalable MP Configuration

Low Cost Dual Processor
2-Way Implementation

- No Memory support or changes required
- Minimal Implementation Specific OS Impact
- All MP functionality is contained on the CPU
- Cache and TLB coherency over local Memory-I/O bus (P-bus)
- Logic handles 3 freq ratios between the processor and the external P-bus

Benchmark Performance

This slide will share the performance of key Benchmarks

SPCE 69 > 130
SPEC92 > 70
FP52 > 130
rFlops (uniproc) = 230
perl-rflop = 200
mips = 115
CERN Benchmark < 14 sec
Wilton 30 ≈ 160
with STW OP ≈ 150
Growth Paths

- Process Improvements
- Faster Cache Chip Technology
- Circuit Enhancement
- More Compiler use of New Features

Summary

- Scalable PA-RISC processor for Workstation and Multiuser
- Continuation of HP Processor Family
- Single Chip Solution
- Frequency as well as feature enhancements
- On Chip Floating Point