AMD 3DNow!™ Technology and the K6-2 Microprocessor

Stuart Oberman
Fred Weber
Norbert Juffa
Greg Favor

California Microprocessor Division
Advanced Micro Devices
OUTLINE

• Motivation for 3DNow! Technology
• Features of the 3DNow! instruction set
• AMD-K6-2 implementation
• 3D graphics performance
• Future
Acceleration of Multimedia Applications

• Multimedia applications have become an integral part of the PC platform
  – Multimedia algorithms are computationally intensive

Application or Game Physics  Geometry transform, Clipping, & Lighting  Triangle Set up  Pixel Rendering

floating point intensive  floating point intensive  floating point & integer intensive  integer intensive
Motivation for 3DNow! Technology

• Why a new technology?
  – Previous focus has been on integer intensive pixel rendering tasks: MMX and 3D graphics hardware
  – 3D graphics performance now limited by floating-point intensive front-end of graphics pipeline
  – New applications require realistic physical modeling

• What is it?
  – New set of instructions to accelerate FP computation
  – Defined in collaboration with leading ISV’s
  – Maximizes the performance of graphics accelerator cards by improving the front-end of graphics pipeline
3DNow! Technology

• Benefits
  – Accelerates most floating-point intensive multimedia operations
    • Graphics pipeline (physics, geometry, setup)
    • Audio processing
3DNow! Technology

• 21 new instructions
  – “LEAN and MEAN” design philosophy
  – Includes only performance critical features

• SIMD floating-point instructions
  – Compatible with IEEE single precision data type
  – Two 32-bit FP values per 64-bit reg/mem operand
  – Uses MMX registers -> avoids x87 register stack
  – No exceptions
  – Limited rounding modes
  – No switching overhead between 3DNow! and MMX
  – Peak throughput of 4 FLOPS per cycle

• No core OS support required
Classes of Instructions - FP

- **Basic arithmetic**
  - PFADD, PFSUB, PFSUBR, PFACC, PFMUL

- **Comparisons**
  - PFCMPEQ, PFCMPGT, PFCMPGE

- **Min/max**
  - PFMIN, PFMAX

- **Conversions**
  - PF2ID, PI2FD

- **Reciprocal and reciprocal square root**
  - PFRCPP, PFRSQRT
  - PFRCPI1, PFRSQI1, PFRCPI2
Classes of Instructions - Non FP

- **Integer**
  - PAVGUSB, PMULHRW

- **Data movement**
  - PREFETCH/PREFETCHW

- **Overhead reduction**
  - FEMMS
Reciprocal and Reciprocal Square Root

- Alternative to “classical” DIV and SQRT
  - Reciprocal and reciprocal square root frequently used in graphics applications
  - Higher performance through reuse of common divisors and radicands

- Choice of reduced (14-15b) or full precision
  - Reduced precision sufficient for many applications and is higher performance
  - Avoid all long latency operations; full precision synthesized from fully-pipelined Newton-Raphson iterations ops
Reciprocal Iteration Instructions

• Reciprocal Newton-Raphson iteration
  – To compute full-precision reciprocal of b using initial approximation $R_0$:
    • $R_{full} = R_0 \times (2 - b \times R_0)$
  – $R_0$ is accurate to about 14 bits, b is a 24 bit number
  – PFRCPIT1 performs $b \times R_0$ rounded to 32 bits, inverts the result (one’s complement), and compresses out leading 8 bits known to be identical, leaving 24 bits
  – PFRCPIT2 expands the previous result to 32 bits, multiplies by $R_0$, adds a fixed bias, and rounds to 24 bits
Reciprocal Accuracy

- Fast approximations
  - PFRCP accurate to 14.9 bits
  - PFRSQRT accurate to 15.8 bits

- Full precision
  - PFRCP, PFRCPIT1, PFRCPIT2 sequence provides IEEE RN result for > 99% of all operands; remaining differ by 1 unit-in-the-last-place
  - PFRSQRT, PFMUL, PFRSQIT1, PFRCPIT2 sequence provides IEEE RN result for > 87% of all operands; remaining differ by 1 unit-in-the-last-place
AMD-K6-2 Microprocessor

- Worldwide launch May 28, 1998
- Implemented in 0.25um CMOS process
- 9.3M transistors on a die of 80 mm²
- New features of AMD-K6-2
  - **Superscalar 3DNow! and MMX units**
    - Dual decode and dual execution pipelines
    - No decode pairing restrictions
    - Only one cycle misalignment penalty on memory accesses
  - **100 MHz Front Side bus**
    - Increases local bus and L2 cache bandwidth by 50%
    - Redesigned I/O timing to allow for low cost 100 MHz motherboard
AMD-K6-2 Multimedia Units

Register X Execution Pipeline

MMX ALU

Shared Resources

FP Add
Recip / RecipSqrt

MMX+FP Multiplier

MMX Shifter

Register Y Execution Pipeline

MMX ALU
## AMD-K6-2 Multimedia Performance

<table>
<thead>
<tr>
<th>Instruction Type</th>
<th>Latency (cycles)</th>
<th>Throughput (cycles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DNow! FP</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3DNow! / MMX integer ALU</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MMX multiply</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K6-2 Performance</td>
<td>PII Performance</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>14 bit reciprocal</td>
<td>2 cycles</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pipelined</td>
<td></td>
</tr>
<tr>
<td>15 bit reciprocal</td>
<td>2 cycles</td>
<td>-</td>
</tr>
<tr>
<td>square root</td>
<td>pipelined</td>
<td></td>
</tr>
<tr>
<td>24 bit reciprocal</td>
<td>6 cycles</td>
<td>~ 17 cycles</td>
</tr>
<tr>
<td></td>
<td>pipelined</td>
<td>non-pipelined</td>
</tr>
<tr>
<td>24 bit reciprocal</td>
<td>8 cycles</td>
<td>~ 28 cycles</td>
</tr>
<tr>
<td>square root</td>
<td>pipelined</td>
<td>non-pipelined</td>
</tr>
</tbody>
</table>
Shared 3DNow! and MMX Multiplier

Data Format Selection

Booth 2 Encoders

Booth Muxes

EX1

Binary Tree

4-2 CSA
4-2 CSA
4-2 CSA
3,2 CSA

17 PP

64 bits

4-2 CSA
3,2 CSA
3,2 CSA

MUX

EX2

64b CPA
64b CPA

Final Result Selection (32b)

MMX Data Alignment

A_H x B_H

Zeroed out

16b

16b

Sign

Extended

A_L x B_L
## 3D Winbench 98 Performance

### 3D Winbench 98 / Windows 95 (DirectX 6.0 optimized for 3DNow!™)

<table>
<thead>
<tr>
<th>Processor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD-K6-2/300</td>
<td>1110</td>
</tr>
<tr>
<td>Pentium II 300</td>
<td>961</td>
</tr>
<tr>
<td>Pentium II 350</td>
<td>1060</td>
</tr>
<tr>
<td>Pentium II 400</td>
<td>1120</td>
</tr>
</tbody>
</table>

Windows 95 OSR 2.1, 32MB DRAM, Maxtor DiamondMax IDE HD, 512K L2 cache, Diamond Viper V330 4MB AGP. AMD-K6 3D processor based system: Microstar 5169 mainboard supporting 100MHz bus. Pentium® II 300 based system: Abit LX6 mainboard supporting 66MHz bus (Pentium II 300 based systems with 100MHz bus not currently commercially available). Pentium II 350 and Pentium II 400 based systems: Asus P2B mainboard supporting 100MHz bus.
Quake 2 Performance

Highest Performance without Sound
Timedemo: DEMO1.DM2

Average Frames Per Second (FPS)

Display Resolution (horizontal x vertical)

- AMD-K6-2, Dual Voodoo2
- AMD-K6-2, Single Voodoo2
- Intel Pentium-II, Dual Voodoo2
- Intel Pentium-II, Single Voodoo2