AMD 3DNow!™ Technology: Architecture and Implementations

Stuart F. Oberman
e-mail: stuart.oberman@amd.com

California Microprocessor Division
Advanced Micro Devices
OUTLINE

• Motivation for 3DNow! Technology
• Features of the 3DNow! instruction set
• K6-2 and K7 implementations
• 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

-浮点运算密集型
-浮点运算密集型
-浮点与整数运算密集型
-整数运算密集型

AMD 64-bit 3DNow!
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

Graphic Cards Accelerates
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, PFCMPPGT, PFCMPPGE

• Min/max
  – PFMIN, PFMAX

• Conversions
  – PF2ID, PI2FD

• Reciprocal and reciprocal square root
  – PFRCP, PFRSQRT
  – PFRCPIT1, PFRSQIT1, PFRCPIT2
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 $R_0$:
    - $R_{\text{full}} = R_0 \times (2 - b \times R_0)$
  - $R_0$ 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
K6 / K6-2 Microarchitectural Features

- Out-of-order superscalar pipeline
  - Fetch and decode up to 4 internal ops per cycle
  - Issue up to 6 ops per cycle to 10 functional units
  - Loads, stores, integer ops, FPU, MMX/3DNow! Ops can be performed out-of-order with respect to each other

- Pipeline is 6 to 8 stages deep
  
  IF | DEC | SC | RF | EX1 | EX2* | EX3* | RET |

- L1 Cache -- 32KByte I and 32KByte D write-back
- Socket 7 processor bus at 66 MHz (K6) or 100 MHz (K6-2)
AMD-K6-2 Multimedia Units

Register X Execution Pipeline

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 pipelined</td>
<td>-</td>
</tr>
<tr>
<td>15 bit reciprocal</td>
<td>2 cycles pipelined</td>
<td>-</td>
</tr>
<tr>
<td>square root</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 bit reciprocal</td>
<td>6 cycles pipelined</td>
<td>~ 17 cycles non-pipelined</td>
</tr>
<tr>
<td>square root</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 bit reciprocal</td>
<td>8 cycles pipelined</td>
<td>~ 28 cycles non-pipelined</td>
</tr>
<tr>
<td>square root</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Shared 3DNow! and MMX Multiplier

Data Format Selection
32 bits

Booth 2 Encoders

Booth Muxes
17 PP

EX1
Binary Tree

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

64 bits

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

MUX

EX2

64b CPA
64b CPA

Final Result Selection (32b)

MMX Data Alignment

Sign
Extended

Zeroed
out

A_h x B_h

16b

16b
3D Winbench 98 Performance

Benchmark System Configuration:
Microsoft® Windows® 95 OSR 2.1, 32MB DRAM, Maxtor DiamondMax IDE HD, 512K L2 cache, Diamond Viper V330 4MB AGP, FAT32, AMD-K6®-2 processor-based system: Microstar 5169 motherboard supporting 100-MHz bus, Pentium® II 300 & 333 processor-based systems: Able LXG motherboard supporting 66-MHz bus (Pentium II 300 & 333 processor-based systems with 100-MHz bus not currently commercially available), Pentium II 350 processor-based systems: Asus P2B motherboard supporting 100-MHz bus.

† AMD-K6-2 processor-based systems enabled with a pre-release Nvidia driver supporting 3DNow!™ technology of AMD-K6-2. Pentium II processor-based systems enabled with Nvidia driver ver. 0236 (optimally configured for 3D WinBench™ 98). Performance results may vary with final production versions of these software components. Nvidia and other leading graphics vendors update drivers regularly to take advantage of new developments such as 3DNow!™ technology.
## Quake II Platform Configurations

<table>
<thead>
<tr>
<th>K6-2 300/100</th>
<th>PII 300/66</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OS:</strong></td>
<td>OS: Win95</td>
</tr>
<tr>
<td><strong>Motherboard:</strong></td>
<td>Motherboard: Asus PB2 (BX)</td>
</tr>
<tr>
<td><strong>BIOS Date:</strong></td>
<td>BIOS Date: 4/2/98</td>
</tr>
<tr>
<td><strong>Memory:</strong></td>
<td>Memory: 64MB</td>
</tr>
<tr>
<td><strong>2D Video:</strong></td>
<td>2D Video: Diamond Viper 330 (RIVA 128)</td>
</tr>
<tr>
<td><strong>3D Video:</strong></td>
<td>3D Video: Diamond VD2 (12MB)</td>
</tr>
<tr>
<td><strong>HDD:</strong></td>
<td>HDD: Maxtor DM 4.3GB</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Motherboard:** Epox 51MVP3E-M

**BIOS Date:** 6/1/98

**Memory:** 64MB

**2D Video:** Diamond Viper 330 (RIVA 128)

**3D Video:** Diamond VD2 (12MB)

**HDD:** Maxtor DM 4.3GB
Quake II Performance - DEMO1

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

640x480 800x600 1024x768
Quake II Performance - MASSIVE1

- Highest Performance without Sound
- Timedemo: MASSIVE1.DM2

<table>
<thead>
<tr>
<th>Display Resolution</th>
<th>AMD-K6-2, Dual Voodoo2</th>
<th>AMD-K6-2, Single Voodoo2</th>
<th>Intel Pentium-II, Dual Voodoo2</th>
<th>Intel Pentium-II, Single Voodoo2</th>
</tr>
</thead>
<tbody>
<tr>
<td>640x480</td>
<td>57.5</td>
<td>56.9</td>
<td>53</td>
<td>48.4</td>
</tr>
<tr>
<td>800x600</td>
<td>57.3</td>
<td>51.4</td>
<td>51.3</td>
<td>48.4</td>
</tr>
<tr>
<td>1024x768</td>
<td>53.3</td>
<td>51.3</td>
<td>51.3</td>
<td>48.7</td>
</tr>
</tbody>
</table>

Average Frames Per Second (FPS)
K6 Family and Future

AMD-K6
- .25 µ process
- Lower Power
- Higher Speeds
- 8.8M Transistors
- 68 mm²
- MMX Enhanced

AMD-K6-2
- .25 µ process
- Bus 100 MHz
- 100 MHz Frontside L2
- Superscalar MMX™
- 9.3M Transistors
- 80 mm²

AMD-K6-3
- .25 µ process
- 100 MHz Bus
- On-chip, Processor
- Speed Backside L2
- 100 MHz Frontside L3
- Superscalar MMX
- 3DNow! Technology
- 9.3M Transistors
- 80 mm²

µP Forum ‘98
October 13
ISSCC ‘99

K7

1H’97 2H’97 1H’98 1Q’99 2Q’99
K7 Microarchitecture
K7 Floating Point and Multimedia Unit
3DNow! Floating Point Datatype

(32 bits x 2) Two packed, single-precision, floating-point doublewords

<table>
<thead>
<tr>
<th>63</th>
<th>32</th>
<th>31</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D0</td>
<td></td>
</tr>
</tbody>
</table>

D0 and D1 each hold an IEEE 32-bit single precision value:

<table>
<thead>
<tr>
<th>31</th>
<th>23</th>
<th>22</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Biased Exponent</td>
<td>Significand</td>
<td></td>
</tr>
</tbody>
</table>