SOC Design: SOC for Graphics

Lan-Da Van (范倫達), Ph. D.
Department of Computer Science
National Chiao Tung University
Taiwan, R.O.C.

Fall, 2018
Source from:

An SoC With 1.3 Gtexels/s 3-D Graphics Full Pipeline for Consumer Applications

Outline

• Introduction
• SOC Architecture
• User Programmable Geometry Engine
• Memory Bandwidth Efficient Rasterization Engine
• Implementation and Comparison Result
• Conclusion
Introduction

• Very high texel rate SoC design
  – 166 MHz full pipeline engine with performance of 33 Mvertices/s and 1.3 Gtexels/s
  – ARM11 RISC processor
  – Video IP
  – Chip in 0.13 um CMOS technology with 7.1x7.0 mm²

• Vertex cache mechanism
  – Reduce the memory bandwidth
SOC Architecture (1/2)
SOC Architecture (2/2)

• Geometry Engine
  – Floating-point engine that supports full programmable vertex and triangle processing

• Rasterization Engine
  – Design to reduce memory bandwidth

• 3D graphics IP has one slave bus interface to receive rendering commands from the RISC and seven bus master interfaces to fetch vertices and texture images or to access depth and frame buffers.

• ARM11 plays a role of stand-alone processor and a video decoder.

• Mixer composes the video data from video processor (VP) and the graphics data rendered by the 3D graphics IP with configurable options.
Geometry Engine Architecture (1/5)
Geometry Engine Architecture (2/5)

• GE architecture
  – A programmable vertex shading engine (VSE)
  – Hardwired primitive assembly (PA) logic
  – A fixed function clipping engine (CE)
  – A programmable triangle setup engine (TSE)

• GE function
  – Vertex transform and lighting (TnL)
  – Clipping
  – Triangle Setup

• PA function
  – The transformed and lit vertices are gathered in the primitive assembly logic to form triangles.
Geometry Engine Architecture (3/5)

• CE function
  – After primitive assembly, the CE performs view-frustum clipping, perspective division and view-port mapping on a triangle stream.
  – Eliminate the triangle parts outside the view-frustum
  – Regenerate the clipped polygon to triangles

• TSE function
  – Transform the triangles into the data required in the hardware rasterizer
  – User defined microcode
  – Triangle-level programmability
Geometry Engine Architecture (4/5)

- **SIMD Unit**
  - Basic Arithmetic Instructions:
    - ADD
    - MUL
  - Comparison instructions
    - MIN, MAX, SGE
  - Inner product instructions:
    - DP4 (dot-product of four-dimensional vectors)
    - DP3 (dot-product of three-dimensional vectors)

- **Special Function Unit**
  - RCP, RSQ
  - LOG, EXP (for base 2)
Geometry Engine Architecture (5/5)

• Latency Issue
  – One clock cycle for a floating-point multiplier
  – Two clock cycles for a two-input floating-point adder
  – Three clock cycles for a four-input floating-point adder cascaded after four parallel multipliers with one throughput
  – Four clock cycles to pass through the multiplier and the four-input adder
  – Seven clock cycles for the transcendental functions and division in the SFU.
    • The coefficients are stored in a 512 byte lookup table per each function for sufficient precision.
Special Function Unit Architecture

- Seven sub-stages
- Each sub-stage can write back to the VOB and register file. => variable write-back latency
- Hazard controller with NOP instruction to the pipeline
Hazard Controller of VSE
Hazard Controller of VSE

- The registers are composed of three fields:
  - 4-bit tag works as a life-time counter and decreases one cycle each cycle time.
  - The destination address of the register file is used for testing RAW hazard between the incoming source addresses and the in-pipeline destination addresses.
  - Destination mask information.
- Data field: quad-float vectors (i.e., 128 bits)
• Add programmability into the setup computation for triangle-level effect
• The 3x3 matrix multiplications are used for derivatives computation.
• The instruction set of the TSE is similar to that of the VSE.
  – Special instructions for setup process
  – DP3
  – PCT: Quantize the floating-point vertex position to the pixel center for sampling pixel data at the pixel center.
• Four pipeline stages in the TSE:
  – Instruction fetch, instruction decoding/operand fetch, execution and write-back.
Performance of VSE and TSE (1/2)

Mvertices/sec for VSE
Mtriangles/sec for TSE

- Transform only: 33.2
- Minimum triangle setup: 13.83
- Single specular lighting: 7.55
- Spot lighting: 4.74
- Anisotropic lighting: 3.95
- Setup for color and depth: 2.96
- Setup for a single texture: 3.02
- 2.18 setup for double textures
Performance of VSE and TSE (2/2)

- The CE consumes about 30 cycles to clip triangles per clip plane.
- The proportion of clipped triangles to total triangles is about 10% for conventional games [18].
- The VSE and TSE perform user-defined operations on vertices and triangles.
- Determine the throughput of the TSE to be almost 70% of the VSE considering the experimental ratio of vertex per triangle and fully clipped out triangle.
- The above figure shows that the performance ratio is closed to 70% as the algorithm complexity of both engine increases.
Rasterization Engine Architecture
Rasterization Engine

- RE architecture
  - Rasterizer
  - Texture Engine (TE)
  - Per-pixel processing engine (PPE)

- Rasterizer function
  - Interpolate all the data of the pixels: screen coordinate of pixels (x,y), color (r,g,b,a), depth z, fog factor f, and two pairs of 2-D texture coordinates (u1, v1), (u2, v2)

- TE function
  - Process the texture mapping on four pixels simultaneously and support two-level multi-texturing and trilinear MIPMAP texture mapping

- PPE function
  - Perform fog-effects, color blending, and alpha, depth, and stencil tests
Concept of the Depth Filter (1/2)

Triangle A marked on the Depth Filter

Triangle A which is rendered previously

Eye

Depth Filter Plane

This pixel is passed to the next pipeline. (not rejected by the Depth Filter)

Triangle B which is under rendering

This pixel is rejected by the Depth Filter

B
Concept of the Depth Filter (2/2)

- Depth filter plane
- One the z axis in the screen space
- Including the overhead of external memory accesses by cache misses, the Depth Filter reduces the total external memory accesses of the graphics pipeline by 58.7%
Cache Architecture of the Depth Filter
Memory Bandwidth Reduction by the Techniques in 3-D Graphics IP

Required memory bandwidth (MB/sec)

- Texel-fetch
- Color-write
- Color-read
- Depth-write
- Depth-read
- Vertex-fetch
- DF overhead

Totally 85.2% reduction

Provided memory bandwidth: 1.33GB/s

280MB/s used by ARM1136JF and video IPs

1.05GB/s used by 3-D graphics IP
Comparison of Memory Bandwidth Reduction

Required memory bandwidth (MB/sec)

25
### TABLE 1
**CHIP SPECIFICATION AND CHARACTERISTICS**

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Technology</td>
<td>0.13 um CMOS 7-Metal</td>
</tr>
<tr>
<td>Supply Voltage</td>
<td>1.2V</td>
</tr>
<tr>
<td>Operating Frequency</td>
<td>166Mhz for IPs and 333Mhz for RISC Core</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>407 mW for fully activated 3-D Graphics IP</td>
</tr>
<tr>
<td>Transistor Counts</td>
<td>17.9M transistors</td>
</tr>
<tr>
<td>Die Size</td>
<td>7.1x7.0 mm2</td>
</tr>
<tr>
<td>Package</td>
<td>256 pin QFP</td>
</tr>
<tr>
<td>3-D Performance</td>
<td>33Mvertices/sec for GE, 1.3Gtexels/sec for RE</td>
</tr>
<tr>
<td>3-D Function Feature</td>
<td>All OpenGL ES operations</td>
</tr>
<tr>
<td></td>
<td>Programmable vertex and triangle processing</td>
</tr>
<tr>
<td></td>
<td>2-level multi-texturing</td>
</tr>
<tr>
<td>Video Interface</td>
<td>BT.656 video input / output</td>
</tr>
<tr>
<td></td>
<td>TFT-LCD output</td>
</tr>
<tr>
<td></td>
<td>Component output</td>
</tr>
<tr>
<td></td>
<td>NTSC/PAL, S-Video Composite output</td>
</tr>
</tbody>
</table>
Conclusion

• Propose a 3-D graphics SOC for consumer electronics
• Full pipeline engine with performance up to 33 Mvertices/s and 1.3 Gtexels/s.
• Programmable geometry engine (due to VSE and TSE) and low-bandwidth rasterization engine.
• Bandwidth Reduction with 85.2%:
  – An a-index two-level texture cache
  – A depth filter