An 8.6mW 12.5Mvertices/s 800MOPS 8.91mm^2 Stream Processor Core for Mobile Graphics and Video Applications

You-Ming Tsao, Chin-Hsiang Chang, Yu-Cheng Lin, Shao-Yi Chien, and Liang-Gee Chen

Graduate Institute of Electronics Engineering and Department of Electrical Engineering,
National Taiwan University, Taipei, Taiwan

Abstract

An 8.6mW stream processor core for mobile applications is implemented with 8.91mm^2 area in 0.18um CMOS technology at 50MHz. The adaptive multi-thread architecture with configurable memory array and geometry-content-aware technique are proposed to reduce power consumption while achieving 12.5Mvertices/s for 3D graphics and motion estimation with search range {H[-24,24),V[-16,16]} for CIF (352x288) 30fps video encoding. (Keywords: stream processor, vertex shader, low power GPU, adaptive multi-thread, configurable memory array and early-rejection-after-transformation)

Introduction

Embedded multimedia functions of 3-D graphics and video coding become a promising trend in mobile devices. It is required to design a low-power and low-cost vertex processor with enough processing speed [1][2][3]. From the system point of view, if the graphics and video coding engines can be integrated as a single processor while sharing the hardware resource, not only the chip area can be reduced, but also the hardware utilization can be further improved.

In this work, we have developed a stream processor [4] for both graphics and video coding applications with low power consumption and high hardware utilization. This design is the first reported chip using a unified architecture to support full Vertex Shader 3.0 model [5] and video encoding features. For this chip, three key techniques, adaptive multi-thread (AMT) architecture, configurable memory array (CMA) and early-rejection-after-transformation (ERAT), are proposed to achieve low power consumption, high performance and high efficiency. It achieves the processing speed of 12.5Mvertices/s for graphics applications and supports full search (FS) motion estimation (ME) for CIF (352x288) 30fps video coding.

Processor Architecture

Fig. 1 shows the proposed hardware architecture. It is based on 2-issue VLIW architecture with SIMD instruction in each slot. When operated in 50MHz, it achieves the performance of 400MFLOPS with two 4-channel floating-point operations executed simultaneously for transforming 12.5M vertices/s and the floating-point datapath can be configured to fixed-point datapth to provide the processing capability of 800MOPS for the proposed video encoding acceleration instruction sets. The CMA can be reconfigured as different kinds of register files in the stream processor. In order to support various power optimization techniques in graphics application programs, both the data forwarding and the clock gating circuits for each pipeline stage and processing element can be both controlled by instructions.

Adaptive Multi-Thread

AMT with data forwarding reduces data hazard conditions to improve the performance with fewer pipeline bubbles, and

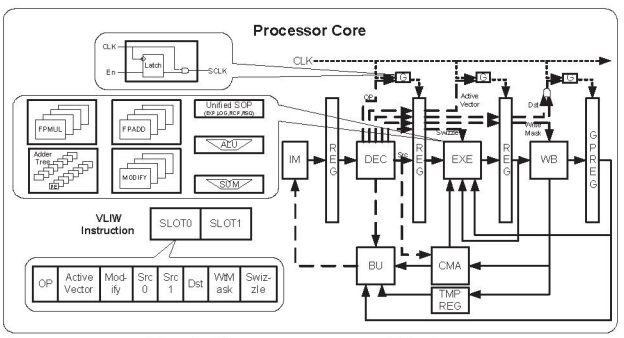

Fig. 1. Processor core architecture.

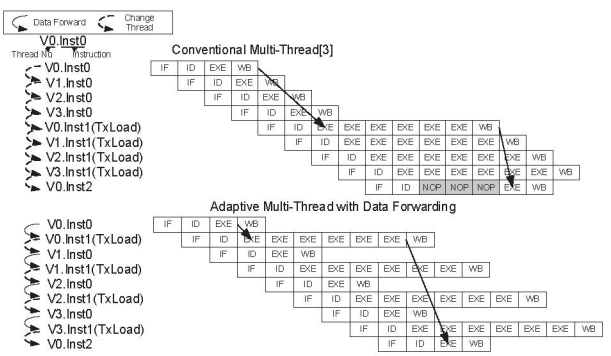

Fig. 2. Adaptive multi-thread schedule.

alleviate the data access of the register files form the datapath to reduce the power consumption. The developed AMT technique efficiently uses the minimum threads to provide the maximum hidden latency ability. It is shown in Fig. 2 that when comparing the AMT technique with the conventional multi-thread techniques [3], fewer pipeline bubbles are introduced and the efficiency of the processor is increased.

Configurable Memory Array

Memory bandwidth plays an important factor in power limited mobile devices. Our proposed stream processor adopts both the cache and tightly couple memory mechanisms to decrease the required memory bandwidth for different applications. It can achieve almost 60% cache hit rate and save the bandwidth and processing power of duplicate vertices. When performing ME, our architecture supports level-C data reuse scheme [6] to save 86% memory bandwidth. As shown in Fig. 3(a), the CMA with 4-channel and 8-bank accessing ability serves as a physical on-chip memory pool, which can be logically configured for different applications to achieve high memory utilization. Compare with the dedicated data buffer architecture [2][3], the CMA provide the more flexibility and efficiency. Fig. 3(b) shows the CMA configuration of the stream cache and the constant register files for vertex processing, and Fig. 3(c) shows the configuration of CMA for ME, where the reference data is

loaded in to the constant register files and the current macro block data is treated as the input stream.

Early-Rejection-After-Transformation

In the graphic pipeline, the vertex processor performs shading operations on every vertex. After sending the vertices to the rendering stage, many primitives will be found to be invisible on the screen by the render processor, which implies that a lot of processing power has been wasted in the vertex processor. A geometry-content-aware technique called early-rejection-after-transformation (ERAT) is developed to reduce the power consumption and increase the performance by rejecting redundant triangles after the transform stage. A dedicated module is designed to detect and reject the redundant triangles belonging to the three types shown in Fig. 4. The work is also the first reported vertex processor embedded with the content-aware technique, which could reduce the power at least 20% when performing realistic lighting programs.

Implementation

This work is implemented using TSMC 0.18um 1P6M process with 8.91mm^2 at 50 MHz. The chip features and micrograph are shown in Fig. 5. The measured power consumption is 8.6mW. Fig. 6(a) shows the results of power reduction when perform phong shading model. It shows that when all the three proposed key techniques are employed, 86% of the power consumption of the VLIW SIMD architecture with gated clock can be reduced. Performance index of Mvertices/s per mW is used to evaluation the power efficiency using the peak performance and power, and it shows that 1.82 times improvement can be achieved when compared with the state-of-the-art vertex processor [3] in Fig. 6(b).

Conclusion

This paper presents an 8.6mW stream processor core with the processing capability of 12.5Mvertices/s for graphics applications, and it can support ME for CIF 30fps video coding. The core area is only 8.91mm^2 to achieve both the demands of graphics and video encoding functions for mobile multimedia applications.

Acknowledgments

The authors would like to thank Chip Implementation Center (CIC) and SMEDIA corp. for chip fabrication. This work was supported by the National Science Council of Taiwan, R.O.C. under Grant NSC95-2221-E-002-371.

References

[1] F. Arakawa, et al., "An embedded processor core for consumer appliances with 2.8GFLOPS and 36M polygons/s FPU," *ISSCC Dig. Tech. Papers*, pp.334--335, 2004.

[2] J. Sohn, et al., "A 50 Mvertices/s graphics processor with fixed-point programmable vertex shader for mobile applications," *ISSCC Dig. Tech. Papers*, pp.192--193, 2005.

[3] C. Yu, K. Chung, D. Kim, and L. Kim, "A 120Mvertices/s multi-threaded VLIW vertex processor for mobile multimedia applications," *ISSCC Dig. Tech. Papers*, pp.408--409, 2006.

[4] U.J. Kapasi, et al., "Programmable stream processors," *Computer*, pp.54--62, Aug. 2003.

[5] K. Gray, *DirectX 9 Programmable Graphics Pipeline*, Microsoft Press, 2003.

[6] J. Tuan, T. Chang, and C. Jen "On the data reuse and memory bandwidth analysis for full-search block-matching VLSI architecture," *IEEE Trans. Circuits Syst. Video Technol.*, pp.61--72, Jan. 2002.

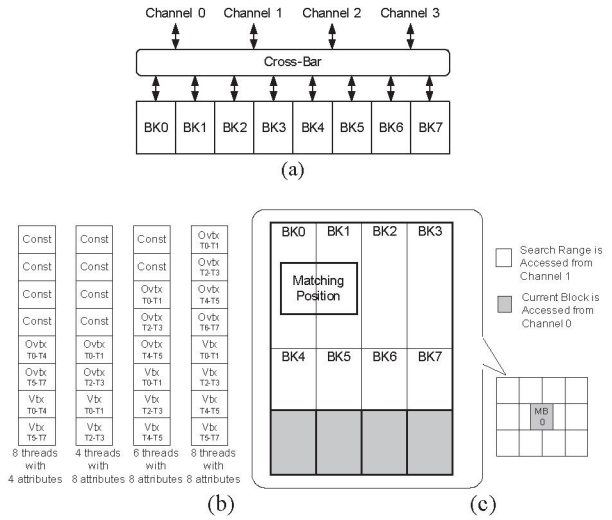

Fig. 3. Data organization in configurable memory array.

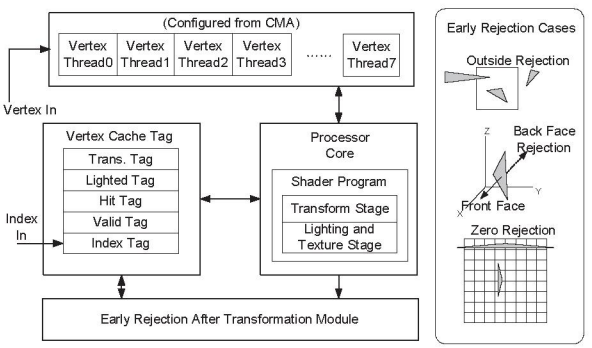

Fig. 4. Block diagram of early rejection after transformation.

Process Technology	TSMC 0.18um CMOS 1P6M
Chip Size	2.7mm x 3.3 mm
Supply Voltage	1.8V
Clock Frequency	50MHz
Power Consumption	(*)8.6 mw
Features	OpenGL ES 2.0 Support Shader Model 3.0 Video encoding IME capability

(*)Shader Program : Specular Light with 20 instructions

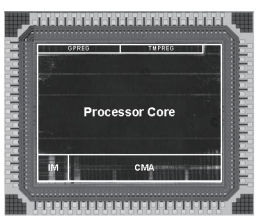

Fig. 5. Chip specification and micrograph.

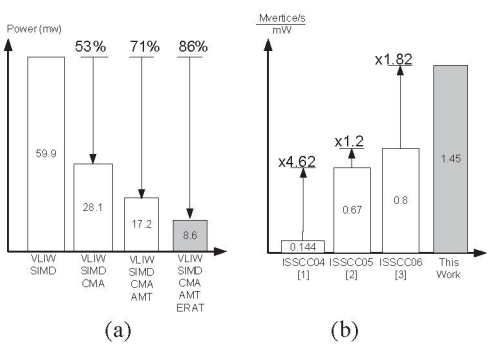

Fig. 6. Power consumption comparison.