9.2 A 80/20MHz 160mW Multimedia Processor integrated with Embedded DRAM MPEG-4 Accelerator and 3D Rendering Engine for Mobile Applications

Chi-Weon Yoon, Ramchan Woo, Jeonghoon Kook, Se-Joong Lee, Kangmin Lee, Young-Don Bae, In-Cheol Park and Hoi-Jun Yoo

Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea

As the number of applications for mobile information terminals grow rapidly, multimedia-processing capability is becoming essential. In addition, power efficiency is required for long battery life and programmability is necessary to adapt functions to various needs [1,2]. This DRAM macro for multimedia applications meets these requirements.

Figure 9.2.1 shows a block diagram of the system. The 80MHz ARM9 compatible RISC has a 5-stage pipeline and includes a 32x32bit MAC unit in its datapath to enhance multimedia processing capability. The MAC unit improves performance up to 23% when executing computation-intensive routines such as DCT/IDCT, compared with the conventional multiplier-only datapath. 0.9CPI and 70MIPS are obtained consuming 0.14mW/MHz. Motion compensation (MC), the most computation- and memory-I/O-intensive part of the overall decoding algorithm, is mapped onto a dedicated hardware to support MPEG-4 video stream@SP QCIF at 15frames/s. Through a 128b internal bus between embedded DRAM frame buffer and logic core, the MC accelerator, composed of 8 processing elements, processes data in parallel at 20MHz. In addition, the integrated DRAM frame buffer, whose structure is tightly coupled with access patterns for MC, eliminates external data I/O and redundant power consumption during data processing. After the RISC pre-processes input polygon data, 3D rendering engine (RE) with Z-compare, smooth-shading, α-blending and double-buffering functions draws a scene with 256x256 resolution at 2.2Mpolygon/s. 3.2GB/s data bandwidth through a 2048b internal memory bus and 640b pixel processor (PP) bus lowers the operating frequency of the 3D RE to 20MHz. To overcome the processing speed and data bandwidth gap between the RISC and the dedicated hardware, a bandwidth equalizer with 2kB dual-port SRAM (32b input, 512b output) is inserted. It accepts data from a 32b bus at 80MHz (ARM to equalizer) and sends them to a 512b-wide bus at 20MHz (equalizer to dedicated hardware). The on-chip DLL synchronizes the 80/20MHz clocks with 10~200MHz operation range.

Figure 9.2.2 shows the architecture of the frame buffer for MC accelerator and its mapping on a screen. Two 128b I/O DRAM macros with 9 banks (512bx128 rows) each are integrated to store previous and current frames. Similar to other graphic applications, tiled mapping is a natural choice since all data processing is performed by block granularity. Row size is optimized to 512b to accommodate one 8x8 block. 9-bank structure is used to minimize cell core activation. To process 1-macroblock (4 blocks), a minimum 4 to a maximum 9 adjacent blocks must be activated. By using distributed nine-tiled block mapping (DNTBM), the 9 adjacent blocks are always placed in different banks, and frequent row changes which consume much power are avoided. In addition, previously activated wordlines in each bank are kept alive until row conflicts occur. Since large spatial localities exist among blocks for successive accesses, DNTBM increases the reusability of previously activated wordlines without penalty, which results in additional power reduction. Although tiled mapping is suited for data processing, it is not appropriate for data transfer to serial access memory (SAM), since only a small part of the block data needs to be transferred at once (Figure 9.2.3). In addition, it is waste power if only a part of the block data is used in motion compensation when the whole wordline is activated. By adopting sub-wordlines with partial core activation, up to 31% reduction of average power per frame is obtained when decoding class A and B streams compared to conventional 1-bank tiled mapping.

3DRE is based on virtually spanning 2D array ViSTA) architecture which contains one edge processor (EP) and 8 pixel processors (PP). It uses only 1/8 the hardware of previous work while maintaining the same screen resolution [1]. As shown in Figure 9.2.4, it virtually spans 2D array through the EP pipeline to render polygons that have 2D spatial locality on screen. The EP calculates the polygon edges from the given vertex and broadcasts them to 8 PP. Then each PP in parallel fills up the internal pixels in the polygon. The memory interface circuit is optimized for read-modify-write transaction. It dynamically reconfigures the 640b PP bus to the 2048b memory bus every cycle while independently controlling the 12x512kb DRAM macros. These DRAM macros consist of a 6Mb frame-buffer and a Z-buffer with 3.2GB/s memory bandwidth. With ViSTA architecture and a wide memory bus, the clock of 3DRE is lowered to 20MHz, with <36mW at 1.5V power supply.

To improve power efficiency of the DRAM macro itself, single bitline write (SBW) is adopted [3]. As shown in Figure 9.2.5, the bitline pair is disconnected from sense amp (SA) after charge-sharing between bitline and cell. Then only the bitline connected to a cell is shorted to SA node to restore cell data and the other bitline (reference bitline) remains disconnected. Therefore, redundant reference bitline transitions in conventional folded-bitline structure are eliminated, resulting in a 20% reduction of power consumption during sensing.

Figure 9.2.6 explains the composition of power consumption in the proposed architecture. High power consumption in I/O transactions is avoided by embedding DRAM on the same chip. Power consumption of RISC, MC accelerator and 3D RE are minimized at 12mW, 4.6mW, and 36mW, respectively, by lowering the operating frequency to 20MHz. Additional power reduction is obtained by adopting low-power techniques such as SBW and DNTBM. The overall system consumes 160mW under the condition that all of the functional units are working.

A chip is implemented using a 0.18μm CMOS EML process with 3 poly and 6 metal layers. A micrograph of the chip is shown in Figure 9.2.7. 1.5V power supply is used for logic core, and 2.5V, 3.3V for DRAM and I/O, respectively. Chip area is 12x7mm², including I/O cells.

Acknowledgements:
The authors thank M.K. Im and. J.H. Lee of Hyundai System IC R&D for chip fabrication. This work was supported by System IC 2010 project of Korea ministry of Science and Technology and ministry of Commerce, Industry and Energy.

References:

Y.H. Park, et al., "A 7.1GB/s Low-Power 3D Rendering Engine in 2D Arrary-Embedded Memory Logic CMOS", ISSCC Digest of Technical Papers, pp. 242~243, Feb., 2000.

T. Nishikawa, et al., "A 60MHz 240mW MPEG-4 Video-Phone LSI with 16Mb Embedded DRAM", ISSCC Digest of Technical Papers, pp. 230~231, Feb., 2000.

J. Kook, et al, "A Low Power Reconfigurable I/O DRAM Macro with Single Bit line Writing Scheme", 26th European Solid-State Circuits Conference, pp.384~387, Sept., 2000.

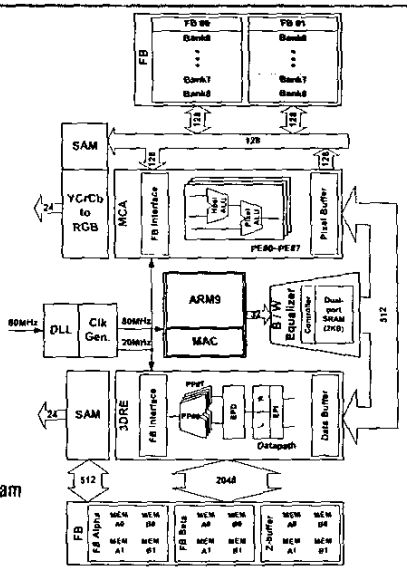


Figure 9.2.1:Block diagram of the proposed heterogeneous architecture.

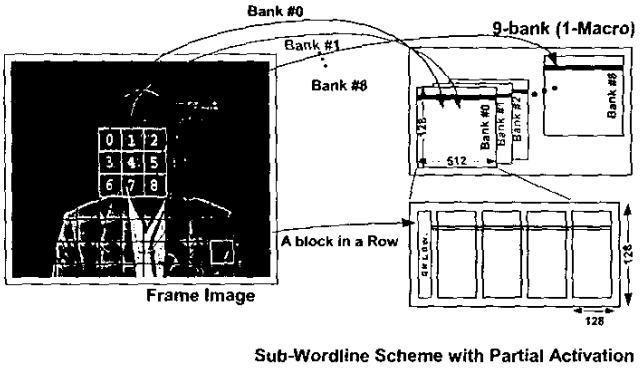


Figure 9.2.2: Distributed 9-tiled block mapping and the architecture of DRAM frame buffer.

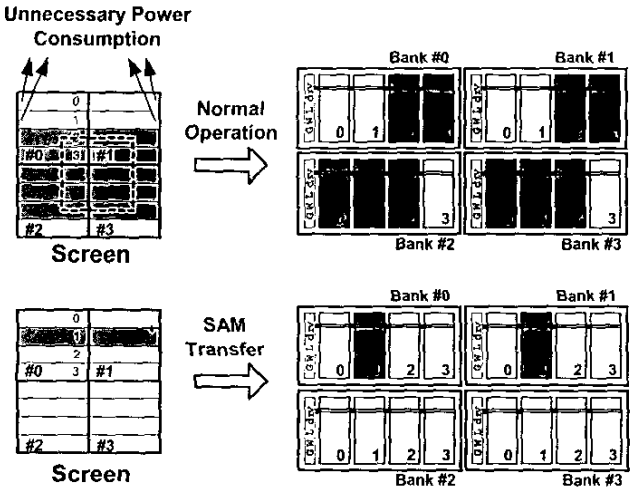
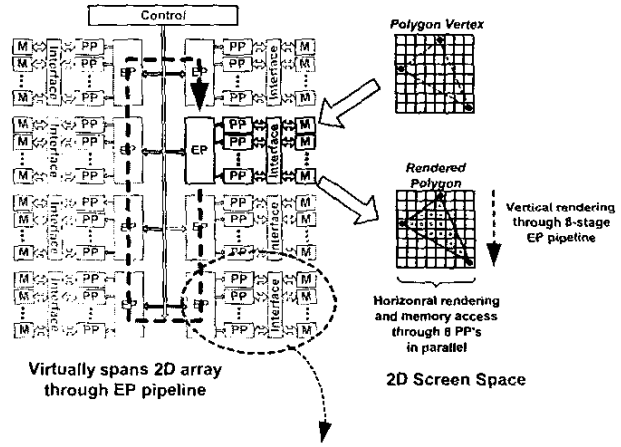


Figure 9.2.3: Partial cell core activation scheme.



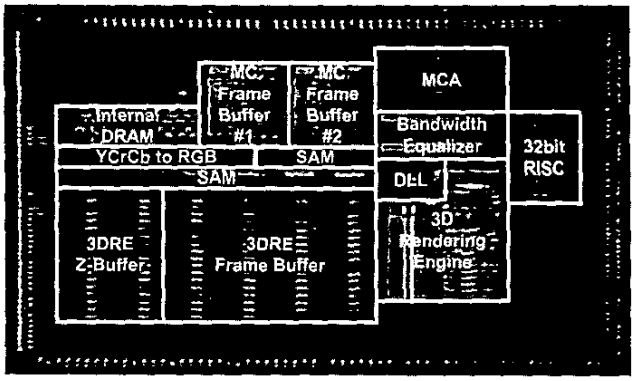


Figure 9.2.7: Die micrograph.

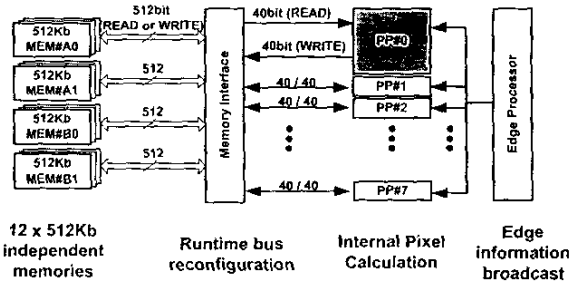


Figure 9.2.4: Virtually spanning 2D array (ViSTA) architecture and its run-time memory bus reconfiguration.

Continued on Page 441

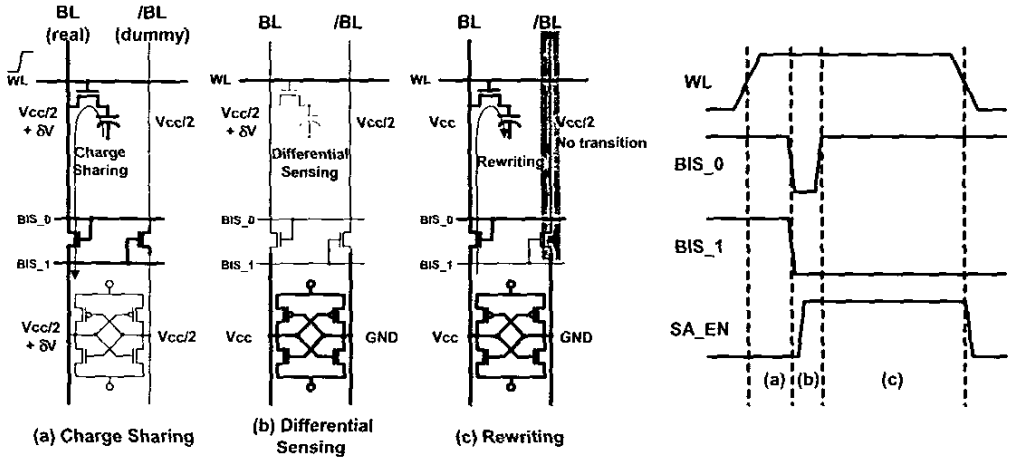
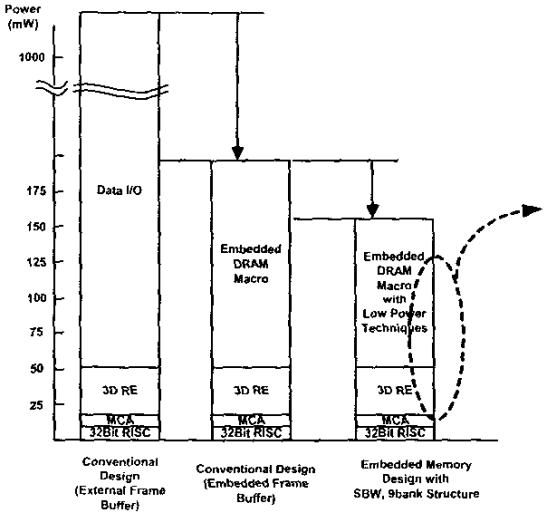


Figure 9.2.5: Operation of single bitline writing (SBW) scheme and its implementation.



	clock (MHz)	Supply (V)	Power (mW)
32Bit RISC	80	1.5	12
MCA	20	1.5	4.6
3D RE	20	1.5	36
FB (MCA)	20	2.5	11.7
FB + Z-buf (3DRE)	80	2.5	84
B. Equalizer	80/20	1.5	4

Figure 9.2.6: Comparison of power consumption of various architectures.

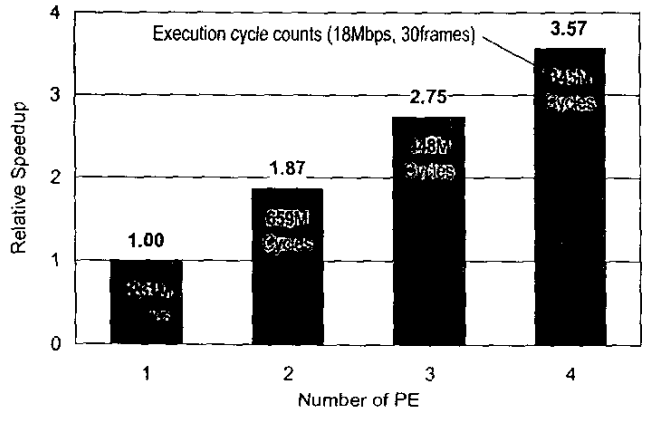


Figure 9.4.7: Performance comparison of MPEG2 (MP@HL) video decoding for one to four PEs.

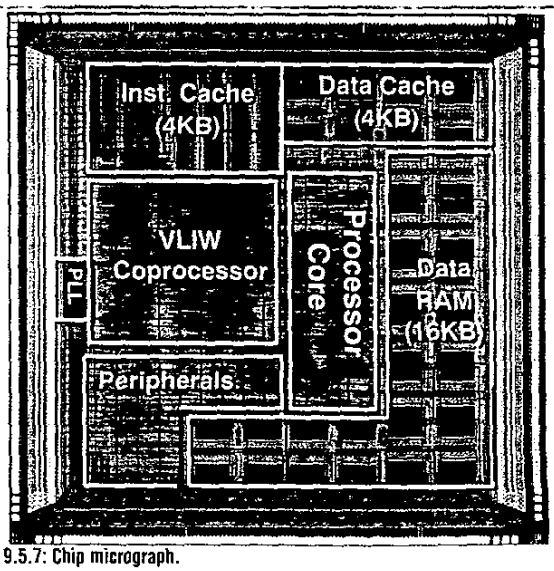


Figure 9.5.7: Chip micrograph.