Extended Abstracts of the 1997 International Conference on Solid State Devices and Materials, Hamamatsu, 1997, pp. 274-275

Invited

PPRAM (Parallel Processing RAM): A Merged-DRAM/Logic System-LSI Architecture

Kazuaki Murakami, Koji Inoue and Hiroshi Miyajima

Department of Computer Science and Communication Engineering, Kyushu University
6-1 Kasuga-Koen, Kasuga, Fukuoka 816 JAPAN
Phone: +81-92-583-7621, Fax: +81-92-583-1338, E-mail: ppram@c.csce.kyushu-u.ac.jp

1. Introduction

Merged DRAM/logic LSIs are rapidly coming to our attention, and are vigorously studied and developed for the following reasons[2]: (1) A large amount of DRAM and logic are now able to be integrated on a same chip in a cost-effective way; (2) Performance gap between processor and DRAM (in terms of bandwidth and latency) is now a severe system-performance bottleneck; (3) Merged DRAM/logic LSIs can meet two conflicting requirements of high performance and low power consumption at a time.

PPRAM, or Parallel Processing RAM, is an architectural framework for such merged memory/logic LSIs and integrates the following onto a single chip[1]: (1) memory (a large amount of DRAM and/or SRAM and/or Flash EEPROM and/or FRAM and/or so on); (2) logic (zero or more general-purpose processor(s) and/or application-specific processor(s) and/or FPGA and/or so on); (3) communication (a network interface based on a common communication protocol).

Many different implementations of *PPRAM* are possible, but they all provide a common network interface (called *PPRAM-Link* [3]). System designers will be able to construct computer/electronic systems of any size, of any functionality, and of any performance, just by choosing the required *PPRAM* chips from those various *PPRAM* implementations and by interconnecting them through the *PPRAM-Link* network.

2. *PPRAM* Solution

PPRAM aims at bringing some paradigm shift shown in Figure 1 to the computer/electronic-system design. *PPRAM* stands on the following three key technologies and will exploit their merits as follows.

- **Merged DRAM/logic LSI** technology will allow us to (1) reduce the power consumption by eliminating the high-capacitance wide bus between DRAMs and microprocessors (MPUs); (2) resolve the memory bottleneck problem by exploiting high on-chip memory bandwidth; (3) improve the memory system performance by utilizing low on-chip DRAM-access latency; (4) optimize the size and organization of on-chip DRAM depending on applications; (5) reduce the off-chip memory-access traffic by utilizing large on-chip DRAM; (6) relieve the bandwidth requirement for inter-chip communication as well; (7) relieve the EMI problem caused by the high-speed wide bus between DRAMs and MPUs, and so on.

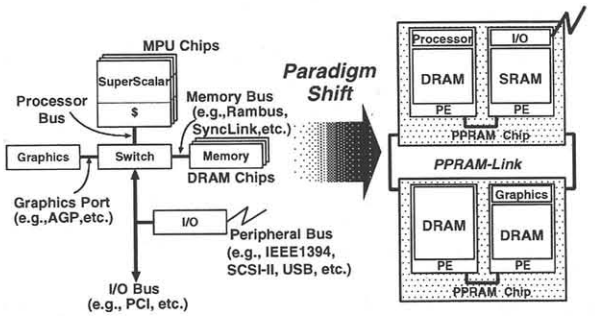

Figure 1: *PPRAM* Paradigm Shift

- **Parallel/distributed processing** technology will allow us to (1) improve the total system performance beyond the limits of instruction-level parallelism by means of exploiting higher-level parallelism on (single-chip and/or multiple-chip) multiprocessor; (2) reduce the design cost by means of simplifying the MPU design with putting multiple simple processors rather than a complex superscalar processor; (3) optimize the power consumption by adjusting the number of active processors depending on the workload; (4) have the designed system scalable in terms of the size, the functionality and the performance; (5) enhance the yield and reliability of chips by exploiting redundant processors, and so on.

- **Standardized high-speed inter-chip communication interface** will allow us to (1) interconnect and inter-operate multiple *PPRAM*-chips supplied by different vendors; (2) port software on various *PPRAM*-based systems; (3) focus on the design of application-specific logic and memory rather than the *PPRAM-Link* interface, and therefore reduce the design costs, and so on.

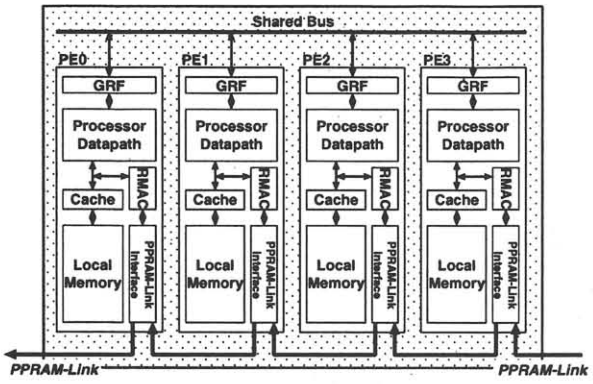

Figure 2: $PPRAM^{\mathcal{R}}$

3. Reference *PPRAM*: $PPRAM^{\mathcal{R}}$

Reference *PPRAM*($PPRAM^{\mathcal{R}}$) is an architectural implementation of *PPRAM*and a counterpart against contemporary high-performance microprocessor architectures[1]. We are currently developing a prototype chip of the $PPRAM^{\mathcal{R}}$in order to show the viability and cost-performance effectiveness of the $PPRAM^{\mathcal{R}}$. Figure 2 shows the block diagram of this chip and Table 1 outlines the chip characteristics.

The chip integrates 256Mb DRAM and four 32b RISC processors, and therefore is referred to as $PPRAM^{\mathcal{R}}$256-4. The 256Mb DRAM is distributed for each processor as local memory (64Mb or 8MB for each) to exploit its inherently high memory bandwidth. A pair of a processor and its local memory is referred to as processing element (PE) or *PPRAM* node. As shown in Figure 2, each PE consists of a processor, an 8MB local memory (DRAM), a 24kB cache memory (SRAM), a remote-memory access controller (RMAC), and a *PPRAM-Link* interface. The cache memory (SRAM) and the local memory (DRAM) are interconnected with each other through 1024 signal lines. Assuming that the DRAM access time is 40ns, the peak memory bandwidth per PE is 1kb/40ns = 25Gb/s $\approx$ 3GB/s.

Each *PPRAM* node is interconnected with other *PPRAM* nodes inside and outside a chip through *PPRAM-Link* . The *PPRAM-Link* provides a high-bandwidth interface for communicating among two or more *PPRAM* nodes by using a collection of fast point-to-point unidirectional links. The *PPRAM-Link* is defined at 1GB/s (16b parallel). The *PPRAM-Link* provides a single global physical address space and remote-memory access capabilities. Besides the *PPRAM-Link* , inside a chip, all the processors share a register file, or GRF (Global Register File). The GRF provides a low-latency communication and synchronization among processors on the same chip.

Table 1: Chip Overview

Number of PEs	4
Local memory	8MB DRAM per PE
I-cache memory	8kB SRAM per PE
D-cache memory	16kB SRAM per PE
Memory bandwidth	3GB/s(LM-D-cache) per PE
Processor logic	500kT 32b RISC
PPRAM-Link I/F	18b/link× 2links 1GB/s per unidirectional link
Clock	100MHz (target)
Process technology	0.25μm CMOS, merged DRAM/logic
Die size	450mm^2

4. Conclusion

Again, Table 1 shows the characteristics of the prototype chip $PPRAM^{\mathcal{R}}$256-4. We will fabricate the full version of $PPRAM^{\mathcal{R}}$256-4 (otherwise its half portion; i.e., $PPRAM^{\mathcal{R}}$128-2) by the end of March 1999. We have already fabricated four different test chips with the help of the VDEC at Univ. of Tokyo.

Specifications for *PPRAM-Link* (physical layer, logical layer, and software API) are now being developed by *PPRAM* Consortium. The first version of *PPRAM-Link* Standard Draft will be published and distributed in public by the end of March 1998.

Acknowledgments

We would like to thank the members of *PPRAM* Consortium and the members of Yasuura/Murakami/Iwaihara Laboratory for their helpful discussions and comments. This work is supported in part by Japanese MESC Grant (09358005) and a research grant from Fujitsu Laboratory. The *PPRAM* Consortium is supported by Fuji Xerox, Matsushita, Mitsubishi, NEC, Oki, Samsung, SONY, Taito, TI, and Toshiba.

References

1) Murakami, K., et al., "Parallel Processing RAM Chip with 256Mb DRAM and Quad Processors," *1997 ISSCC Digest of Technical Papers*, pp.228–229, Feb. 1997.

2) Patterson, D., et al., "A Case for Intelligent RAM: IRAM," *IEEE Micro*, pp.34–44, Apr. 1997.

3) *PPRAM* Consortium, *http://www.ppram.or.jp/*