Multi-Bank Register File for Increased Performance of Highly-Parallel Processors

Koh Johguchi[1], Ken-ichi Aoyama[1], Tetsuya Sueyoshi[1], Moto Maeda[2], Hans Jürgen Mattausch[1], Tetsushi Koide[1], Tetsuo Hironaka[2] and Kazuya Tanigawa[2]

1 Research Center for Nanodevices and Systems, Hiroshima University, Higashi-Hiroshima, 739-8527 Japan
2 Department of Computer Engineering, Hiroshima City University, Hiroshima, 731-3194 Japan

1. Introduction

Exploitation of parallelism on the instruction or program- thread level is a widely used method for improving processor performance. To support the continuing trend towards higher parallelism, access ports and entry numbers of the processor's register file must be increased sufficiently. For instance, a processor design which supports 8-issue/4-thread Simultaneous Multi-Threading (SMT) needs to have a 24 port, 512 entry register file [1]. The hard to overcome problems associated with such large-port-number register files are enormously increased access time, area and power consumption, if the conventional multi-port-cell architecture is used [2]. Here we report, how these problems can be mitigated, without degrading the processor's cycle-per-instruction (CPI) performance, by a multi-bank structure of the register file.

2. Multi-Bank Register File Architecture

Figure 1 shows the global structure of the investigated superscalar processor architecture with multi-bank register file [3], which features a specialized register-access scheduler, and includes an appropriate rename stage and access queues with 2-3 entries. To quantify the effect of the multi-bank register file we discuss an architecture implementation with the MIPS10000 instruction set architecture, 4 instruction fetches, 7 pipeline stages as well as a 4-bank register file with 8 read ports, 4 write ports, 128 entries and 32 bit word length.

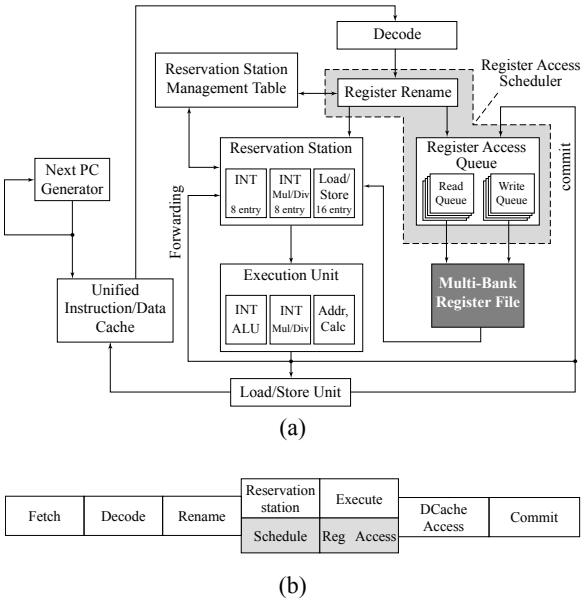

Fig. 1: (a) Overall block diagram of the superscalar-processor with MIPS10000 instruction set architecture and 4 parallel instruction fetches which utilizes the proposed multi-bank register file.
(b) 7-stage-pipeline architecture with decoupling of register access and instruction execution.

We apply a distributed crossbar architecture called Hierarchical Multi-port-memory Architecture (HMA) [2] for the multi-bank register file. Bank pre-charge and access up to the final bank-internal decoder stage occur simultaneously during the 1st phase of the clock cycle while bank-internal access is carried out during the 2nd phase of the clock cycle. Each of the 4 banks contains 2-port SRAM cells with separate read/write ports. A 1-to-8 read-port and a 1-to-4 write-port converter are used as interfaces to the upper hierarchy level. Therefore, each bank can accept 1 read and 1 write access simultaneously in each clock cycle. More simultaneous bank accesses would result in bank-access conflicts and delay the processing of the related instructions. To avoid these access-conflict situations, an efficient register access scheduling methodology utilizes register renaming, access-number reduction by forwarding as well as combination of accesses to the same register, and register access queues for enabling out-of-order access. Examining these scheduling methods with benchmark

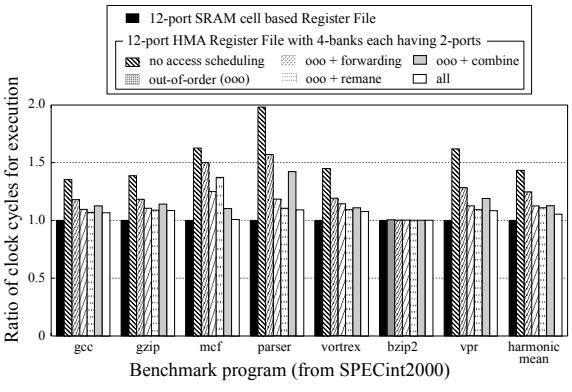

Fig. 2: Evaluation of the register access scheduling efficiency on the basis of clock-cycle numbers for program execution.

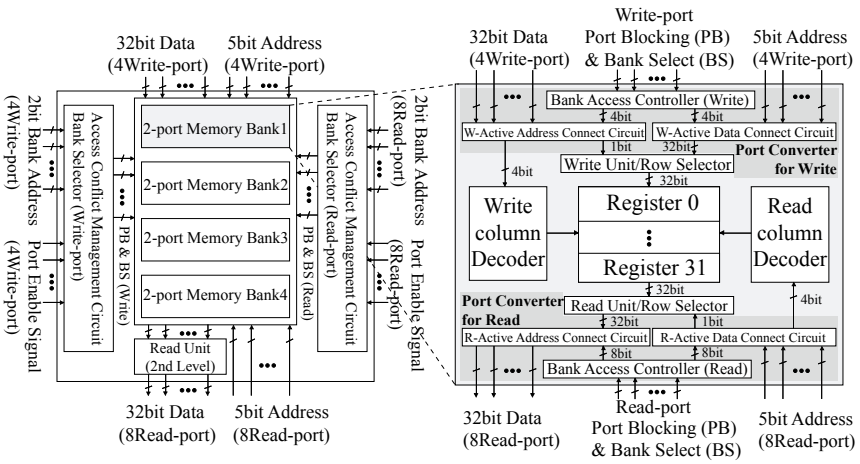

Fig. 3: Architecture of the designed 12-port multi-bank HMA register file with 4 banks, 128 registers and 2 ports per SRAM cell.

simulations, shows that the proposed multi-bank register file is capable of achieving a cycle based instruction-processing performance nearly equivalent (< 5% average degradation) to an ideal 12-port register file (Fig. 2).

Figure 3 shows the block diagram of the 8read/4write-port register file design. The circuits for the access-time-critical read/write path of the designed register file are shown in Fig. 4 (read-access path highlighted by fat lines). Read and write path are divided up to the 2-port SRAM cells.

In order to optimize the access time, the HMA register file uses a new access method with hidden pre-charge. This method hides the pre-charge phase by carrying out access port selection, address decoding, etc. simultaneously during pre-charge of banks, and reduces thus the clock cycle time (Fig. 5). When clock is "0", the register file access up to the bank's word-line driver is operated simultaneously with bank pre-charging. The remaining access part, starting with word-line driver activation, is carried out when clock is "1". The HMA design can implement this new access method easily because the timing of signals is easy to control in each of the hierarchies.

The layout-design of the multi-bank register file, including the 2-port SRAM cells, is carried out in 4-metal standard logic CMOS with 200 nm drawn gate length. The fabricated register file, shown in the chip photo of Fig. 6, has an area of 803 × 498 μm^2 (0.39 mm^2). Figure 7 shows the read-access wave forms as measured with an electron-beam tester during the 1st phase of the clock up to the final bank-internal decoder stage (Fig. 7a) and

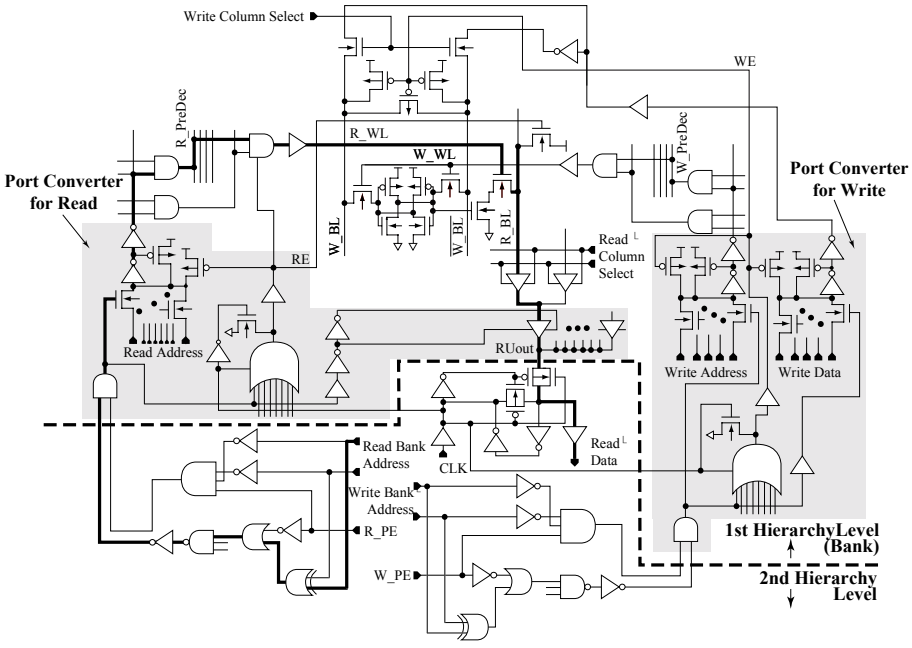

Fig. 4: Read- and write-access path of the designed 12-port multi-bank HMA register file. The critical read-access path is highlighted by fat lines.

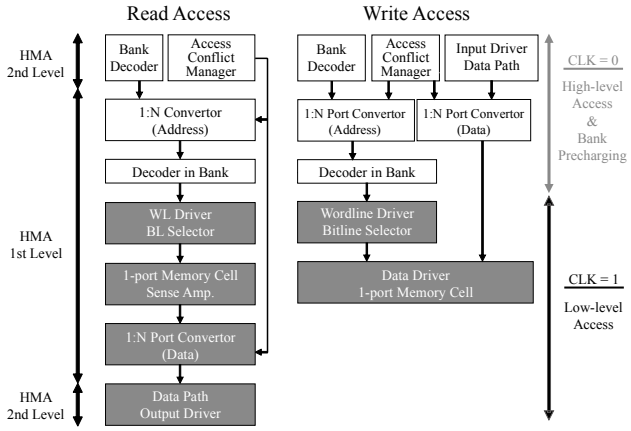

Fig. 5: Access method with hidden pre-charge, as realized in the designed register file.

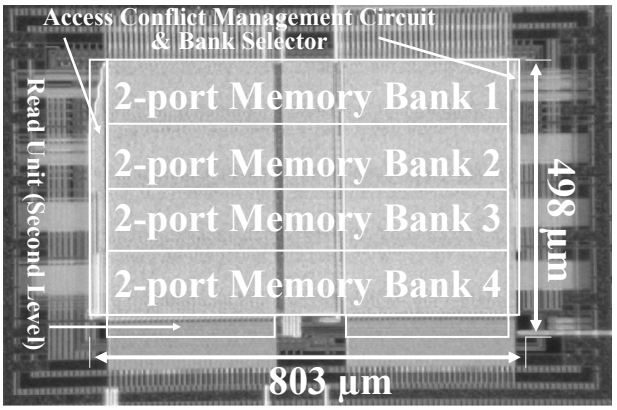

Fig. 6: Chip photo of the fabricated HMA register file in 200nm CMOS.

during the 2nd phase of the clock until data output (Fig 7b). Because the bank-internal read enable signal RE and the read pre-charge signal R_PreDec are allowed to rise at the same time, the CLK transition "0" to "1" can occur 0.4 ns before R_PreDec rises (see Fig. 7c). Thus the minimum cycle time is 2.4 ns if an appropriate clock-duty ratio is used, resulting in a maximum operation frequency of 417 MHz for the fabricated register file.

3. Performance Comparison

The achieved performance data of the designed multi-bank HMA register file are compared in table 1 with conventional multi-port-cell based designs, namely, with our 12-port-cell design in the same technology and with a recently reported high-performance 16-port-cell design in 110 nm CMOS technology [4]. The direct comparison in same technology is based on the simulation data and shows that the multi-bank register file allows 95 % higher clock frequency and 73 % smaller area. The indirect comparison with the 16-port-cell design is carried out based on measured data and by applying constant electric-field scaling to obtain an estimate based on equal design rules. The multi-bank design is estimated to have 37 % higher clock frequency, 58 % smaller area consumption and 52 % smaller power dissipation. The access-scheduling overhead of the multi-bank register file does not degrade the clock-frequency advantages if appropriate pipelining is applied, but has an effect on power-dissipation and area consumption. These issues have been investigated on the basis of a Verilog-HDL design and a gate-level synthesis. The results suggest that power-dissipation and area-consumption advantages remain at a reduced level of 10-20 % and 5-10 %, respectively.

4. Conclusion

In this paper, a 12-port multi-bank type register file has been proposed and evaluated by a design, fabrication and simulation study. From the simulation result, the average degradation of the instruction-processing performance of a superscalar processor with MIPS10000 instruction set architecture and 4 parallel instruction issues is less than 5%, compared to a conventional 12-port cell register file. On the other hand, it is verified by a full-custom design and fabrication study that the proposed multi-bank register file has about twice the maximum clock frequency and less than a third of the

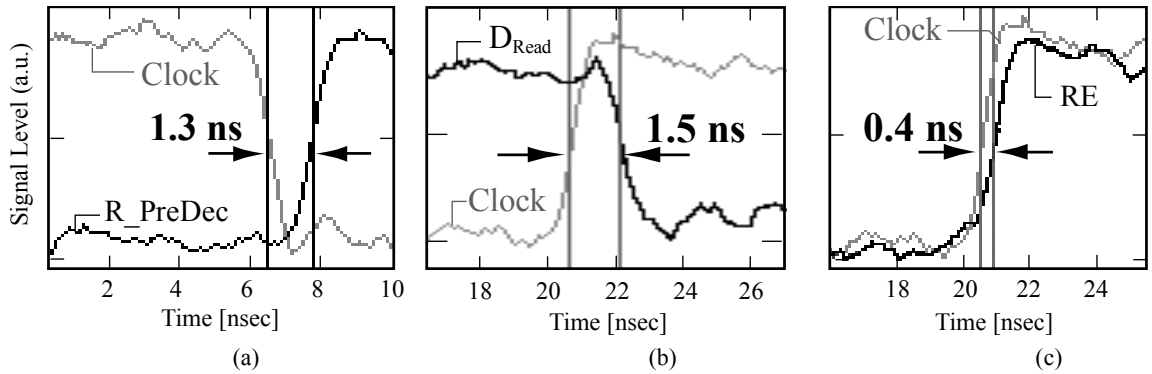

Fig. 7: Read-access of the fabricated HMA register file as measured by an electron-beam tester. (a) Up to the final bank-internal decoder stage during the 1st clock phase. (b) Up to data output during the 2nd clock phase. (c) Delay between clock and bank-internal read enable signal. R_PreDec, D_{Read} and RE signals are the output of the 1st stage of the wordline decoder for read access, the final output of read data from the register file and the internal read enable signal, respectively.

Table I: HMA register file design data and comparison with conventional high-performance multi-port-cell register file designs.

	Multi-bank Register File (HMA)	Conventional Multi-port-cell Register File	Multi-bank Register File (HMA), **estimated**	Conventional Multi-port-cell Register File[4]
Technology	200nm L_{gate} 5 metal CMOS	200nm L_{gate} 5 metal CMOS	110nm L_{gate} 4 metal CMOS	110nm L_{gate} 4 metal CMOS
Supply voltage	1.8 V	1.8 V	1.2 V	1.2 V
Access ports	12 (8r, 4w)	12 (8r, 4w)	16 (10r, 6w)	16 (10r, 6w)
Registers	128	128	34	34
Word length	32 bit	32 bit	64 bit	64 bit
Core area	0.39 mm^2	1.43 mm^2	**0.21 mm^2**	0.5 mm^2
Max operation frequency	640 MHz (simulated) 417 MHz (measured)	330 MHz (simulated)	**1140 MHz (from simulation)** **746 MHz (from measurement)**	545 MHz (measured)
Power dissipation	210 mW @500 MHz (simulated)	105 mW @330MHz (simulated)	**106 mW** @500 MHz	220 mW @500 MHz

area of a conventional 12-port cell register file. Consequently, the overall instruction processing performance including the effect of access conflicts has an improvement potential of about 90% in the same CMOS technology due to the application of a multi-bank register file. The fabricated test chip of the 12-port register file with 4 banks was measured with an electron-beam tester and found to operate at about 35% lower clock frequency than simulated. The reason for this difference is presently under investigation and maybe due to a mistake in the test-chip preparation for the electron-beam tester. Nevertheless, even at this lower measured clock-frequency an overall improvement in the instruction-processing performance of more than 20% due to the 12-port multi-bank register file remains. This performance improvement due to a multi-bank register file becomes larger with further increasing port number of the register file.

Acknowledgment

The guidance of R. Zheng, the University of Tokyo, was essential for obtaining successful electron-beam tester measurements. This research was sponsored by Semiconductor Technology Academic Research Center (STARC), Yokohama, Japan, and partly supported by the 21st Century COE program, Nanoelectronics for Terabit Information Processing, Japanese Government, and the JSPS research fellowship, 1605246. The test-chip has been fabricated in the chip fabrication program of VDEC, the University of Tokyo in collaboration with Hitachi Ltd., Dai Nippon Printing Co. and Cadence Design Systems, Inc.

References

[1] R. P. Preston, R. W. Badeau, D. W. Bailey, S. L. Bell, L. L. Briro, et al., "Design of an 8-wide superscalar RISC micro-processor with simultaneous multithreading," Dig. Of Tech. Papers, ISSCC2002, pp. 334-335, Feb. 2002.

[2] H. J. Mattausch, K. Kishi and T. Gyoten, "Area-efficient multi-port SRAMs for on-chip data-storage with high random-access bandwidth and large storage capacity," IEICE Trans. Electron., Vol. E84-C, No. 3, pp. 410-417, 2001.

[3] T. Saito, M. Maeda, T. Hironaka, K. Tanigawa, T. Sueyoshi, K. Aoyama, et al., "Design of Superscalar Processor with Multi-Bank Register File," Proc. of 2005 IEEE Internati onal Symposium on Circuits and Systems (ISCA2005), pp. 3507-3510, May. 2005.

[4] N. Tzartzanis, W. W. Walker, H. Ngyuyen, A. Inoue, "A 34Word x 64b 10R/6W write-through self-timed dual-supply-voltage register file," Dig. of Tech. Papers, ISSCC2002, pp. 416-417, Feb. 2002.