

A 2-stage-pipelined 16 port SRAM with 590 Gbps random access bandwidth and large noise margin

Koh Johguchi^{a)}, Yuya Mukuda, Ken-ichi Aoyama,
Hans Jürgen Mattausch, Tetsushi Koide

Research Center for Nanodevices and Systems, Hiroshima University,
1–4–2 Kagamiyama, Higashi-Hiroshima, Hiroshima 739–8527, Japan
a) johguchi@hiroshima-u.ac.jp

Abstract: A 90 nm CMOS, 64 Kbit, 1.16 GHz, 16 port SRAM with multi-bank architecture realizing 590 Gbps random access bandwidth, 41 mW power dissipation at 1 GHz and 0.91 mm² (13.9 μm²/bit) area consumption is reported. Compared to conventional 16 port SRAM data, area and power consumption are reduced by factors 16 and 5, respectively, while maximum clock frequency is about a factor 2 higher.

Keywords: SRAM, multi-port memory, distributed crossbar, multi-stage sensing

Classification: Integrated circuits

References

- [1] H. J. Mattausch, K. Kishi, and T. Gyohten, “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.
- [2] K. Johguchi, et al., “Multi-bank register file for increased performance of highly-parallel processors,” *Proc. of ESSCIRC2006*, pp. 154–157, 2006.
- [3] L. Chang, D. M. Fried, J. Hergenrother, J. W. Sleight, R. H. Dennard, R. K. Montoye, L. Sekaric, S. J. McNab, A. W. Topol, C. D. Adams, K. W. Guarini, and W. Haensch, “Stable SRAM cell design for the 32 nm node and beyond,” *Dig. of 2005 VLSI Symp. Technol.*, pp. 128–129, June 2005.
- [4] K. Takeda, Y. Hagihara, Y. Aimoto, M. Nomura, Y. Nakazawa, T. Ishii, and H. Kobatake, “A read-static-noise-margin-free SRAM cell for low-VDD and high-speed applications,” *IEEE JSSC*, vol. 41, no. 1, pp. 113–121, 2006.
- [5] K. Zhang, K. Hose, V. De, and B. Senyk, “The scaling of data sensing schemes for high speed cache design in sub-0.18 μm technologies,” *Dig. of 2000 VLSI Symp. Circuits*, pp. 226–227, 2000.
- [6] N. Tzartzanis, W. W. Walker, H. Nguyen, and A. Inoue, “A 34Word x 64b 10R/6W write-through self-timed dual-supply-voltage register file,” *ISSCC Dig. Tech. Papers*, pp. 416–417, 2002.

1 Introduction

Recently, multi-core technology became widely used for microprocessor and SoC applications. In these integrated systems high port number embedded memories with large random access bandwidth and sufficient storage capacity are urgently needed for realizing common data storage or inter-unit communication capability, e.g. by unified cache memories in multi-core processors and buffer memories of chip-integrated network switches between an SoC's functional units, respectively. However, ideal multi-port memory, based on the conventional multi-port cell architecture leads to a large penalty in silicon area, access time and power consumption due to the increase of port-related signal lines, which blow-up the size of bit-storage cells. Therefore, bank-based multi-port architecture has been developed, which can drastically reduce area consumption by using 1- or 2-port banks and a distributed crossbar switch for large-port-number capability [1]. At the same time, memory-access time and power dissipation are also reduced substantially. Access-conflicts, which may happen when a bank-based architecture is adopted, are avoidable by an access scheduling which takes account of the bank structure [2].

2 Architecture

We propose a 16-port memory architecture with distributed crossbar, which we call Hierarchical Multi-port memory Architecture (HMA) [1, 2]. The bank modules of the 1st hierarchy level consist of a 2-port SRAM core into which the 1-to-8 read-port and write-port converters of the distributed crossbar are integrated. The employed 8 transistor 2-port SRAM cell has independent read and write ports and completely decouples the bit line from the internal storage node. This renders the storage cell static noise margin free and leads to high read stability [3, 4]. Therefore, it allows the coexistence of large access bandwidth and low-voltage operation in the same design, i.e. the selection of two operating modes (low-power or high-speed mode) via the supply voltage.

3 Circuit and Layout Design

A large signal sensing scheme [5] was implemented to further increase the reliability of the 16-port SRAM accesses. The concept of this scheme is to provide always large signals by dividing the read access path and inserting buffers. Using this method, read and write operation are very stable under high-speed operation, but area consumption is somewhat increased. The read access path of the 2-port SRAM core is shown in Fig. 1 (a). We use a 2-stage sensing scheme for reading data from 2-port SRAM cells within an accessed bank. Local bitlines are connected to only 8 SRAM cells (1st stage) and 4 local clusters are connected to the global bitlines (2nd stage).

Figure 1 (b) shows the schematic diagram of a part of the bank-internal 1-to-8 read-port converter, which adopts dynamic CMOS technology with DOMINO logic, and the repeater structure for bank columns and rows until the output latches. When the clock signal changes from "0" to "1," the read-port select signals SR_i are activated, and only if the read-data is "1" the

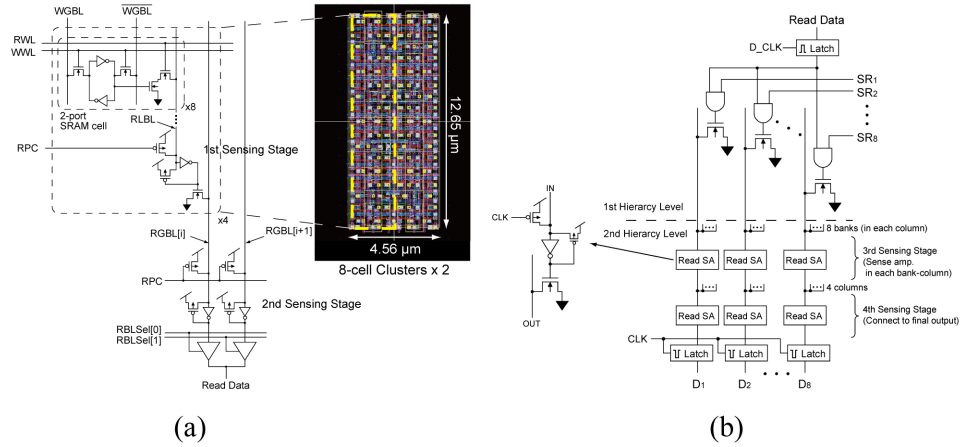


Fig. 1. Read-data path of the 2-port SRAM. (a) is within the 2-port SRAM core, and (b) in the 1-to-8 read-port converter of the bank and the repeater structure until the output latch.

corresponding pull-down NMOS transistor is activated to lower the potential of the corresponding 3rd stage sensing line for bank columns. Final read-port outputs are determined in the 4th stage sensing, which summarizes the bank-column results for each port.

The 16-port SRAM (see Fig. 2 (a)) is designed in 90 nm logic-CMOS technology with 6 metal layers, doesn't applying special SRAM-cell design rules, has a storage capacity of 64Kbit, a word-length of 32 bit and occupies a silicon area of only 0.91 mm², or 13.9 μm² per bit. In comparison to a previously reported 2.125 Kbit, 0.5 mm², 16-port SRAM design [6] in 110 nm CMOS with conventional 16-port SRAM storage cells which has 230 μm² area consumption per bit, this represents a bit-area reduction by a factor 16.5. Figure 2 (b) shows the layout of a 2-Kbit 2-port bank with two 1-to-8 port converters. The bank size is 0.185 mm², which includes the internal 2-port SRAM, the port converters of the distributed crossbar and the bank controller.

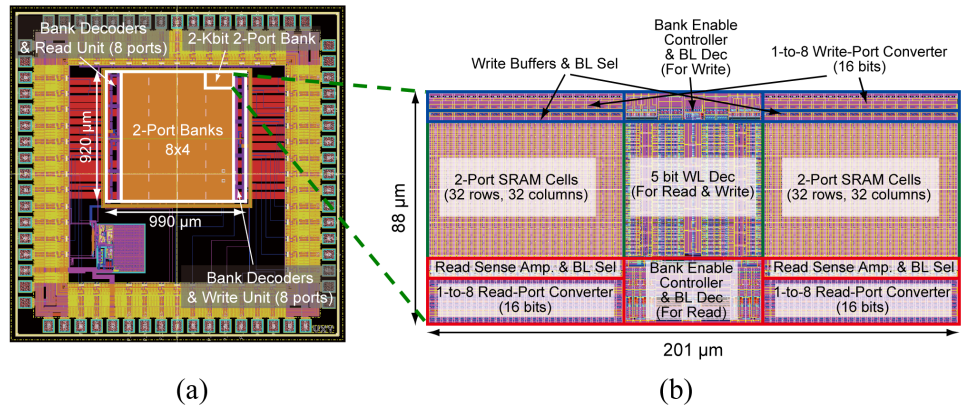


Fig. 2. Layout of the designed 16-port, 64-Kbit SRAM (a) and a 2-Kbit 2-port bank with 1-to-8 read- and write-port converters (b).

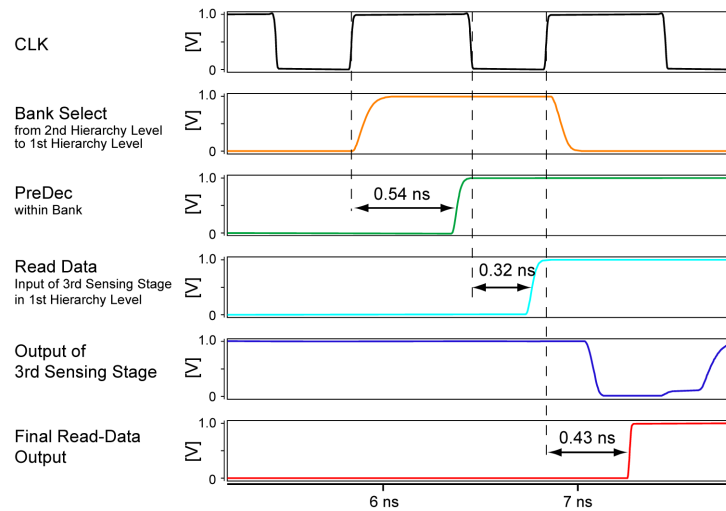


Fig. 3. Simulated wave form of the test chip at 1 GHz frequency. The clock cycle duty is 40% to 60% and the maximum operating frequency is 1.16 GHz.

Simulated maximum operating frequency and power dissipation of our 16-port SRAM design, determined with a 40% to 60% clock duty cycle from a layout-extracted net list, are more than 1 GHz (1.16 GHz) and 41 mW at 1 GHz, respectively (Fig. 3). This maximum clock frequency is a factor 2 higher than for the 16-port register file reported in [6], although we have realized a more than 30 times larger storage capacity. Also the power dissipation of 41 mW at 1 GHz compares favorable to the 5.3 times larger value of 220 mW at 500 MHz for the conventional design, even at our substantially higher clock frequency and larger storage capacity. Access latency of our 16-port SRAM is 2 clock cycles due to the application of a 2-stage pipeline. In the first clock cycle, bank decoding and conflict arbitration (clock “0”) as well as bank-internal control-signal generation and wordline pre-decoding (clock “1”) are carried out. In the second clock cycle, word-line driver activation and bank-internal read/write access (clock “0”) as well as the read-data transfer to the output ports via the 3rd and 4th large-signal sensing stages (clock “1”) are executed.

Due to its modular 2-dimensional structure, the architecture of the reported 16-port SRAM can be easily scaled to higher port as well as bank numbers, with only a marginal degradation of the maximum operating frequency. Consequently, random access bandwidth levels beyond the 1 Tbps threshold are achievable by a small redesign with a 32-port version of our architecture and design approach.

4 Conclusion

This paper reported a 16-port SRAM design, which realizes the highest published random access bandwidth for SRAMs by using a multi-stage sensing scheme and a distributed crossbar memory architecture with a 2-Kbit bank based on 2-port SRAM cells. Area per bit, power consumption and maxi-

mum clock frequency are about a factor 16, 5 and 2, respectively, better than previous 16-port SRAM designs.

Acknowledgments

The VLSI chip in this study has been fabricated through the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo, with the collaboration by Semiconductor Technology Academic Research Center (STARC), Fujitsu Limited, Matsushita Electric Industrial Company Limited., NEC Electronics Corporation, Renesas Technology Corporation, and Toshiba Corporation, Simucad Design Automation Inc., Mentor Graphics Corporation, and Cadence Design Systems Inc. This research was supported by STARC, Japan, by the 21st Century COE program of the Ministry of Education, Culture, Sports, Science and Technology, Japanese Government, and by a JSPS Research Fellowships for Young Scientists, 1605246.