A 9-ns HIT-Delay 32-kbyte Cache Macro for High-Speed RISC

KAZUTAKA NOGAMI, TAKAYASU SAKURAI, MEMBER, IEEE, KAZUHIRO SAWADA, KENJI SAKAUE, YUICHI MIYAZAWA, SHIGERU TANAKA, YOICHI HIRUTA, KATSUTO KATOH, TOSHINARI TAKAYANAGI, TSUKASA SHIROTORI, YUKIKO ITOH, MASANORI UCHIDA, AND TETSUYA IIZUKA, MEMBER, IEEE

Abstract—**A 32-kbyte cache macro with an experimental reduced instruction set computer (RISC) is realized. A pipelined cache access is proposed to realize a cycle time shorter than the cache access time. A double-word-line architecture combines single-port cells, dual-port cells, and CAM cells into a memory array to improve silicon area efficiency. The cache macro shows 9-ns typical clock-to-HIT delay owing to several new circuit techniques, such as a new section word-line selector, a dual transfer gate, and 1.0-μm CMOS technology. It supports multi-task operation with logical addressing by a selective clear circuit. The RISC includes a double-word load/store instruction using a 64-bit bus to fully utilize the on-chip cache macro. A new test scheme enables measurement of the internal signal delay. The test device is designed based on the unified design rules (UDR) scalable through multigenerations of process technologies down to 0.8 μm.**

I. INTRODUCTION

CPU PERFORMANCE has been increasing with the improvement of CPU architecture, circuit and process technologies. Recently, reduced instruction set computer (RISC) architecture has accelerated CPU performance [1]. A RISC architecture simplifies hardware and pipeline stages, and realizes less cycles per instruction and a shorter cycle time than a complex instruction set computer (CISC) architecture.

Owing to these features, a RISC architecture requires very wide memory bandwidth, which is a key problem in the RISC system. For this reason, high-performance cache memory is required in the RISC architecture.

Several attempts have been made to enlarge memory bandwidth by including cache memory on the same chip [2]. On-chip cache size has reached up to 12 kbytes [3].

Manuscript received July 27, 1989; revised September 28, 1989.

K. Nogami, K. Sawada, S. Tanaka, Y. Hiruta, K. Katoh, T. Takayanagi, Y. Itoh, and T. Iizuka are with the Semiconductor Device Engineering Laboratory, Toshiba Corporation, 1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan.

T. Sakurai is with the University of California, Berkeley, CA 94720 on leave from the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki 210, Japan.

K. Sakaue, T. Shirotori, and M. Uchida are with Toshiba Microelectronics Corporation, 1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan.

Y. Miyazawa is with Stanford University, Stanford, CA 94305 on leave from the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki 210, Japan.

IEEE Log Number 8932957.

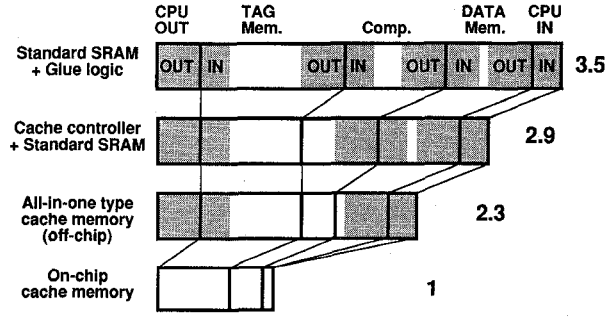


Fig. 1. Cycle-time comparison among various integration types of cache memory based on the same Si technology level.

However, the formerly reported cache size is not sufficient to obtain the maximum performance of RISC architecture.

In this paper, a 32-kbyte cache macro is described. This cache macro combines high hit ratio and high clock rate operation to provide sufficient effective memory bandwidth for a high-speed RISC.

Section II introduces the merits of an on-chip cache memory and shows the requirement for on-chip cache memory. An overview of this cache macro is given in Section III. In Section IV, pipelined cache operation is proposed, which is one of the key techniques to reduce cycle time. Memory core architecture is described in Section V. In Section VI, several new circuits are explained. Process technology is described in Section VII. Performance of the test device is summarized in Section VIII, and Section IX is dedicated to the conclusion.

II. WHY ON-CHIP CACHE MEMORY?

Cache memory is a reliable method to improve memory bandwidth. Among several approaches to organize cache memory, on-chip cache memory is the most effective approach. Fig. 1 shows a cycle-time comparison among various integration types of cache memory in a TTL I/O system. On-chip cache memory achieves more than twice the operational frequency of any other type of off-chip cache memory because it eliminates the inter-chip communication delay, denoted IN/OUT in the figure. The cycle-

0018-9200/90/0200-0100$01.00

time difference between on-chip and off-chip cache memory has become increasingly larger as Si-technology has advanced, because the cycle time of on-chip cache memory has been decreasing due to the improvement of process technology, while the cycle time of off-chip cache memory has reached the interface limitation, which is about 50 MHz in TTL interface. Moreover, on-chip cache memory easily employs a wide bus width. On the other hand, in the off-chip cache memory it is not easy to employ a wide bus width, which increases board area and system cost and causes serious noise problems.

There are several requirements in on-chip cache memory. Effective memory bandwidth in a hierarchical memory system with on-chip cache memory is expressed by following formula:

effective memory bandwidth
= (hit ratio) × (cache memory bandwidth) + (1 − hit ratio) × (off-chip memory bandwidth)

cache memory bandwidth
= (cache operation frequency) × (on-chip bus width)

off-chip memory bandwidth
= (off-chip memory access frequency) × (off-chip bus width)

From these equations, high hit ratio, high-speed operation, and wide bus width are necessary for high-performance cache memory. The hit ratio depends mainly on cache size. Therefore, a large high-speed on-chip cache memory with wide bus width is required in a RISC system.

III. Cache Overview

Table I summarizes system level features of the newly developed cache macro and an experimental RISC on a single chip. This cache macro was designed for the feasibility study of a large scale on-chip cache macro, so we adopted the simplest way of cache configuration. The cache macro is organized as a 32 kbyte direct-mapped configuration of data/instruction unified cache. A cache size of 32 kbytes is the largest ever reported. Bus operation is synchronous. The RISC includes a double-word load/store instruction. With the double-word load/store instruction, the RISC can handle two sequential words at a time using a 64-bit data bus, which doubles the performance in processing two consecutive 32-bit-data and/or 64-bit data.

Fig. 2 shows the cache configuration. Each cache line has four process ID (PID) bits, four VALID bits, 17 TAG bits, and 16 bytes of DATA. Using four PID bits, 16 different processes can share a cache macro. The cache macro has 2K lines.

Fig. 3 shows a block diagram of the cache macro. The memory core is separated into two blocks containing DATA part and TAG, VALID, and PID parts. The main word-line buffer is placed between the DATA part and other parts. This buffer isolates the capacitance of the DATA main word line from the TAG main word line, thus enhancing the speed of the TAG, VALID, and PID parts.

A HIT signal is generated by the readout data from TAG, PID, and VALID parts. Hit signal generation is a critical path in cache access, so that the DATA part is allowed to be a little slower than the TAG part.

Fig. 4 shows the timing chart of this cache macro. The cache macro operates synchronously and supports pipeline addressing to reduce the cycle time. The write operation takes two clocks in order to prevent a collision between the read and write data.

Two-way or four-way set associative cache will be easily organized by using the basic cells developed in this cache macro to improve hit ratio.

TABLE I
System-Level Features

Cache Macro	
Cache size	32Kbyte (data/instruction unified)
Configuration	Direct mapped
Operation	Synchronous
Address space	4G byte
Process ID	4bit (CAM cell)
Valid bit	1bit/word (selectively clearable)
Line size	16byte
RISC	
# of instructions	46
Pipeline stages	4
# of registers	32 (1 write port + 2 read port)
Load instruction	2 cycles
Store instruction	3cycles
Most of other inst's	1cycle
Branch instruction	Delayed jump

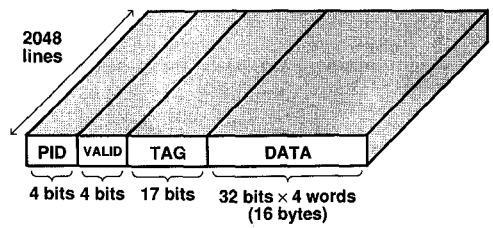


Fig. 2. Cache configuration.

IV. Pipelined Operation

In this cache macro, pipeline operation of cache access is proposed. Pipeline operation is widely used in logic devices, because pipeline operation is a reliable method to improve cycle time. However, pipeline operation has been scarcely adopted in a cache memory or synchronous memory because it is difficult to divide cache access or memory access into multiple pipeline stages.

Fig. 5(a) shows pipeline stages and delay factors of cache access. Cache access delay consists of row decoding, word-line driving, sensing, comparison between address and TAG, and data output. Conventionally, synchronous memory access starts from row decoding. In the pipelined cache memory, delay time is allotted to two pipeline stages.

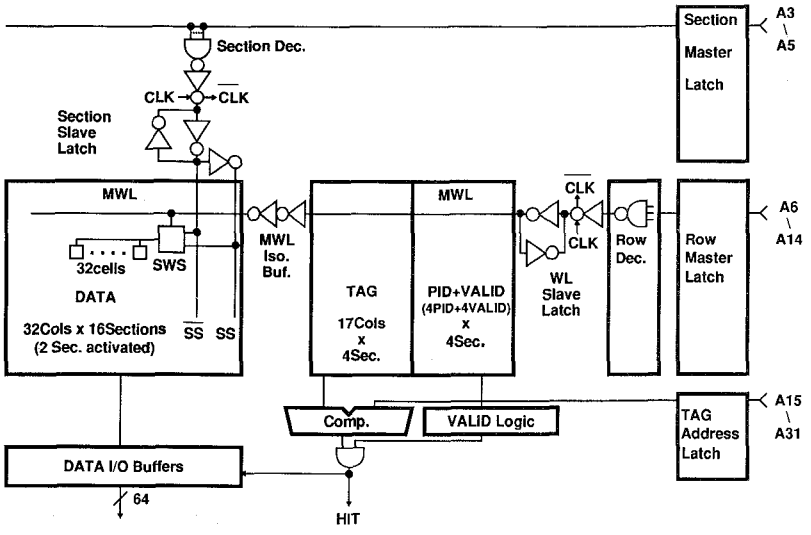


Fig. 3. Block diagram of cache macro.

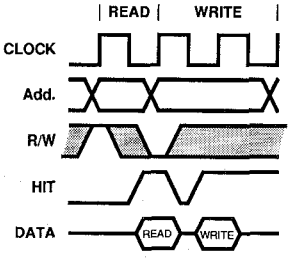


Fig. 4. Timing chart of the cache operation.

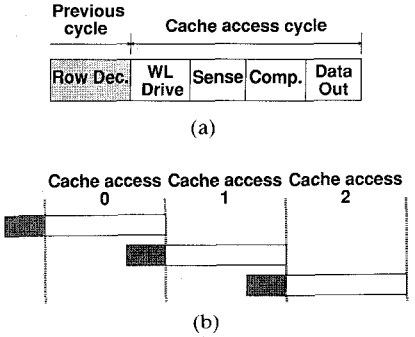


Fig. 5. Pipeline operation of cache access. (a) Pipeline stages and delay components of cache access. (b) Consecutive cache access flow.

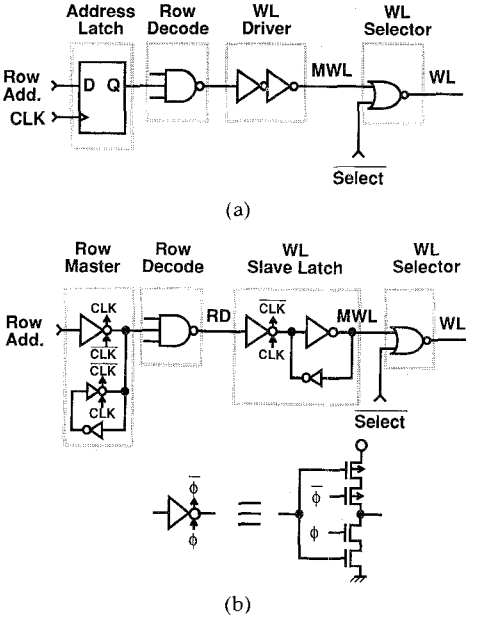


Fig. 6. Word-line select circuit. (a) Conventional. (b) Word-line slave latch.

The first stage is row address decoding and the second stage is a sequence of word-line drive to data out. In the fast SRAM, the delay from address input to word-line drive is about a half of the total delay. Therefore, pipelined row decoding is very effective in improving cycle time. The cache access cycle starts from word-line drive, and row decode is done in the previous cycle of cache access.

The sequence of cache access cycles is shown in Fig. 5(b). This figure shows that, using this pipeline operation, the cache macro can achieve shorter a cycle time than the cache access.

For this pipeline operation, a word-line slave latch is required, as shown in Fig. 6 in comparison with the conventional word-line select circuit. In the conventional word-line select circuit, row decoding starts after the row address is latched. In the synchronous memory, the address is set up before the clock assertion. Therefore, row decoding is kept waiting until the clock assertion.

This word-line slave latch, as shown in Fig. 6(b), is the key technique of the pipelined cache memory. The unique feature of this circuit scheme is that the row decoder is placed between the master and slave latches. Every word line has a slave latch with a clocked CMOS inverter. Therefore, a slave latch was designed to have the same pitch as that of the memory cell. Master latches are in the address buffers. Master and slave latches are transparent latches and triggered by complementary clock signals. These latches act as pipeline latches between the memory access stage and the previous pipeline stage, and a part of the address decoding time is merged into the previous pipeline cycle. The cache access is a critical path in the

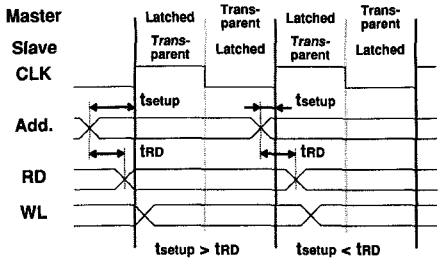


Fig. 7. Timing chart of word-line selection. The first cycle shows the case of the address setup time being longer than the row decoding time. The second cycle shows the case of the address setup time being shorter than the row decoding time.

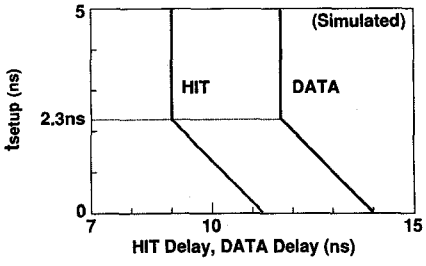


Fig. 8. Plot of address setup time versus HIT and DATA delay.

CPU system, therefore, this delay improvement directly enhances the system performance.

A timing chart of word-line selection is shown in Fig. 7. There are two cases of word-line selection. The first case is for an address setup time that is longer than row decoding time, as shown in the left portion of this chart. The second case is for an address setup time shorter than row decoding time, as shown in the right portion of this chart. In the first half of the cycle, address signals are latched by master latches at the rising edge of the clock signal. Then slave latches are transparent. In the second half of the cycle, row decode signals are latched and word lines are locked by slave latches at the falling edge of clock signal. Master latches are transparent and address signals are transmitted to the row decoder.

When address setup time is longer than row decoding time, row decode finishes before clock assertion, and word-line selection is controlled by the clock. In this case, cache cycle time is independent of address setup time. When the address setup time is shorter than row decoding time, row decode signals become valid within the first half of the cache access cycle, and then the word line is selected. In this case cache cycle time depends on address setup time. In both cases, the whole or a portion of row decoding time is merged into the previous pipeline cycle.

Furthermore, the word-line slave latch does not require any additional timing conditions compared with conventional circuits. Therefore, we can use the word-line slave latch without any special considerations. In other words, the address generation circuit of the CPU has relaxed constraints.

Fig. 8 shows the relation between address setup time, HIT delay, and DATA delay. When address setup time is longer than 2.3 ns, HIT and DATA delay are constant.

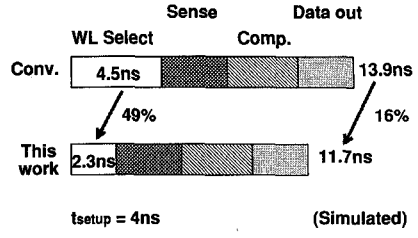


Fig. 9. DATA delay comparison.

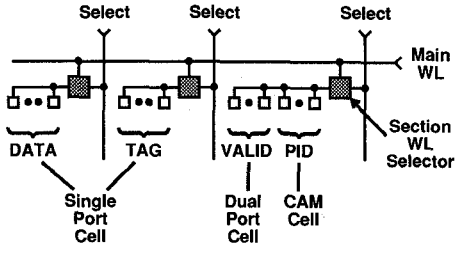


Fig. 10. Double-word-line architecture. Single-port cell, dual-port cell, and CAM cell are designed to have the same word-line pitch.

When address setup time is shorter than 2.3 ns, HIT and DATA delay become larger as address setup time decreases. The turning point of 2.3-ns address setup time is identical with row decoding time. In most applications, address setup time is set to be more than 2.3 ns. Therefore, row decoding time is fully eliminated from the cache access cycle in most applications.

A clock-to-DATA delay comparison between the pipelined cache memory and the conventional cache memory is shown in Fig. 9. This comparison is done using SPICE simulation. A 4-ns address setup time is assumed. Clock-to-word-line delay decreases from 4.5 to 2.3 ns. This 49-percent reduction is achieved by the word-line slave latch. Clock-to-DATA delay also decreases from 13.9 to 11.7 ns, which corresponds to a 16-percent reduction. The cache access is a critical path in the CPU system. Therefore, this delay improvement directly enhances the system performance.

V. Core Architecture

Fig. 10 shows the double-word-line architecture employed in this cache macro. Using this double-word-line architecture, TAG, VALID, PID, and DATA parts are combined into a single memory array. Conventionally, DATA part and other parts have been separated into two memory arrays. However, in a large-scale cache memory, both memory arrays become large, so it is difficult to draw the floor plan with limited Si resources.

Double-word-line architecture has been used in a high-density standard static RAM [4] and an off-chip cache memory [5], [6]. However, this double-word-line architecture is different from the conventional one. In this cache macro, single-port cell is used in the DATA and TAG parts, a dual-port cell in the VALID part, and a CAM cell in the PID part. To utilize a double-word-line architecture, each cell was designed to have the same word-line pitch. A

combination of memory arrays of different cells is effective in improving silicon area efficiency.

Another advantage of the double-word-line architecture is power savings. Power consumption is one of the most serious limitations in memories with wide bus width, because a large number of bits are read out at a time. In the conventional architecture, all memory cells connected to a selected row are activated and consuming power. In this architecture, a minimum number of cells are activated in an operation by selecting only four section word lines connected to a decoded row, so the power consumption is minimized.

VI. Circuit Design

Several novel circuit technologies are employed in this cache macro.

A new word-line selector to achieve high-speed core operation is described in Section VI-A. A dual transfer gate scheme is treated in Section VI-B, which dissolves bump-down delay. A selective clear circuit to support logical addressing is discussed in Section VI-C. Lastly, a test circuit to measure internal signal delay is described.

A. Word-Line Selector

Fig. 11 shows the core circuitry in this cache macro. The new word-line selector is shown in the right portion. This new word-line selector consists of two CMOS inverters and one section word-line shortening transistor. When the word-line selector is selected, the inverters are activated and drive the section word lines. When unselected, the section word line is discharged through the NMOS FET of the inverter or shortening transistor.

Conventionally, a CMOS NOR gate has been used in high-density RAM's. In conventional word-line selectors, serially connected PMOS FET's with slightly increased size are required for high-speed word-line drive and that increases the select signal and main word-line capacitances. This section word-line selector needs 2.5 transistors per section word line, less than the four transistors of a conventional section word-line selector. To keep the same transition time of the word line, the size of PMOS FET's in the new word-line selector is about half that in the conventional word-line selector. Therefore this circuit reduces the capacitances of the select line and main word line by 25 and 40 percent, respectively. This contributes to the speed-up of cache access by 10 percent.

B. Dual Transfer Gate

The dual transfer gate circuit is also shown in Fig. 11. The dual transfer gate consists of a PMOS transfer gate assigned for read operation, and an NMOS transfer gate assigned for write operation. One transfer gate is activated at a time. The bit line is precharged to the V_{cc} level by a PMOS precharge circuit.

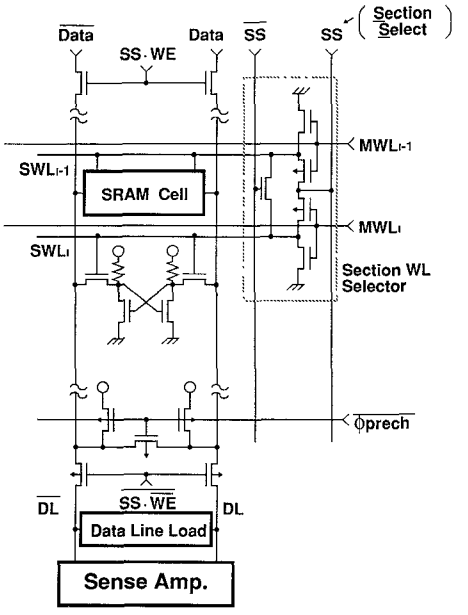


Fig. 11 Core circuitry.

Usually, an NMOS bit-line load and NMOS transfer gate are used in standard SRAM's. In such a circuit, bit-line overprecharge caused by bump-down of the supply voltage causes access time delay [7]. Therefore, the bit-line pull-down load is activated in a selected bit line. In the cache memory, such a bit-line pull-down load cannot be adopted. Since a large number of bits are read out at a time, a bit-line pull-down load causes large power consumption.

The PMOS transfer gate and bit-line V_{cc} precharge are free from the V_{cc} bump-down problem because the bit-line signal is transmitted to the sense amplifier even after the V_{cc} bump-down.

C. Selective Clear Circuit

Fig. 12 shows the selective clear technique. Selection is done by discharging the MATCH line. When the "process" is switched in the multi-task system, MATCH lines are precharged, and the process ID comparison is done using CAM cells. Then CLEAR is asserted. If the process ID matches, MATCH lines maintain the precharge level, so the clear port of the dual-port cell is activated, and the corresponding VALID bits are selectively cleared. A flush clear function is also available in this circuit by eliminating the process ID comparison from the selective clear function.

A clear operation can be done independently of normal cache operation, because a complete clear operation can be executed without using the normal operation port. This is useful for the logical address cache to achieve high-speed process switching.

A current limit transistor is connected to the bit line of the clear port for improving the reliability. If a large number of VALID cells are cleared without this transistor, a large peak current flows into the bit line of the clear

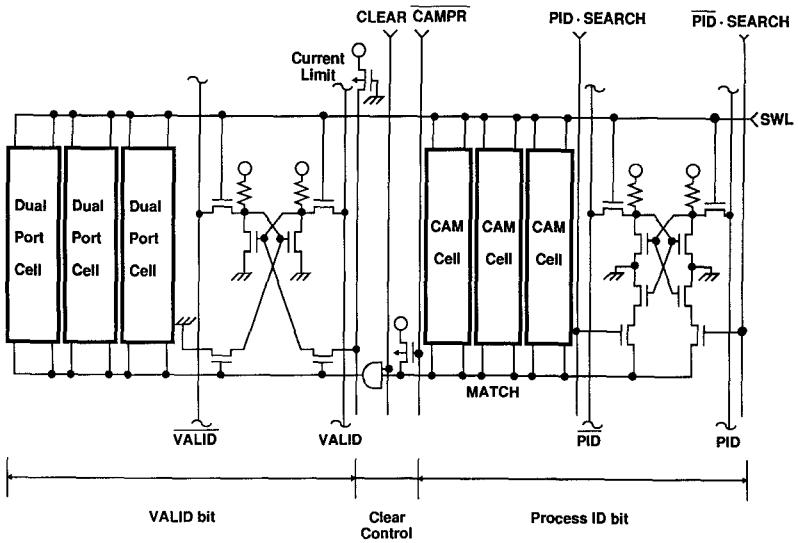


Fig. 12. Selective clear circuit.

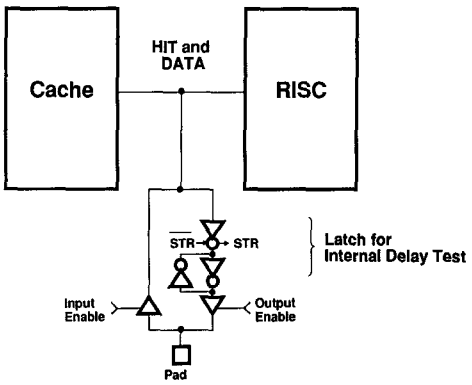


Fig. 13. Test circuit.

TABLE II
DEVICE FEATURES

Technology	Double Al & double poly twin-well CMOS
Design rule	1.0μm basic rule
MOSFETs	0.8μm gate length
Cell size	
SRAM cell	9.2μm × 13.8μm
Dual-port cell	20.0μm × 13.8μm
CAM cell	25.1μm × 13.8μm
Cache macro size	8.7mm × 8.7mm
Chip size	14.5mm × 10.8mm
Strobe to HIT	9ns (typical)
Strobe to DATA	12ns(typical)
Cycle freq.	80MHz (typical)

port, which will cause the Al line to open from electromigration.

This cache macro employs the polysilicon-load CAM cell, because the polysilicon-load CAM cell is 40 percent smaller than the pure CMOS CAM cell.

D. Test Circuit

For accurate evaluation of very-high-speed devices, on-chip test circuits are useful rather than an external expensive LSI tester. Moreover it is impossible for an external tester to measure the internal signal delay, which is much shorter than the interface delay between the device and the tester.

Fig. 13 shows a test circuit to measure the internal HIT and DATA delay. A latch is included in the PAD I/O buffer. When the strobe (STR) goes to ONE, the clocked CMOS inverter becomes high impedance and the present state of the bus is latched. By varying the STR timing, it is possible to find the fastest successful timing, which is the real internal output timing of the cache macro. The STR is wired to all test circuits. The signal-to-signal delay is measured as the difference of STR timing. The size of this test circuit is 0.02 mm^2, which is sufficiently small to measure multiple internal signals.

VII. SI PROCESS TECHNOLOGY

The whole layout was done under the unified design rules (UDR) [8]. The UDR is a common layout rule scalable to multigenerations of logic LSI process technologies. Therefore, this cache macro and various types of memory cells are available for a memory macro embedded in the logic devices.

The test device was fabricated using a double-aluminum and double-polysilicon twin-well CMOS technology. The basic design rule of 1.0 μm, and 0.8-μm gate length MOSFET's were used. Resistor-load memory cells were used for single-port cells, dual-port cells, and CAM cells. Four additional masks were added to the standard logic process to make the resistor-load memory cells.

VIII. PERFORMANCE

Device-level features are listed in Table II. Single-port cells, dual-port cells, and CAM cells were designed with the same row pitch. Single-port cell size is slightly larger

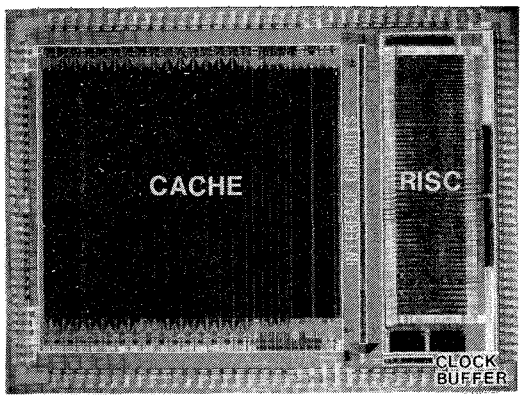


Fig. 14. Chip microphotograph. Chip size is 10.8 mm×14.5 mm, and cache macro size is 8.7 mm^2.

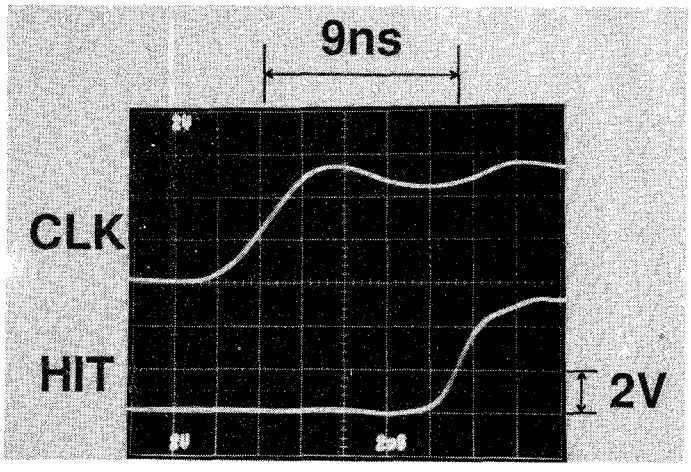


Fig. 15. Measured internal waveforms.

than the standard SRAM cell with same design rule because of the compatibility with the unified design rules.

Fig. 14 shows a chip microphotograph of the test device. An experimental RISC is implemented and interface circuits are placed between the cache macro and the RISC. Chip size is 10.8 mm×14.5 mm, and the cache macro size is 8.7 mm×8.7 mm.

Fig. 15 shows measured internal waveforms. Clock-to-HIT delay is 9 ns. Data buffers are activated by the HIT signal with 3-ns delay, so DATA is accessed in 12 ns. This measurement demonstrates that this cache macro has the capability of 80-MHz operation.

IX. Conclusion

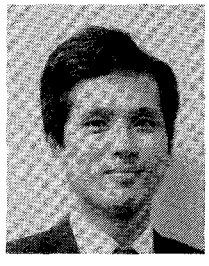
A 32-kbyte cache macro was realized using a 1.0-μm double-aluminum double-poly twin-well CMOS process with 0.8-μm MOSFET's. Pipelined cache access was proposed to realize a fast cycle time. Double-word-line architecture improved silicon area efficiency by combining different memory cells into an array. A 9-ns HIT delay was achieved by several new circuit techniques including a new section word-line selector and a dual transfer gate. The test device was designed under the unified design rules and is available in multigeneration process technology down to 0.8 μm.

Acknowledgment

The authors would like to express their deep appreciation to H. Yamada and Y. Unno for their encouragement throughout this work.

References

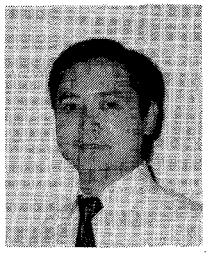
[1] D. Patterson and C. Sequin, "A VLSI RISC," *Computer*, pp. 8–21, Sept. 1982.
[2] M. Horowitz *et al.*, "A 32b microprocessor with on-chip 2KByte instruction cache," in *ISSCC Dig. Tech. Papers*, Feb. 1987, pp. 30–31.
[3] L. Kohn and S.-W. Fu, "A 1,000,000 transistor microprocessor," in *ISSCC Dig. Tech. Papers*, Feb. 1989, pp. 54–55.
[4] T. Sakurai *et al.*, "A low-power 46-ns 256-kbit CMOS SRAM with dynamic double word line," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 578–585, Oct. 1984.
[5] T. Sakurai *et al.*, "A circuit design of 32KByte integrated cache memory," in *Proc. Symp. VLSI Circuits* (Tokyo), Aug. 1988, pp. 45–46.
[6] K. Nogami *et al.*, "Architecture and design methodology of 32KByte integrated cache memory," in *Eur. Solid-State Circ. Conf. Dig. Tech. Papers*, Sept. 1988, pp. 98–101.
[7] H. Shimada, Y. Tange, K. Tanimoto, and M. Shiraishi, "An 18ns 1Mb CMOS SRAM," in *ISSCC Dig. Tech. Papers*, Feb. 1988, pp. 176–177.
[8] T. Kuroda *et al.*, "Unified design methodology and device architecture for multi-generation ASIC application," in *CICC Tech. Dig.*, May 1988, pp. 25.7.1–4.



Kazutaka Nogami was born in Oita, Japan, on May 19, 1959. He received the B.S. and M.S. degrees in applied physics from the University of Tokyo, Tokyo, Japan, in 1982 and 1984, respectively.

In 1984 he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he has been engaged in the research and development of the 1-Mbit VSRAM, integrated cache memories, and an on-chip cache macro. His current interests include application-specific memories and the memory hierarchy of the computer system.

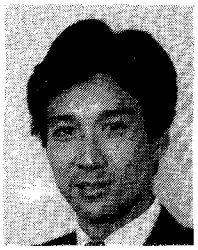
Mr. Nogami is a member of the Institute of Electronics, Information and Communication Engineers of Japan.



Takayasu Sakurai (S'77–M'78) was born in Tokyo, Japan, on January 10, 1954. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1976, 1978, and 1981, respectively. His Ph.D. work was on electronic structures of a Si–SiO_2 interface.

In 1981 he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he was engaged in the research and development of CMOS dynamic RAM and 64-kbit, 256-kbit SRAM, 1-Mbit virtual SRAM, cache memories, and an RISC with on-chip large cache memory. During the development, he also worked on modeling of wiring capacitance and delay, a new soft-error free memory cell, new memory architectures, new hot-carrier resistant circuits, arbiter optimization, and gate-level delay modeling. Since 1988 he has been a Visiting Scholar at the University of California at Berkeley, doing research in the field of computer-aided design of VLSI's. His present interests are in application-specific memories, VLSI processors, and CAD for VLSI's.

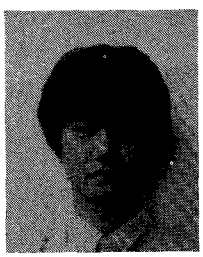
Dr. Sakurai is a member of the Institute of Electronics, Information and Communication Engineers of Japan and the Japan Society of Applied Physics.



Kazuhiro Sawada was born in Hyogo, Japan, on March 25, 1957. He received the B.S. and M.S. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1980 and 1982, respectively.

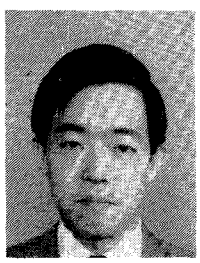
In 1982 he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he has been engaged in the research and development of the 256-kbit SRAM, the 1-Mbit virtual SRAM, the 1-Mbit DRAM embedded 72K-gate array, integrated cache memory, and on-chip large cache macro.

Mr. Sawada is a member of the Institute of Electronics, Information and Communication Engineers of Japan.



Kenji Sakaue was born in Niigata, Japan, on April 17, 1959. He received the B.S. degree in electronic engineering from the Chiba Institute of Technology, Chiba, Japan, in 1982.

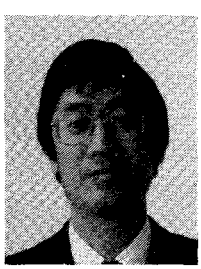
In 1982 he joined the Toshiba Microelectronics Corporation, Kawasaki, Japan, where he is engaged in the research and development of CMOS logic LSI's.



Yuichi Miyazawa was born on July 13, 1958. He received the B.S. and M.S. degrees in electronic engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1981 and 1983, respectively.

In 1983 he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he was engaged in the design of logic VLSI's, including a digital signal processor and a graphic processor. Since 1989 he has been a Visiting Scholar at Stanford University, Stanford, CA, studying digital signal processing.

Mr. Miyazawa is a member of the Institute of Electronics, Information, and Communication Engineers of Japan, and the Japan Society of Applied Physics.



Shigeru Tanaka was born in Tokyo, Japan, on January 1, 1952. He received the B.S., M.S., and Ph.D. degrees in physics from Tokyo University, Tokyo, Japan, in 1974, 1976, and 1979, respectively.

In 1980 he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he has been engaged in the research and development of the IIL device, the CMOS/SOS device, and the graphics processor. His current interests include high-performance microprocessors.

Dr. Tanaka is a member of the Information Processing Society of Japan and the Institute of Electronics, Information and Communication Engineers of Japan.

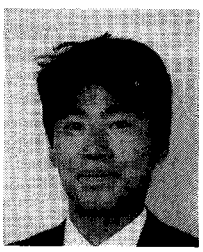
Yoichi Hiruta was born in Fukushima, Japan, on September 6, 1955. He received the M.S. degree in physics in 1981 and the Ph.D. degree in electronic material science in 1985, both from Shizuoka University, Shizuoka, Japan. His doctoral research was on the analysis of the optical absorption bands of electro-chromically colored WO_3 and MoO_3 films, using the thermo-modulation technique.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1984, where he was involved in the research and development of the process for sub- and half-micrometer CMOS technology. He is now engaged in packaging technology for VLSI's.

Dr. Hiruta is a member of the Physical Society of Japan and the Japan Society of Applied Physics.

Katsuto Katoh was born in Shimane, Japan, on October, 25, 1967. He graduated from the Masuda Technical High School, Shimane, Japan, in 1986.

In 1986 he joined the Toshiba Corporation, Kawasaki, Japan. From 1986 to 1987 he attended a one-year technical training program at the Toshiba Computer School, Kawasaki, Japan. He joined the Semiconductor Device Engineering Laboratory in 1987, where he was involved in the research and development of the process for sub- and half-micrometer CMOS technology. He is now engaged in packaging technology for VLSI's.



Toshinari Takayanagi was born in Aichi, Japan, on December 2, 1962. He received the B.S. and M.S. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1985 and 1987, respectively.

In 1987 he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he has been engaged in the research and development of CMOS LSI memory.

Tsukasa Shirotori was born in Nagano, Japan, on December 27, 1963. He received the B.S. degree in chemistry from the Kanagawa Institute of Technology, Kanagawa, Japan, in 1986.

In 1986 he joined Toshiba Microelectronics Corporation, Kawasaki, Japan. He then joined Toshiba's Semiconductor Device Engineering Laboratory, where he has been engaged in the research and development of CMOS LSI memory.

Mr. Shirotori is a member of the Institute of Electronics, Information and Communication Engineers of Japan.

Yukiko Itoh was born in Hokkaido, Japan, on December 22, 1966. She graduated from the Hakodate Commercial High School, Hokkaido, Japan, in 1985.

In 1985 she joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where she has been engaged in the development of the 1-Mbit VSRAM, integrated cache memories, and an on-chip cache macro.

Masanori Uchida was born in Tokyo, Japan, on December 26, 1964. He received the B.S. degree in precision mechanics from Tokai University, Kanagawa, Japan, in 1988.

In 1988 he joined the Toshiba Microelectronics Corporation, Kawasaki, Japan. He then joined Toshiba's Semiconductor Device Engineering Laboratory, where he has been engaged in the research and development of CMOS LSI memory.

Tetsuya Iizuka (M'79) received the B.S. degree in applied physics, and the M.S. and Ph.D. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1970, 1972, and 1975, respectively.

In 1975 he joined the Toshiba Research and Development Center, Kawasaki, Japan, where he was engaged in research on IIL and CMOS/SOS devices and in the development of dynamic RAM testing methodologies. In 1979 he joined a newly organized laboratory, the Semiconductor Device Engineering Laboratory (SDEL), of the Toshiba Corporation. He developed the first generations of 16K, 64K, 256K, and 1-Mbit CMOS SRAM's. He first applied BiCMOS sense amplifiers for 64K SRAM. He developed a new concept of SRAM called the virtually static RAM (VSRAM), an advanced concept of PSRAM. He developed application-specific memories (ASM's), such as the 1-Mbit DRAM embedded in 72K SOG and integrated cache memories. He worked on hot-carrier resistant logics. In 1981 he stayed at Hewlett-Packard Laboratories, Palo Alto, CA, as a Visiting Engineer. He studied CMOS latchup modeling and bird's beak free isolation MOS analysis. He is currently a Deputy Senior Manager, Advanced Logic and SRAM Department, SDEL. Since 1984 he has been a Lecturer at the University of Tokyo. He has also been a Lecturer at Hokkaido University since 1988.

Dr. Iizuka is a member of the Institute of Electronics, Information and Communication Engineers of Japan. He has served as a Program Committee Member for the 1985–1987 Symposiums on VLSI Technology, as an Integrated Circuit Subcommittee Member of the 1985 and 1986 IEDM's, and as a 1989–1990 ISSCC Technical Committee Member.