

# A 35 ns Cycle Time 3.3 V Only 32 Mb NAND Flash EEPROM

Yoshihisa Iwata, Ken-ichi Imamiya, Yoshihisa Sugiura, Hiroshi Nakamura, Hideko Oodaira,  
Masaki Momodomi, Yasuo Itoh, Toshiharu Watanabe, *Member, IEEE*, Hitoshi Araki,  
Kazuhito Narita, Kazunori Masuda, and Jun-ichi Miyamoto

**Abstract**—A 32 Mb NAND type flash EEPROM has been developed with 0.425  $\mu\text{m}$  CMOS technology. A 35 ns cycle time is achieved by adopting a pipeline scheme. A boosted word-line scheme and a program verify operation achieving tight threshold voltage ( $V_{th}$ ) distribution of programmed cells reduce read-out access time. Multiple block erase operation is realized by adopting erase block registers. All functions are operable with a single 3.3 V or 5 V power supply.

## I. INTRODUCTION

HIGH density flash memories are attracting much interest as the media for mass storage in portable computers and battery-powered systems. Offering high-speed programming and low-power consumption, NAND flash EEPROM's are well suited for these portable applications [1], [2]. As opposed to hot electron injection used in other architectures, the F-N tunneling current is so small that a large number of cells can be programmed or erased simultaneously. To comply with the demand of the flash type mass storage, a 32 Mb NAND type flash EEPROM [1] has been developed with 0.425  $\mu\text{m}$  CMOS technology.

To take the advantage of the NAND flash architecture, a pipeline scheme was adopted resulting in a 35 ns cycle time for data read-out or programming data-load. The boosted word-line scheme and tight  $V_{th}$  distribution of the programmed cells reduce page-read access time. In order to achieve tight programmed cell  $V_{th}$  distribution, a program verification step is required to suppress excessive programming. The novel data-register circuit [3], [4] simplifies program verification. A multiple block erase mode is convenient in many mass storage applications. In this 32 Mb NAND EEPROM, the multiple block erase mode was implemented using the erase block register [5]. All functions are operable with 3.3 V only or 5 V only power supply.

This paper describes the novel circuit techniques used in this EEPROM for the program verify operation, the pipeline scheme, and the multiple block erase operation.

Manuscript received July 28, 1995.

Y. Iwata, K. Imamiya, Y. Sugiura, Y. Itoh, T. Watanabe, H. Araki, K. Narita, K. Masuda, and J. Miyamoto are with the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Saiwai-ku, Kawasaki 210, Japan.

H. Nakamura and H. Oodaira are with the ULSI Research Center, Toshiba Corporation, Saiwai-ku, Kawasaki 210, Japan.

M. Momodomi is with the Memory Division, Toshiba Corporation, Saiwai-ku, Kawasaki 210, Japan.

IEEE Log Number 9415147.

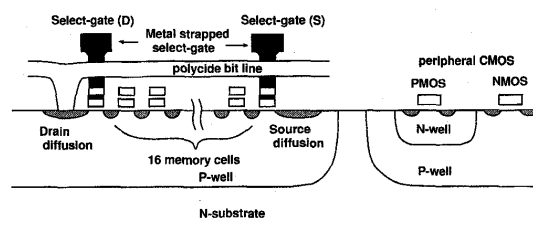


Fig. 1. Cross sectional view.

## II. PROCESS FEATURES AND THE NAND CELL

Fig. 1 shows the cross section of the NAND structure cell and peripheral CMOS logic. This device is fabricated with N-substrate double well 0.425  $\mu\text{m}$  CMOS technology. The P-well for the cell area is isolated from the P-well for the peripheral CMOS circuits. The peripheral PMOS transistors are located in N-wells enclosed within the peripheral P-well, because the N-substrate is biased at 20 V.

Both the gate length and the gate width of the memory cell are 0.425  $\mu\text{m}$ . The NAND string size is  $1.445 \times 17.255 \mu\text{m}$ , and the cell size is  $1.556 \mu\text{m}^2$  per bit. The cells are fabricated by self-aligned double poly-silicon technology. The tunnel-oxide thickness is around 10 nm. The inter-poly-silicon dielectric is an oxide-nitride-oxide (ONO) stack with a capacitive equivalent oxide thickness of 18 nm. The drain diffusion is low concentration because the drain is biased at 8 V on the condition of inhibit programming. We use the polycide bit-line to reduce the capacitive noise between the neighboring bit-lines [4]. Since polycide thickness is thin compared with Al, the capacitive coupling between the neighboring bit-lines is small. The capacitive noise is reduced to 40% compared to an Al bit-line.

Transistors with different oxide thickness are used in peripheral circuits. CMOS logic circuits use 16 nm gate oxide, whereas transistors for generating and transferring severe program/erase voltages use 42 nm gate oxide. The process features are shown in Table I.

Table II shows the cell bias conditions. Bias conditions are almost the same as previous NAND devices [1], [2]. The difference is that the P-well for the cell area is biased at  $-0.5$  V to raise the  $V_{th}$  of the parasitic word-line transistor to 20 V or more. The drain, source, gate, and back-gate of the cell are set to internal voltage, power supply voltage, or 0 V

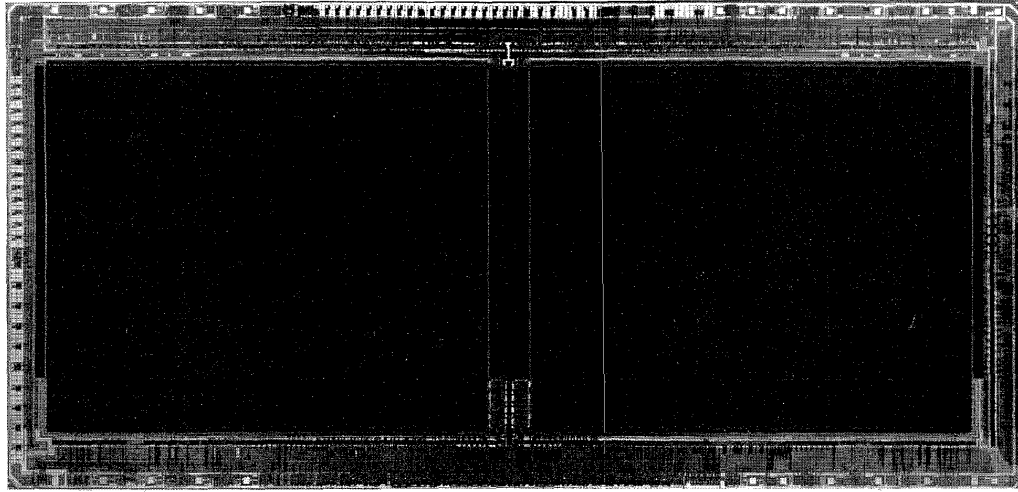


Fig. 2. Die photograph.

TABLE I  
PROCESS FEATURES

Technology:	0.425 $\mu$ m N-sub double well CMOS
	2poly,1polycide,1Al
Memory cell:	gate width/length 0.425 $\mu$ m/0.425 $\mu$ m
	tunnel oxide 10nm
	interpoly 18nm
	size 1.556 $\mu$ m <sup>2</sup>
Peripheral:	gate length 0.68 $\mu$ m(NMOS), 0.76 $\mu$ m(PMOS)
	gate oxide 16nm(8V), 42nm(20V)

TABLE II  
BIAS CONDITION

	Read	Erase Cycle		Program Cycle	
		Erase	Verify	Program	Verify
SELECT-GATE(D)	4.5	20	4.5	10	4.5
SELECT-GATE(S)	4.5	20	4.5	0	4.5
WORD-LINE(selected)	0	0	0	19	0.5
WORD-LINE(pass)	4.5	0	0	10	4.5
WORD-LINE(unselected)	0	20	0	0	0
BIT-LINE('0')	VCC(result)	F	0(result)	0	VCC(result)
BIT-LINE('1')	0(result)	F	0(result)	8	0(result)
SOURCE	0	20	0	0	0
P-WELL	0	20	0	-0.5	0

F-Floating

in program or erase operation, preventing miss-programming or miss-erasing.

### III. ORGANIZATION AND DEVICE FEATURES

Fig. 2 shows the die photograph. The die size is 7.14  $\mu$ m  $\times$  14.62  $\mu$ m. The 32 Mb memory array is divided into 2 planes,

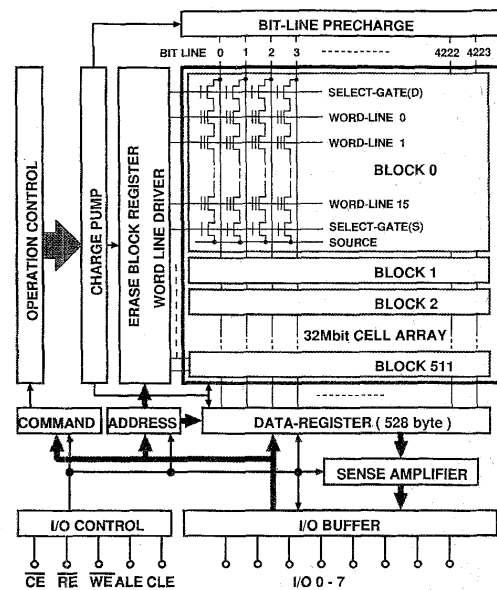


Fig. 3. Block diagram of the 32 Mb NAND EEPROM.

4096  $\times$  4224 bits each. The logical block diagram is shown in Fig. 3. The cells are organized in 512 blocks of 16 pages each. One page consists of 528 bytes or 512 bytes. Either 528 byte organization or 512 byte organization can be selected by an optional pin. In 528 byte organization, an extra 16 bytes are accessible continued with normal 512 byte access.

Each bit-line is connected to its data-register which also acts as a sense amplifier for that bit-line. The data-register temporarily saves data for serial read operation or for page programming. The data is serially transferred from or to an 8 b I/O port. The operations synchronize with external signal

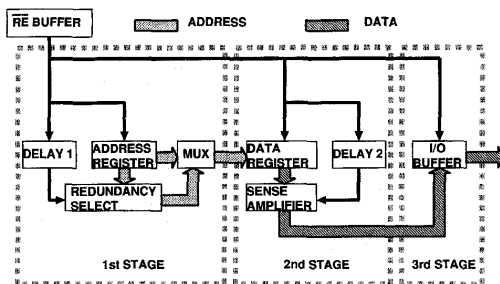


Fig. 4. Block diagram of the pipeline scheme.

RE for reading-out or WE for data-loading. The register holds a full 528 byte page. The unit of erasing is a block. Each block has a block register whose status indicates erase permit or inhibit of the block in the multiple block erase function. The erase block register is located within the row decoder.

This chip is controlled by commands. The commands, the 1st address, and programming data are all multiplexed on a single set of I/O ports, ALE (Address Latch Enable) and CLE (Command Latch Enable) signals determine the type of information entered. After the 1st address is input, subsequent addresses are generated internally as needed for serial read or data-load operation, toggling RE or WE, respectively.

#### IV. READ

Read operation in this memory is divided into two parts. One part is page read, in which the data is transferred from the memory cells to the data-registers. All 528  $\times$  8 cells connected to the same word-line are read simultaneously. The other part is serial read, in which the data is serially transferred from the data-registers to the I/O buffers. The technique of high speed read operation is described as follows.

##### A. Serial Read

In order to realize high-speed operation, the 32 Mb NAND EEPROM utilizes a three stage pipeline scheme. The block diagram and timing diagram of the pipeline are shown in Figs. 4 and 5, respectively. The read-out operation from the data-registers to the I/O buffers can be divided into three stages. Each stage is synchronized with RE. The first stage is address generation and column redundancy selection check. The second stage is column selection and read out from the selected data-register output. The third stage is data output from the I/O buffers. These three stages can be performed in parallel on the neighboring three addresses. When the data from address  $N$  is output, the data from address  $N + 1$  is sensed and address  $N + 2$  is generated.

The longest operation in these three stages determines serial cycle time. Since the third stage is the output operation only, the other two are discussed. In the first stage, the address generation is synchronized with RE. The column redundancy select circuit cannot be activated until the correct address arrives from the column address register. In the second stage, a second sense amplifier with complementary-pair input from the data-register is used to speed up the data-transfer from

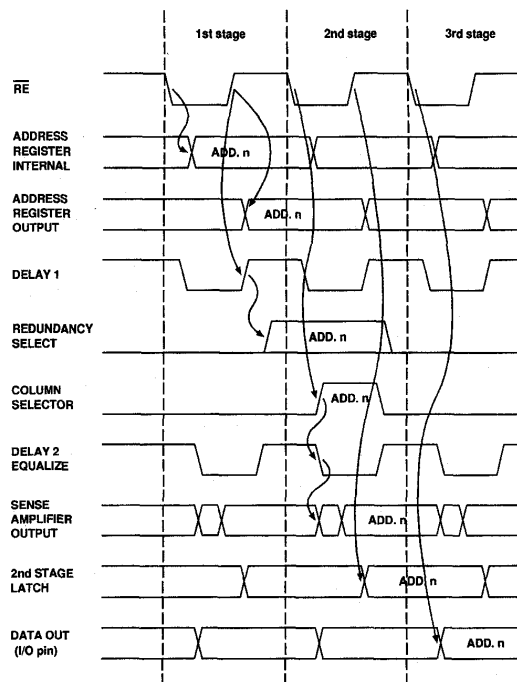


Fig. 5. Timing diagram of read-out pipeline.

the data-register to I/O buffer. The column selection signal is synchronized with RE. The sense amplifier should not be activated until the column selection signal arrives to the selected data-register. In both cases, additional timing delay from RE is necessary to ensure correct device operation. However, this may unnecessarily increase the cycle time.

Self-consistent delay circuits such as DELAY1 and DELAY2 in Fig. 4 can minimize the redundant timing delays. The DELAY1 circuit is a column address register simulator circuit designed using the same circuit elements as the real address register that consists of an ordinal binary counter. Furthermore, the interconnection between the DELAY1 and the redundancy select circuit is made identical to the one between the address register and the redundancy select circuit. The redundancy circuit is activated by simulated address arrival. The DELAY2 circuit is a logical OR function of all column selection signals. This automatically synchronizes the second sense amplifier and complementary-pair equalize circuit with the column selection signals.

The internal timing is generated self-consistently with parasitic delay, so that the pipeline has a wide operation margin against power supply voltage, temperature, and process fluctuation with no redundant time margins. A Shmoo plot for read-out cycle time at 25°C is shown in Fig. 6. A 35 ns cycle time can be achieved successfully under 3.3 V power supply.

##### B. Page Read

Fast data transfer from the cells to the data-registers is another key issue for high-speed read. At the beginning of this

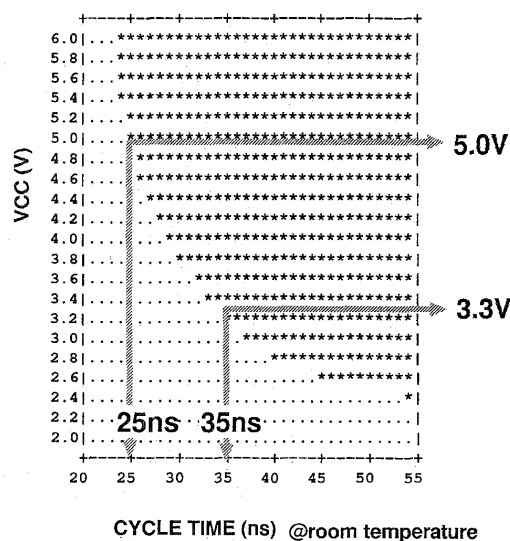


Fig. 6. Shmoo plot of serial read cycle.

operation, all the bit-lines are precharged. The pass word-lines are driven to a high level and the selected word-line is set to 0 V. After the precharge is accomplished, the select-gate connects NAND memory cells to the bit-lines, discharging them to ground level or not, according to the programmed data. All the bit-lines are sensed by the data-registers.

The read access time depends on cell current which is the function of the  $V_{th}$  of pass cell and the voltage of select-gates and pass word-line. The select-gates and pass word-lines are boosted to 4.5 V independent of the power supply voltage to obtain sufficient cell current. Sufficient cell current is also ensured by program verification, achieving tight  $V_{th}$  distribution. An  $8.5 \mu s$  page access time is achieved by a combination of these techniques. A Shmoo plot at  $25^\circ C$  is shown in Fig. 7. The access time is almost independent of power supply voltage because the internal timing of page read operation is determined by the delay circuit shown in Fig. 8. The delay time is determined by the product of  $R$  and  $C$ .

#### V. PROGRAM

Program operation is divided into three parts. The first is data-load, in which the data is serially transferred from I/O buffers to the data-registers. The second part is program. All  $528 \times 8$  cells connected to the same word-line are programmed simultaneously. The third part is program verify. In the NAND flash structure, the memory cells are connected in series so the  $V_{th}$  of each programmed memory cell must be kept between the selected word-line voltage and the pass word-line voltage. If the  $V_{th}$  of the pass cell in the NAND series exceeds the pass word-line voltage, other cells in same NAND series cannot be read out. In order to suppress excessive programming and to obtain tight  $V_{th}$  distribution for high-speed page-read, program verification is required.

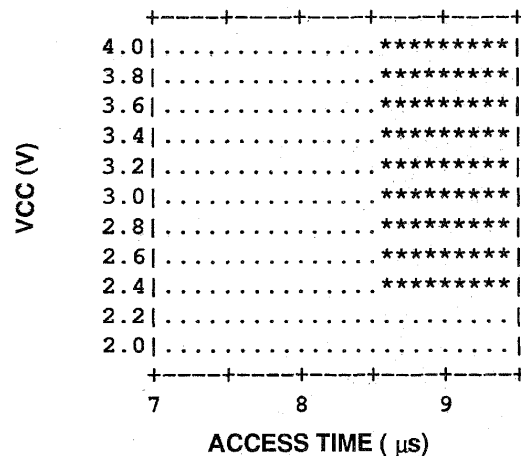


Fig. 7. Shmoo plot of page-read access time.

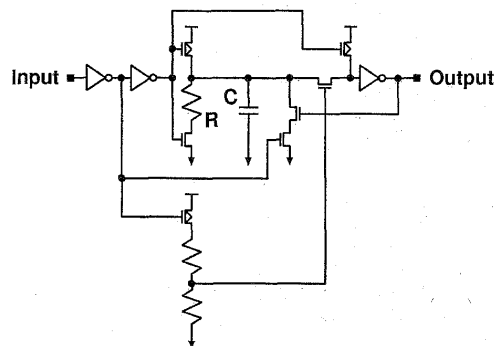


Fig. 8. RC delay circuit.

#### A. Data-Load

Since the data-load operation is performed after the 1st address is provided, a two-stage pipeline scheme is adopted. The timing diagram is shown in Fig. 9. The first stage is the same as serial read operation, and the DELAY1 acts in the same manner. The second stage is data-loading to the data-registers. The data-load has loose timing compared with serial read operations. A 35 ns cycle time can be achieved successfully under 3.3 V power supply.

#### B. Program Verification

A simultaneous program verification of all  $528 \times 8$  cells connected to the same word-line was previously introduced [2]. The states of the data-registers change with ongoing programming and verification as shown in Fig. 10. At first, the data to be programmed is serially loaded to the data-registers in a data-load cycle initiated by the data-load command. "0" represents the program state, and "1" represents the program-inhibit state. By program command, the cells are programmed according to the contents of the registers. During program verification, the data-register state of the successfully pro-

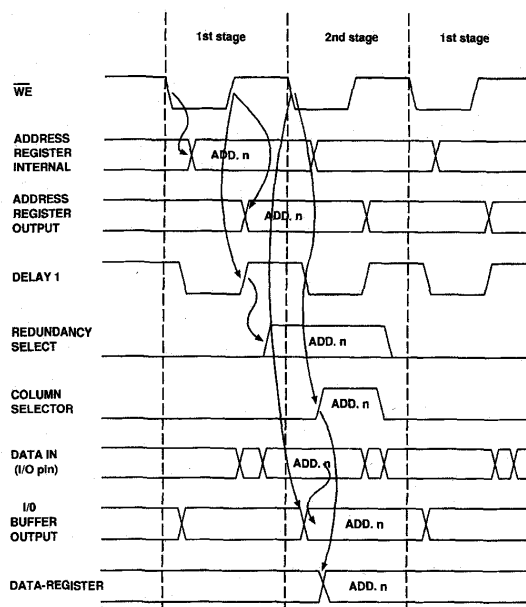


Fig. 9. Timing diagram of data-load pipeline.

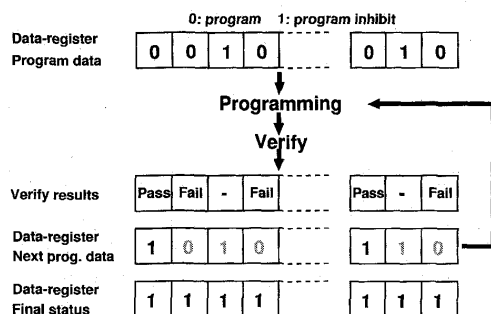


Fig. 10. Program sequence.

grammed cell is changed from "0" to "1." The programming sequence cycles until all the registers are set to "1." Since the cell whose data-register is changed from "0" to "1" is not over-programmed in subsequent program cycles, tight  $V_{th}$  distribution is achieved.

In this 32 Mb NAND EEPROM, the program verification step is simplified by the novel data-register [3] shown in Fig. 11. Using a previous technique [2], the verify operation was complex, consisting of verify read operation, latched data feed-back, and data modification for the next programming step. With the new circuit, the data for the next programming cycle is modified at the same time as the sense operation, and the verify read operation remains almost the same as a normal read operation. The operation is described as follows.

The data-register state controls the bit-line for cell programming activated by the "LOAD" signal. A low level (0 V) at N1 corresponds to programming (electron injection),

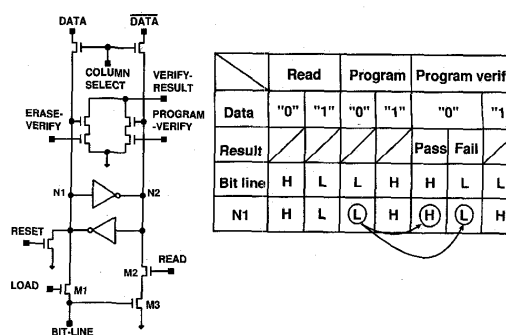
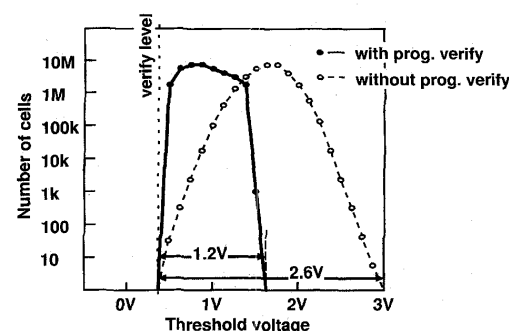


Fig. 11. The data-register.

Fig. 12. Measured programmed cell  $V_{th}$  distribution.

and a high level (8 V) corresponds to programming inhibit. The program inhibit voltage of 8 V is generated internally as is the program voltage of 19 V. After programming is performed, with a 19 V word-lines, the programmed cells  $V_{th}$  are verified, and the state of the data-registers corresponding to sufficiently programmed cells are changed to the program inhibit state. The program verify operation is the same as a normal read operation although the selected word-line voltage is 0.5 V to guarantee sufficient  $V_{th}$  of programmed cells, and the initial state of the register in normal read is determined by the "RESET" input. The bit-line voltage is detected by transistor M3. The bit-line of a sufficiently programmed cell is kept at a high level to set N2 to 0 V once activated by the "READ" signal. Thus, the data-register status of a sufficiently programmed cell changes to the program inhibit state. The "VERIFY-RESULT" input is obtained by a wired OR function of the register output, N2. The state of this signal determines whether another re-programming operation is required or not. The programmed cell  $V_{th}$  distribution width is reduced from 2.6 V to 1.2 V by this verify scheme as shown in Fig. 12.

## VI. MULTIPLE BLOCK ERASE

The multiple block erase algorithm is shown in Fig. 13 [5]. Block address registers are set in repeated block address input cycles. The erase sequence is initiated by the erase command. The word-line voltage is set to 0 V for erase and about 20 V for erase inhibit according to the erase block register. The P-well

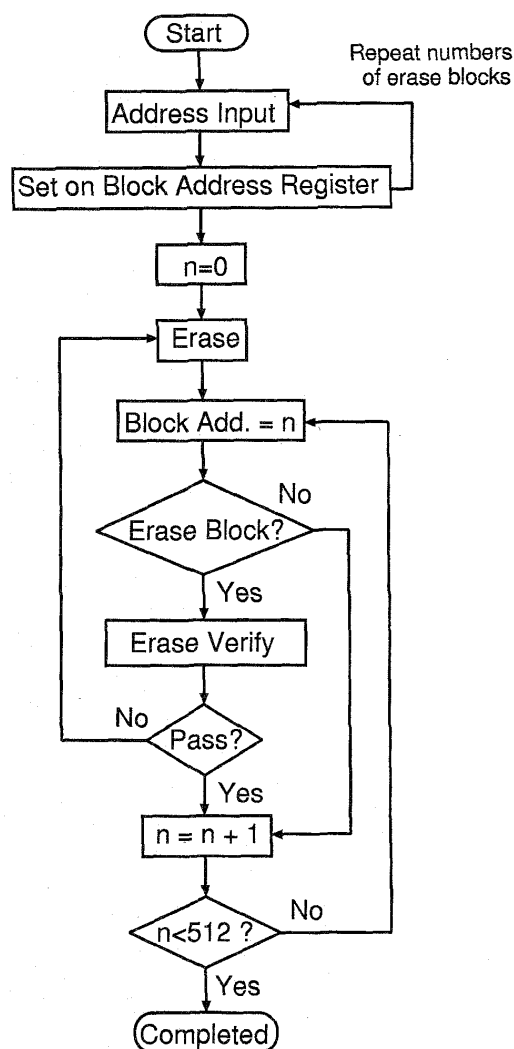


Fig. 13. Multiple block erase algorithm.

of the cell array is biased at about 20 V. Any combination of blocks can be erased simultaneously because of the low F-N tunneling currents.

The erase-verify operation is performed block-by-block after the erase operation according to the erase block register. The condition of the erase block register is read out from block 0 to block 511 in order. If the condition is erase inhibit, the condition of the next block address is read out. If the condition is erase permit, a verify read is performed. In erase verify, all the word-lines of the selected block are set to 0 V allowing all  $16 \times 4$  k cells to be verified simultaneously. This scheme reduces erase verify time. The verify result is detected by a wired OR function of all the data-register outputs in the same manner as program verify. If the verify result is pass, the condition of next block address is read out. If the verify result

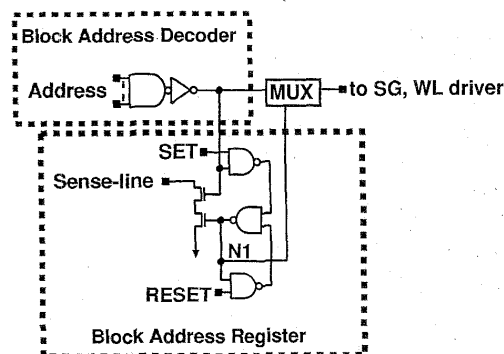


Fig. 14. Erase block register.

TABLE III  
CHIP CHARACTERISTICS

Organization:	528bit $\times$ 8 $\times$ 8k pages
Erase block size:	8k byte
Die size:	7.14mm $\times$ 14.62mm
Power supply voltage:	3.3V-only or 5V-only
I/O cycle time:	35ns
Serial access time:	28ns
Page access time:	8.5 $\mu$ s
Program time:	400ns/byte
Erase time:	2.5ms one block 5.0ms whole chip
Current(program):	30mA
Current(Program):	< 1 $\mu$ A

is fail, all of the erase permit block is re-erased. This step is repeated until block 511 is checked and erased if necessary.

The erase block register is shown in Fig. 14. During a block address input cycle, "SET" is "H" and "RESET" is "L." If "N1" of the block selected by the address input is "H," the condition of the block is erase permit. "N1" of unselected blocks is kept "L" to inhibit erase. After sense-line precharge is accomplished, the state of "N1" is read out by discharging the sense-line or not according to the state of "N1" at address selection.

## VII. CONCLUSION

A 32 Mb NAND type flash EEPROM has been successfully developed using a 0.425  $\mu$ m CMOS process. Table III shows the memory characteristics. A serial cycle time of 35 ns at

3.3 V is accomplished using a three-stage pipeline for read-out and a two-stage pipeline for data-load. Program time is 400 ns per byte, and a whole chip can be erased in 5 ms. The programmed cell threshold voltage spread is 1.2 V using program verification and a novel data-register. A power supply voltage is 3.3 V only or 5 V only can be selected by metal mask option. The cell bias conditions of a 5 V device is equal to that of the 3.3 V device, except for the precharge voltage of bit-line that is set to the power supply voltage in read operation.

The 32 Mb NAND EEPROM serves of portable mass storage applications such as solid state disk with high-speed programming and erasing and low-power consumption.

#### ACKNOWLEDGMENT

The authors thank H. Minagawa, E. Kozuka, N. Suzuki, Y. Tatsumi, Y. Takeyama, H. Oota, T. Tanaka, H. Nakai and T. Yamamura for circuit design, H. Hagiwara, M. Abe, and N. Matsui for process design, and K. Hashimoto and N. Endo for their encouragement.

#### REFERENCES

- [1] M. Momodomi *et al.*, "A 4-Mb NAND EEPROM with tight programmed  $V_{th}$  distribution," *IEEE J. Solid-State Circuits*, vol. 26, pp. 492-496, Apr. 1991.
- [2] T. Tanaka *et al.*, "A quick intelligent program architecture for 3 V-only NAND-EEPROM's," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1992, pp. 20-21.
- [3] J. Miyamoto *et al.*, U.S. Patent pending.
- [4] H. Nakamura *et al.*, "A novel sense amplifier for flexible voltage operation NAND flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1995, pp. 71-72.
- [5] H. Nakai *et al.*, U.S. Patent ~5,371,702.



**Yoshihisa Iwata** was born in Aichi, Japan, on May 16, 1960. He received the B.S. degree in physics from Nagoya University, Nagoya, Japan, in 1983.

He joined the Toshiba Research and Development Center, Kawasaki, Japan, in 1983, where he was engaged in the circuit design of DRAM's and EEPROM's. In 1990, he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he has been engaged in the research and development of EEPROM's.



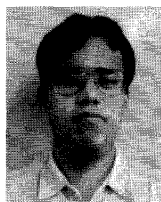
**Ken-ichi Imamiya** was born in Tokyo, Japan, on July 5, 1961. He received the B.S. degree in physics from Tokyo University, Tokyo, Japan, in 1986.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1986, where he has been engaged in the circuit design of EPROM's and EEPROM's.



**Yoshihisa Sugitara** was born in Aichi, Japan, on June 23, 1965. He received the B.E. and M.E. degrees in material science from Tsukuba University, Ibaragi, Japan, in 1988 and 1990, respectively.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1990, where he has been engaged in the circuit design of EEPROM's.



**Hiroshi Nakamura** was born in Yamaguchi, Japan, on August 16, 1965. He received the B.E. and M.E. degrees in electronic engineering from Osaka University, Osaka, Japan, in 1988 and 1990, respectively.

He joined the Toshiba ULSI Research Center, Kawasaki, Japan, in 1990. Since then, he has been engaged in the circuit design of EEPROM's.



**Hideko Oodaira** was born in Aomori, Japan, on December 31, 1967. She graduated from Hirosaki Technical High School, Japan, in 1986.

She joined the Toshiba Research and Development Center, Kawasaki, Japan, in 1986. Since then, she has been engaged in the circuit design of EEPROM's at the Toshiba ULSI Research Center, Kawasaki, Japan, in 1990.



**Masaki Momodomi** was born in Fukuoka, Japan, on March 11, 1958. He received the B.E. degree in electronic engineering from Kyushu University, Japan, in 1980.

He joined the Toshiba Research and Development Center, Kawasaki, Japan, in 1980, where he was engaged in the circuit design of DRAM's and EEPROM's. He is now with the MOS Division, Toshiba Corporation, Kawasaki, Japan, and is engaged in the development of EEPROM's.

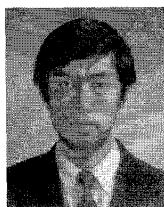


**Yasuo Itoh** was born in Osaka, Japan, on August 1, 1953. He received the B.E. and M.E. degrees in electronic engineering in 1977 and 1979, respectively, and the Ph.D. degree in electrical engineering in 1982, all from Osaka University, Japan.

He joined the Toshiba Research and Development Center, Kawasaki, Japan, in 1982, where he was engaged in the circuit design of DRAM's and EEPROM's. In 1990, he joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, where he has been engaged

in the research and development of EEPROM's.

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



**Toshiharu Watanabe** (M'92) was born in Niigata, Japan, on July 7, 1956. He received the B.S. and M.S. degrees in applied physics from the University of Tokyo, Japan, in 1980 and 1982, respectively. He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1982, where he has been engaged in the development of high density DRAM's and EEPROM's.



**Kazunori Masuda** was born in Fukuoka, Japan, on February 11, 1970. He graduated from Kokura Technical High School, Japan, in 1988.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1988, where he has been engaged in the development of high density EPROM's and EEPROM's.



**Hitoshi Araki** was born in Tokyo, Japan, on September 4, 1962. He received the B.E. and M.E. degrees in electronic engineering from Sizuoka University, Japan, in 1985 and 1987, respectively.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1987, where he has been engaged in the development of high density EPROM's and EEPROM's.



**Kazuhito Narita** was born in Aomori, Japan, on August 29, 1967. He graduated from Towada Technical High School, Japan, in 1986.

He joined the Semiconductor Device Engineering Laboratory, Toshiba Corporation, Kawasaki, Japan, in 1982, where he has been engaged in the development of high density EPROM's and EEPROM's.



**Jun-ichi Miyamoto** He received the B.S., M.S., and Ph.D. degrees in electronic engineering from Kyoto University, Kyoto, Japan in 1973, 1975, and 1989, respectively.

In 1975, he joined Toshiba Research and Development Center, Kawasaki, Japan, and worked on the circuit design of ECL memory and bipolar PROM. Since 1982, he has been engaged in the development of 64 Kb CMOS static RAM with bipolar sense amplifiers at the Toshiba Semiconductor Device Engineering Laboratory (SDEL). From 1985 to 1986,

he was a visiting research scholar of the Computer Systems Laboratory at Stanford University, Stanford, CA, where he designed an application specific memory. After returning to the SDEL, he designed 4 Mb and 16 Mb EPROM and managed both NAND and NOR type Flash Memory development. Currently, he is a Chief Specialist with the Advanced Memory Development Department at SDEL.

Dr. Miyamoto is a member of the Institute of Electronics, Information and Communication Engineers of Japan. He served on the program committee of the 1992 and 1993 Symposium on VLSI Circuits.