TP 2.5: A 1Mb 2-Transistor/bit Non-Volatile CAM Based on Flash-Memory Technologies

Tohru Miwa, Hachiro Yamada, Yoshinori Hirota, Toshiya Satoh, Hideki Hara

NEC Corporation, Kanagawa, Japan

A 1 Mb content-addressable memory LSI based on flash technologies (flash CAM) has memory cells consisting of a pair of flash memory cell transistors. 10.34μm² cell and 42.9mm² die are attained with 0.8μm design rules. The flash CAM can be searched for masked binary data. Read access time and search access time are 115ns and 145ns, respectively, with a 5V supply voltage. Power dissipation is 200mW at 3.3MHz. The flash CAM cell consists of two floating-gate transistors. This structure is in strong contrast to the comparator-added-storage structure of 17-transistor SRAM-based cells or five-transistor two-capacitor of DRAM-based cells. In addition to non-volatility, flash CAMs also feature on-board programmable/erasable memory.

A CAM cell array using floating-gate transistors is discussed in [4]. However, LSIs based on that type of array are impractical. The complicated multi-valued logic required means that margins are insufficient. This problem is solved by the flash CAM, based on binary logic rather than on multi-valued logic.

In addition to the cell structure, four circuit approaches contribute to performance: 1) small-area, low-power, current differential sense amplifier for highly-parallel search operations; 2) search management circuit for on-chip result management and extra long word search; 3) high-speed priority encoder with disable signal bypass; and 4) word-line redundancy circuits for higher production yields.

Figure 1 shows the flash CAM cell configuration. The threshold voltage of each floating gate transistor (M1, M2) is programmed to be either under 3V or over 7V depending on the value of the data stored. For an operation, each bit line is set at 0V or 5V depending on the value of the data being searched for along that line. Figure 1 shows memory cell transistor action for all possible combinations of stored and search data values. When any transistor is ON, a non-match is detected. Matches are detected only when both transistors are OFF.

The small-area, low-power, nine-transistor sense amplifier for highly-parallel is shown inFigure 2. It is a latch-type current-differential amplifier. Because it contains no current mirror amplifier, power dissipation is quite low. Current detection is as follows. First, the selected word lines are precharged to about 1V and the output nodes (TOUT, BOUT) are equalized at high (/Precharge = low and latch = low). After /Precharge goes high, the current difference is amplified to the voltage difference between the output nodes. Finally, the sensed result is latched. After detection, the sense amplifier holds the result and dissipates no power until the next detection.

In search in the left array, as shown in Figure 3, each search sense amplifier compares the current on the left word line with the reference current on the right word line. Each word line has a reference memory cell that leaks half current that an ON cell transistor does. In the flash CAM, 253 search sense amplifiers are located between the right and left cell arrays. They occupy only 0.7% of the die area in total. Each occupies 10.55μm (four-word-line space) x 110μm. Reference cells are the same size as normal cells and are laid out in the cell arrays. In wafer test, threshold voltages of reference cells are set at 4V by read/write circuits.

CAMs perform retrieval formula processing and extra long word search with on-chip parallel execution of logic functions on search results. A search management circuit, shown in Figure 4, performs a logic function on the output of the search sense amplifier (I) and the previous search result (O) (retained) (Figure 4). It retains the result as a new search result (O').

As search becomes more parallel, delay in the priority encoder, that arbitrates among multiple matched words, constitutes a larger part of total search operation delay. A high-speed priority encoder is used, consisting of a 258-input priority selector and an address encoder. The priority selector consists of 33 serially-connected sub-circuits (32-eight-input and one two-input priority selectors). It is provided with eight four-sub-circuit-skip bypaths in addition to sequential paths. The bypass reduces encoding delay from 34ns to 19ns.

In the flash CAM LSI, 256 + 2 (redundancy) search sense amplifiers are laid out in a line from the top of the chip to the bottom, and physical addresses are assigned to each, as shown in Figure 5a. In an ordinary redundancy scheme, when a failed word-line is detected, its amplifier will be replaced by one of the two bottom-most (redundancy) amplifiers. This could create a logical inconsistency, however, when multiple matches are detected. Word line redundancy circuits avoid such inconsistencies. The circuit in Figure 5b generates logical addresses on the basis of the physical addresses in the sense amplifiers. When a failed word line is detected, it simply subtracts one or two from subsequent physical addresses as appropriate to create logical addresses. This covers a failure by shifting all amplifiers below it up one position, maintaining proper order and avoiding inconsistency. The circuit in Figure 5c operates as a decoder to reverse this process by adding one or two to logical addresses as appropriate. It recreates physical address from adjusted logical address.

Figure 6 is a 1Mb CAM chip micrograph. Word organization is 128b x 8k words, and array organization is 1k word lines x 1k bit lines x 2 planes. 50k peripheral transistors and 2M floating gate transistors are integrated in 7.14x6.01mm². Figure 7 shows shmoo plots of read access time and search access time versus power supply voltage (Vcc). At Vcc of 5V, read access time is 115ns and search access time is 145ns. Chip characteristics are summarized in Table I.

Acknowledgments:

The authors thank members of Microelectronics Labs. for discussion, and M. Koike, Y. Okuyama, M. Takada, M. Fukuma for support.

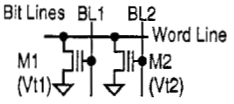
References:

[1] Ogura, T., et al., "A 20kbit associative memory LSI for artificial intelligence machines," IEEE J. Solid-State Circuits, Vol. 24, pp. 1014-1020, Aug., 1989.

[2] Yamagata, T., et al., "A 288-kb fully parallel content addressable memory using a stacked-capacitor cell structure," IEEE J. Solid-State Circuits, Vol. 27, pp. 1927-1933, Dec., 1992.

[3] Wade, J. P., C. G. Sodini, "A Ternary Content Addressable Search Engine," IEEE J. Solid-State Circuits, Vol. 24, pp. 1003-1013, Aug., 1989.

[4] Aragaki, S., et al., "A high-density multiple-valued content addressable memory based on one transistor cell," IEICE Trans. Electron., Vol. E76-C, No. 11, pp. 1649-1656, 1993.



Stored Data (Vt1, Vt2)	Search Data (VBL1, VBL2) "1" (5V, 0V)	"0" (0V, 5V)	"don't care" (0V, 0V)
"1" (>7V, <3V)	OFF, OFF (match)	OFF, ON (non-match)	OFF, OFF (match)
"0" (<3V, >7V)	ON, OFF (non-match)	OFF, OFF (match)	OFF, OFF (match)
"masked" (>7V, >7V)	OFF, OFF (match)	OFF, OFF (match)	OFF, OFF (match)

Figure 1: Flash CAM cell configuration and action.

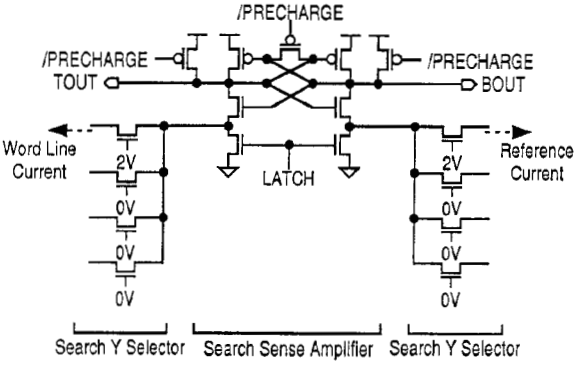


Figure 2: Search sense amplifier circuit configuration.

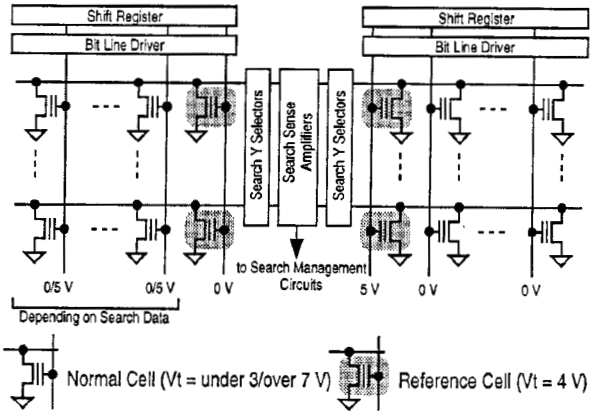
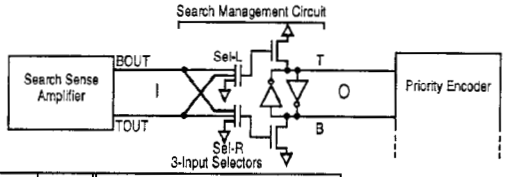


Figure 3: Search operation in the left array.



Sel-L	Sel-R	Logic Operation	
BOUT	TOUT	Through	O'=I
TOUT	BOUT	Not	O'=not(I)
BOUT	GND	AND	O'=and(O, I)
GND	TOUT	OR	O'=or(O, I)
TOUT	GND	SUB	O'=and(O, not(I))
GND	BOUT		O'=or(O, not(I))
GND	GND	HOLD	O'=O

I: output of the search sense amplifier
O: previous search result
O': new search result

Figure 4: Search management circuit.

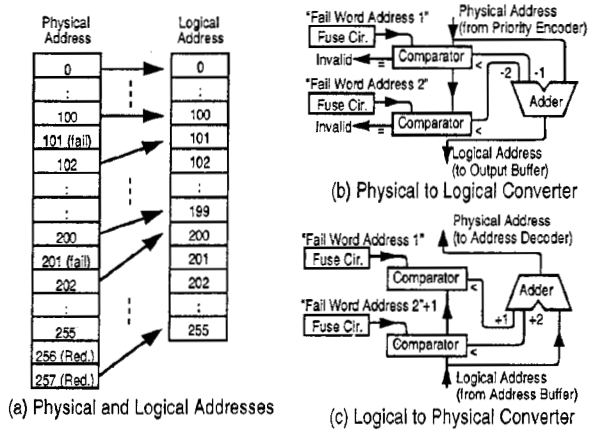


Figure 5: Word line redundancy circuit.

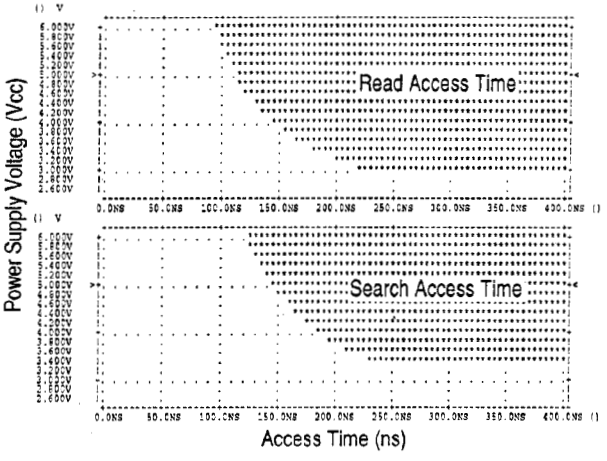


Figure 7: Read and search access time versus power supply voltage.

Figure 6, Table 1: See page 414.

(Continued from page 39)

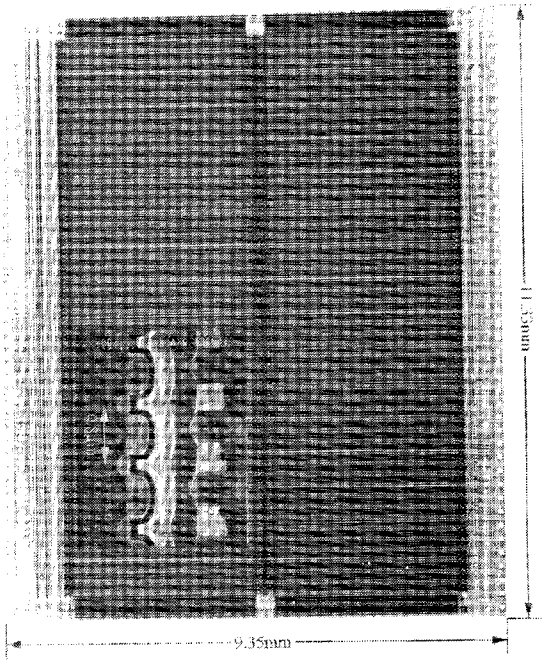


Figure 8: Chip and self-aligned AND cell micrograph.

Density	128Mb
Organization	x8
Powers supply	2.5V
Chip area	105.9mm²
Current dissipation (read)	25mA
1/4 sector access time	500ns
Burst mode cycle time	20ns
Sector (512B) erase time	1ms (typ.)
Sector program time	1ms (typ.)
Memory cell	Self-aligned AND, 0.8x0.5μm²
MOS Tox	10nm(Vcc, Vp) / 25nm(VH)
Lg	0.45/0.5μm(nMOS/pMOS,Vcc,Vp)/ 0.8μm(VH)
Vth	0.25V(nMOS, Vcc/Vp/VH) / - 0.25V(pMOS, Vcc/Vp/VH)
Technology	0.25μm CMOS, p-substrate, 3- well 3-poly, 1-polycide, 3-metal

Table 1: Chip performance.

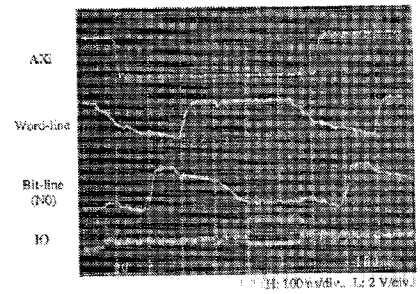


Figure 9: Operating waveforms.

TP 2.5: A 1Mb 2-Transistor/bit Non-VolatileCAM Based on Flash-Memory Technologies

(Continued from page 41)

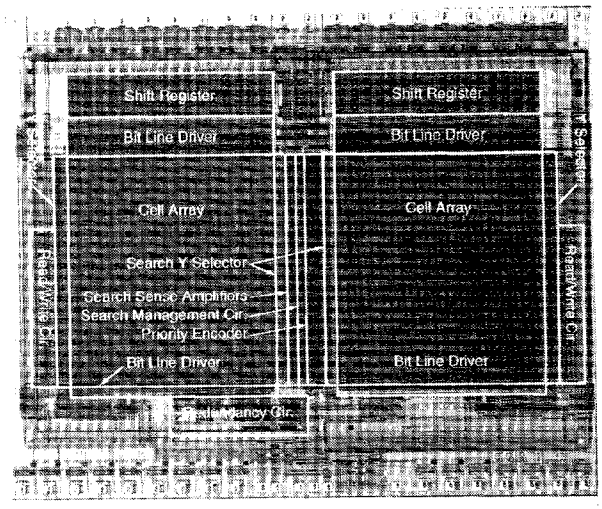


Figure 6: Chip micrograph.

Organization	1Mb (128b x 8k word)
Supply voltages	Vcc (5V), Vpp (12V)
Technology	0.8μm double-poly and double-metal CMOS flash memory
Die size	7.14x6.01mm²
Cell size	4.7μm (bit-line pitch x 2) x2.2μm (word-line pitch), W/L = 1.0μm/0.75μm
Read-access time	130ns
Search-access time	145ns
Read-active power	50mW + 160mW (I/O)(300ns cycle)
Search-active power	140mW (300ns cycle)

Table 1: Chip characteristics.