4-Mb MOSFET-Selected Phase-Change Memory Experimental Chip

F. Bedeschi[1], R. Bez[2], C. Boffino[3,4], E. Bonizzoni[3,4], E. Buda[1], G. Casagrande[1], L. Costa[1], M. Ferraro[1], R. Gastaldi[1], O. Khouri[1], F. Ottogalli[2], F. Pellizzer[2], A. Pirovano[2], C. Resta[1,4], G. Torelli[3], and M. Tosi[2]

[1] STMicroelectronics, Memory Product Group, via Olivetti, 2 – 20041 Agrate Brianza, Italy
[2] STMicroelectronics, Central R&D, via Olivetti, 2 – 20041 Agrate Brianza, Italy
[3] Department of Electronics, University of Pavia, via Ferrata, 1 – 27100 Pavia, Italy
[4] Studio di Microelettronica, STMicroelectronics & University of Pavia, via Ferrata, 1 – 27100 Pavia, Italy

Abstract:

This paper presents a 4-Mb Phase-Change Memory experimental chip using an MOS transistor as a cell selector. A cascode bit-line biasing scheme allows read and write voltages to be fed to the storage element with adequate accuracy. The chip was integrated with 3-V 0.18-µm CMOS technology and experimentally evaluated. A read access time of 45 ns was measured together with a write throughput of 5 MB/s, which represents an improved performance as compared to present NOR Flash memories. Cell current distributions on the 4-Mb array proved chip functionality and a good working window, thus demonstrating the feasibility of a stand-alone Phase-Change Memory with standard CMOS fabrication process.

1. Introduction

Today high-performance portable equipments demand for non-volatile memories featuring higher and higher read/write speed and endurance. Phase-Change Memory (PCM) [1], [2] is a very promising technology to meet these requirements. In fact, PCMs ensure very fast read and reprogramming operations as compared to currently dominant Flash memories, together with high endurance and excellent compatibility with standard CMOS fabrication processes. A further key advantage is very fine write granularity, as any bit can be independently reprogrammed with no need for block erasing.

This paper presents a 4-Mb PCM experimental chip developed to demonstrate the feasibility of a PCM device fabricated by using a standard CMOS process. As shown in Fig. 1, the cell selector is implemented by using an *n*-channel MOSFET. Even though a substrate *pnp* bipolar junction transistor (BJT) can also be employed as a cell selector [2], [3], the use of an MOS device reduces the number of lithographic masks required, thus ensuring lower process cost. In addition, this choice eliminates the problem of the cumulative array leakage current due to the reverse-biased base-to-emitter junction of unaddressed BJT selectors. Furthermore, implementing the cell selector with the same (NMOS) device type used in peripheral circuits, allows research efforts to be concentrated on cell technology development.

The proposed experimental chip was integrated with a 3-V 0.18-µm CMOS technology and experimentally evaluated, thus allowing the technology performance to be assessed.

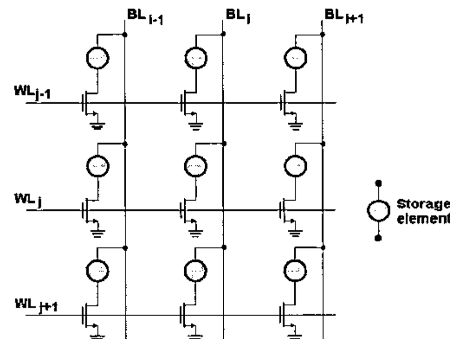

Fig. 1 – Detail of PCM array configuration (*WL* = wordline; *BL* = bitline).

2. Phase-change storage element and chip architecture

In PCMs, also referred to as Ovonic Unified Memories (OUMs), the storage device is made of a thin film of chalcogenide alloy (in our case, $Ge_2Sb_2Te_5$, GST). This material can reversibly change between an amorphous (high impedance, RESET state) and a polycrystalline (low impedance, SET state) phase when thermally stimulated, thus allowing information storage. The phase conversion of a storage element is obtained by appropriately heating (by means of electrical pulses applied to a suitable heater element) and then cooling a small, thermally isolated portion of the chalcogenide material. Once the chalcogenide material melts, it completely loses its crystalline structure. When rapidly cooled, the chalcogenide material is locked into its amorphous state (to this end, the cooling operation rate has to be faster than the

0-7803-8480-6/04/$20.00 ©2004 IEEE

crystal growth rate). To switch the memory element back to its crystalline state, the chalcogenide material is heated to a temperature between its glass transition temperature and its melting point temperature. In this way, nucleation and micro-crystal growth occur in several ns, thus leading to a (poly)crystalline state and, hence, to a percolation path for the conduction.

From above, it is apparent that the storage element can be modelled as a programmable resistor (high resistance = logic 0; low resistance = logic 1). Reading a cell basically consists in measuring the resistance of the addressed storage device. To this end, a predetermined voltage is forced across the storage element of the selected cell, and the ensuing current flow is sensed. In practice, the cell current is compared to a reference current provided by an identical cell programmed to a suitable resistance value. To obtain the best sense accuracy, the reference cell is located within the memory array and is identically biased. This choice minimizes boundary effects and allows thermal tracking between the reference and the array cells.

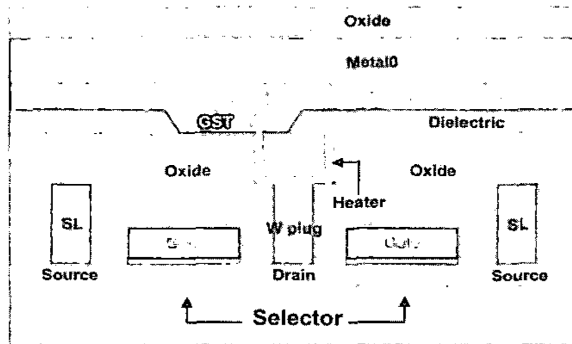

Fig. 2 – Schematic cross-section of the PCM/MOS array.

A schematic cross-section of the cell array along one bitline (which is realized in the lowest metal level, referred to as metal0) is depicted in Fig. 2. The storage element consists of a small portion of GST alloy deposited in a rectangular microtrench, and is in physical contact with a thin vertical metallic heater [4]. The latter is connected to the drain region of the MOSFET selector by means of a tungsten plug, while the selector gate is connected to the metal2 wordline (not shown). The tungsten source line (SL) is grounded at the borders of the array. A metal1 strap every 64 cells is provided to minimize the overall line resistance. Indeed, the SL resistance affects the gate-to-source voltage of the MOSFET selector in a different way depending on the cell position in the array. This effect particularly impacts on the accuracy and reliability of programming operations, during which a considerable amount of current flows through the storage element. Fig. 3 shows a SEM (Scanning Electron Microscope) microphotograph of the cell cross-section. Tab. 1 summarizes the read and write currents and voltages for selected (Sel) and unselected (No Sel) cells, which are arranged in the memory array as depicted in Fig. 1.

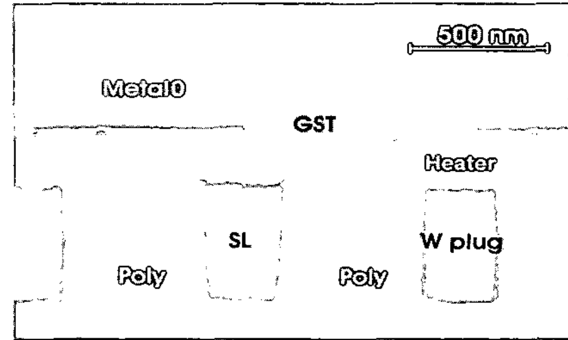

Fig. 3 – SEM microphotograph of the cell cross-section.

		READ		SET		RESET	
		V(V)	I(μA)	V(V)	I(μA)	V(V)	I(μA)
WL	Sel	1.8	0	3	0	3	0
	No Sel	0	0	0	0	0	0
BL	Sel	0.4	0-80	1.5	300	2.7	600
	No Sel	0	0	0	0	0	0

Tab. 1 – Current and voltages for selected (Sel) and unselected (No Sel) cells.

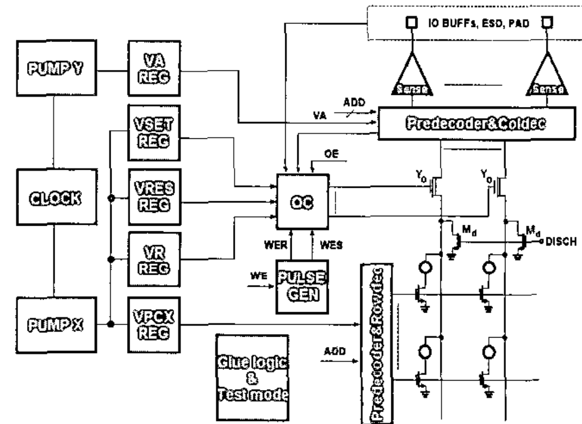

Fig. 4 – Schematic block diagram of the experimental chip.

A schematic block diagram of the proposed experimental chip is illustrated in Fig. 4. The memory is organized in a single 4-Mb array (2048 rows ~ 2048 columns). Natural transistors Y_O (which operate in the saturation region) regulate the bitline voltage to ~400 mV during reading, and to ~1.5 V and ~2.7 V during the SET and the RESET phase, respectively, to force the required current through the cell in each operation (Tab. 1). In read mode, the chosen cascode bitline biasing approach allows fast bitline precharge and sensing. Furthermore, during read operations, the bitline voltage has to be adequately low, accurate, and stable in order not to disturb the state of the cell. In this respect, the adopted bitline biasing technique prevents the risk of spurious SET pulses since the cascode

structure rejects noise injection from the column decoder supply line, referred to as V_A in Fig. 4.

SET and RESET operations are performed without resorting to program&verify and erase&verify techniques [5], [6], so as to reduce reprogramming time. It is therefore apparent that very accurate electrical pulses have to be applied to the addressed bitlines and, hence, to the selected cells. In write mode, a pulsed, regulated bitline voltage is obtained by simply controlling the gates of transistors Y_O (no DC current drawing from the corresponding regulator is required). Noise rejection ensured by the cascode structure also guarantees high accuracy of the applied write voltage. Transistors M_d discharge all bitlines after any read and write operation.

The Operation Control (OC) block provides the read and write regulated voltage signals to the gate of transistors Y_O. Conventional structures [7] were adopted to regulate SET, RESET, and read voltages, referred to as V_{SET}, V_{RESET}, and V_{READ}, respectively, as well as voltages V_{PCX} (wordline decoder supply voltage) and V_A (bitline decoder and cell current supply voltage).

Reprogramming operations require voltages higher than the nominal supply V_{dd}. For this reason, two voltage tripler charge pumps [8] were integrated. Charge pump X supplies voltage regulators for V_{SET}, V_{RESET}, V_{READ}, and V_{PCX}, while a separated voltage tripler (charge pump Y) is devoted to regulator V_A in order to achieve the driving capability required to provide the cell current in write operations.

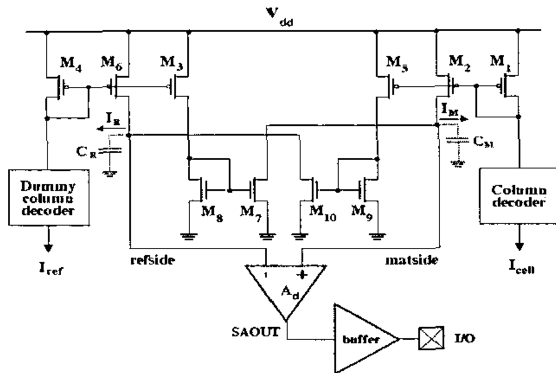

Fig. 5 – Circuit diagram of the sense amplifier.

A fully symmetrical sense amplifier topology (Fig. 5) [9] was developed to ensure zero systematic offset together with adequate rejection of disturbs due to capacitive coupling with noisy substrate, power supply, and ground. After bitline precharge and equalisation (circuitry not shown in the figure), the current differences $I_M = I_{cell} - I_{ref}$ and $I_R = I_{ref} - I_{cell}$ are taken (where I_{cell} and I_{ref} are the addressed cell current and the reference current). The current differences I_M and I_R are integrated on the parasitic capacitances of nodes *matside* and *refside*, C_M and C_R, respectively. The ensuing voltages are then compared by means of block A_d, which produces a latched output signal *SAOUT*. The latter is fed to the I/O pad by means of an output buffer.

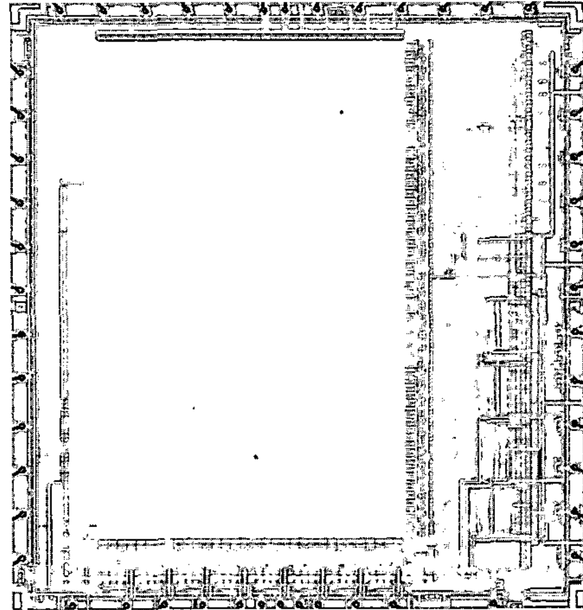

Fig. 6 – Microphotograph of the experimental chip.

3. Experimental results

The proposed experimental chip was integrated by using a single-poly, single-well, 3-V 0.18-μm CMOS technology. Fig. 6 shows a chip microphotograph. The nominal supply voltage V_{dd} is 1.8 V. V_A was set to 3.3 V so as to correctly bias the column decoder during both read and program operations.

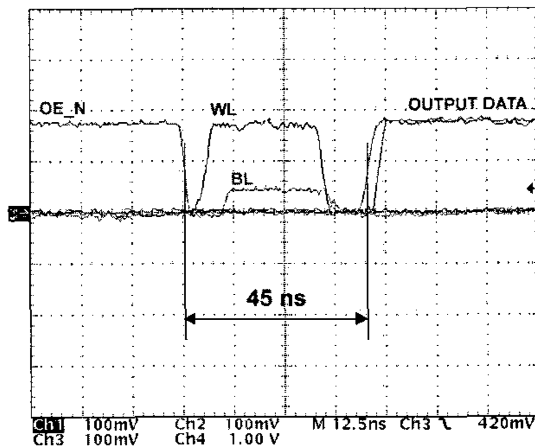

Fig. 7 – Measured voltage waveforms when reading a SET cell (*OE_N*, *WL*, and *BL*: active probes, attenuation by a factor of 10).

Fig. 7 illustrates the measured voltage waveforms when reading a SET cell (*WL* = wordline; *BL* = bitline; *OE_N* = output enable, active low; *OUTPUT DATA* = I/O pin; cell current = 80 μA; reference current = 30 μA). The read access time is 45 ns.

Fig. 8 shows the measured voltage waveforms in consecutive RESET and SET operations (*WE* = write enable, active high; *INPUT DATA* = I/O pin). During the RESET operation, the falling edge of the addressed wordline voltage has to be very sharp so as to allows the melted GST material to be rapidly cooled, thus correctly amorphizing the cell. This operation, referred to as quenching, is carried out by keeping the wordline fall time within few ns (in our case, 2 ns). From Fig. 8, a RESET pulse of 40 ns and a SET pulse of 150 ns were demonstrated. In the proposed experimental chip, a write parallelism of 8 was implemented. The write throughput, which is determined by the 200-ns SET time (SET pulse + 50 ns due to circuitry delay), is therefore 5 MB/s. The achieved write throughput represents a strong improvement with respect to the current NOR Flash memory performance. A write throughput of 10 MB/s (on the same order as for NAND Flash memories) can be obtained by increasing the write parallelism up to 16, at the cost of an increase in the current drawn from the power supply V_{dd} (30 mA, also taking the efficiency of the charge pumps into account).

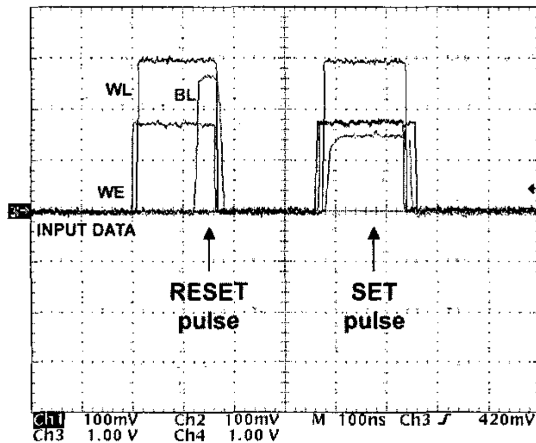

Fig. 8 – Measured voltage waveforms during SET and RESET operations (*WL, BL*: active probes, attenuation by a factor of 10).

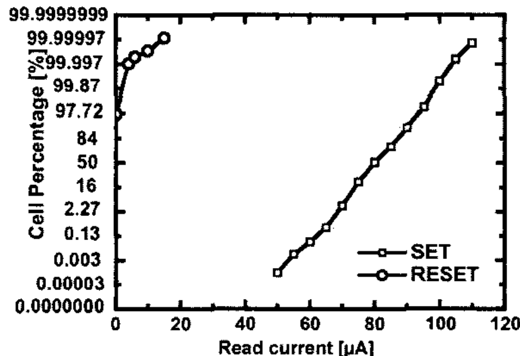

Fig. 9 – Read currents distributions of a 4-Mb array after subsequent full-SET and full-RESET operations.

Several measurements were also successfully performed to assess the whole chip functionality. Fig. 9 illustrates the cell current distributions referred to the whole 4-Mb array after subsequent full-SET and full-RESET operations. The achieved current window is more than adequate for information storage. The experimental results confirm the whole functionality and the feasibility of a stand-alone Phase-Change Memory with standard CMOS fabrication process.

4. Conclusions

A 4-Mb Phase-Change Memory experimental chip using an MOS device as a cell selector, integrated with a 3-V 0.18-µm CMOS technology, has been presented. A read access-time of 45 ns and a write throughput of 5 MB/s were measured, demonstrating improved performance as compared to currently dominant NOR Flash memories. Experimental data of cell current distributions on the 4-Mb array have been provided, proving the chip functionality and the feasibility of a stand-alone Phase-Change Memory with a standard CMOS fabrication process.

References

[1] S. Tyson, G. Wicker, T. Lowrey, S. Hudgens, and K. Hunt, "Nonvolatile, high density, high performance phase-change memory", *Proc. 2000 IEEE Aerospace Conference*, vol. 5, pp. 385-390, March 2000.

[2] M. Gill, T. Lowrey, and J. Park, "Ovonic Unified Memory – A high performance nonvolatile memory technology for stand-alone memory and embedded applications", *2002 IEEE Solid-State Circuits Conference Dig. Tech. Pap.*, vol.1, pp. 458-459, Feb. 2002.

[3] W. Y. Cho, B.-H. Cho, B.-G. Choi, H.-R. Oh, S.-B. Kang, K.-S. Kim, K.-H. Kim, D.-E.Kim, C.-K. Kwak, H.-G. Byun, Y.-N. Hwang, S.-J. Ahn, G.-T. Jung, H.-S. Jung, and K. Kim, "A 0.18 µm 3.0 V 64Mb non-volatile phase-transition random-access memory (PRAM)", *2004 IEEE Solid-State Circuits Conference Dig. Tech. Pap.*, vol. 47, pp. 40-41, Feb. 2004.

[4] F. Pellizzer, A. Pirovano, F. Ottogalli, M. Magistretti, M. Scaravaggi, P. Zuliani, M. Tosi, A. Benvenuti, P. Besana, S. Cadeo, T. Marangon, R. Piva, A. Spandre, R. Zonca, A. Modelli, E. Varesi, T. Lowrey, A. Lacaita, G. Casagrande, P. Cappelletti, and R. Bez, "Novel µtrench phase-change memory cell for embedded and stand-alone non-volatile memory applications", to appear in *Proc. 2004 Symposium on VLSI Technology*.

[5] G. Torelli and P. Lupi, "An improved method for programming a word-erasable EEPROM", *Alta Frequenza*, vol. LII, pp. 487-494, Nov./Dec. 1983.

[6] V. N. Kynett, M. L. Fandrich, J. Anderson, P. Dix, O. Jungroth, J. A. Kreifels, R. A. Lodenquai, B. Vajdic, S. Wells, M. D. Winston, and L. Yang, "A 90-ns one million erase/program cycle 1-Mbit Flash memory", *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1259-1264, Oct. 1989.

[7] G. A. Rincon-Mora and P. E. Allen, "Low-voltage, low quiescent current, low drop-out regulator", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 1, 36-44, Jan. 1998.

[8] M. Zhang, N. Llaser, and F. Devos, "Improved voltage tripler structure with symmetrical stacking charge pump", *Electronics Letters*, vol. 37, pp. 668-669, May 2001.

[9] F. Bedeschi, E. Bonizzoni, O. Khouri, C. Resta, and G. Torelli, "A fully symmetrical sense amplifier for non-volatile memories", to appear in *Proc. 2004 IEEE International Symposium on Circuits and Systems*.