14.8 A 45nm 1Gb 1.8V Phase-Change Memory

Corrado Villa[1], Duane Mills[2], Gerald Barkley[2], Hari Giduturi[2], Stefan Schippers[1], Daniele Vimercati[1]

[1]Numonyx, Agrate Brianza, Italy
[2]Numonyx, Folsom, CA

Floating-gate Flash memories have been able so far to satisfy the market requirements, especially for the portable equipments, and to be the mainstream non-volatile memory (NVM) technology [1]. Projecting into the next decade, though, there are several limitations that must be faced to further scale the floating-gate concept. The increasing complexity of floating gate scaling has left room for the investigation of alternative NVM concepts. Phase-change memory (PCM) technology is the only one of the proposed alternative technologies that is demonstrating the capability to enter in the broad NVM market and to become mainstream in the next decade [2]. PCM, based on the property of chalcogenide materials that have two stable states, amorphous and crystalline, with different electrical resistivity values, provides a new set of features interesting for novel applications, combining features of NVM and DRAM and being at the same time a sustaining and a disruptive technology. PCM can be exploited by the memory system resulting from the convergence of consumer, computer and communication electronics. In this paper, we present a read-while-write 45nm [3] 1Gb non-volatile memory based on phase-change $Ge_2Sb_2Te_5$ (GST) alloy. The device implements 3 specifications that are compatible with existing NOR products for embedded and wireless market with no mux, address/data mux (ADMux) and address/address/data mux (AADMux) interfaces. An extended command set that allows bit-alterability is also included.

The memory is organized into 256 tiles of 4Mb each. Sixteen X-aligned tiles compose a partition of 64Mcells, with 16 total partitions. Two partitions can be simultaneously active, one being read and the other being written. A total of 256 read and 256 verify sense amplifiers are placed at the bottom side of the array to read the cells and implement the read-while-write functionality. In the sense amp region, a bank of program loads is placed to pulse the array's cells. The redundancy resources system is composed of spare columns, spare rows and spare tiles. To further improve reliability an error-correction scheme with less then 10ns correction time is embedded in the device.

The storage element is composed of a heater on the bottom and the GST material on the top. A PNP bipolar transistor selects the cell. The bitline (BL) is connected to the cell GST material. The PCM array is built out of a basic structure made of 4 cells, with 4 emitter contacts and 1 base contact. A TEM of the basic structure is shown in Fig. 14.8.1. A biasing system guarantees that during a read or write operation any BJT parasitic effect has negligible impact. The WL driver is composed of a pull-up resistor and a NMOS pull-down. The BL is biased by the sense amplifier during read or verify operation, by the program load during pulsing and floated or soft controlled by a leaker in all the other cases. In standby state, all the WLs are de-selected and biased by the pull-up resistor to the de-selection voltage (V_{HX}) that is around 1.2V. Each BL results in an intermediate voltage between V_{HX} and ground depending on the current balance of reverse- and forward-biased PNP BJT selectors (soft-driven). A cell is selected during a read operation by selecting the corresponding WL and biasing the corresponding BL to a V_{READ} voltage (around 1.2V). This voltage is driven by the sense amplifier allowing a current in the BL depending on the set/reset status of the cell. A cell is selected during a write operation (set/reset) by driving the corresponding WL to ground, leaving all the other WLs connected to the pull-up resistor at a V_{HX} voltage (around 5V in program) and biasing the corresponding BL to a V_{PROG} voltage (around 5V) through the program load, shown in Fig. 14.8.2. An aggregate tile leakage current due to all the reverse-biased junctions is present both in a read and program status. It is much higher during program than during read. A mechanism to reduce leakage in program by keeping active a subset of tiles is available.

The program load is able to inject a current or force a voltage depending on the device configuration. In the voltage-forcing approach, a cell-position-compensation system is implemented to reduce the parasitic effect. The WL compensation is based on adding a resistive path to ensure that all the cells have the same equivalent parasitic resistance value. The BL compensation sets the program load voltage depending on the row address.

To program a pattern of data, a sequence of SET pulse, RESET pulse, SET verify and RESET verify is given after a pre-read phases. In case the verify fails, the sequence is repeated using larger pulses. Figure 14.8.3 shows typical waveforms of the program algorithm starting from a pre-read phase and with some program-verify steps. The granularity of the power injection in the cell is the result of the trade-off between programming speed and setting precision. The pulses have a trapezoidal shape with parameters that are controlled by the algorithm in height and slope for discharge. RESET pulses can be as short as 60ns while SET pulses can be configured in a range of several hundreds ns (discharge ramp).

Figure 14.8.4 is a schematic of the sense amplifier, which is a differential I/V converter, plus comparator. It is designed to bias the BL directly from V_{CC}. The differential structure rejects any read-while-write and supply noise occurring during the cell-read operation. A BL-discharge-after-read feature is implemented to reduce the error introduced by the WL rise. For a multiple set cells pattern on the same WL, the higher current injected in the selected WL by the set cells increases the voltage, affecting weak set cells and reset cells value. Lowering the corresponding BL after a set cell is detected, reduces this effect. The reference current is generated by a synthesizer that uses a temperature sensor, which is implemented inside the memory and tracks the temperature variation, maximizing the read window.

In order to be robust to the high temperatures of the soldering process, special data, like device configuration, redundancy information, algorithm parameters, etc., are stored in fuses, where the blown/not-blown state is retained even when the high temperatures of the soldering process are applied. To facilitate testing, a PCM array shadows the fuses, allowing temporary values to be written; the final configuration of the device is then frozen by copying the PCM shadow array contents into the fuses.

Figure 14.8.5 summarizes the main characteristics of the device and Figure 14.8.6 shows the die micrograph.

Acknowledgements:
We gratefully acknowledge the work done by the whole Design Team, CAD support, Engineering Team and R&D Technology Development Team.

References:
[1] R. Bez, et al. "Introduction to flash memory", *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489-502, April 2003.
[2] R. Bez and G. Atwood, "Chalcogenide phase change memory: scalable NVM for next decade?", *NVSMW* 06., pp. 12-15, 2006
[3] G. Servalli, "A 45nm Generation Phase Change Memory Technology", *IEDM 2009*, session number 5.7.

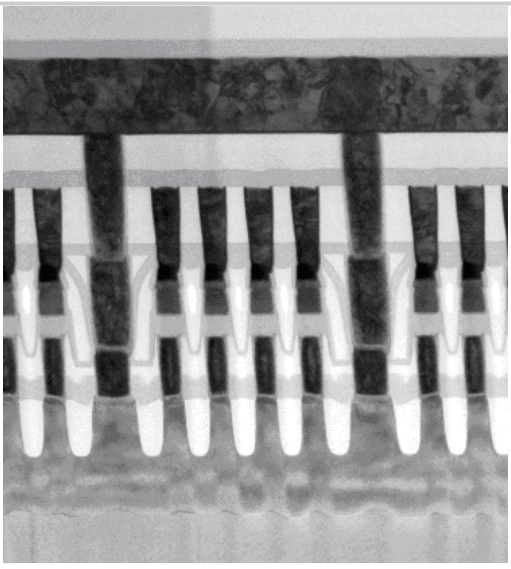

Figure 14.8.1: TEM of four-cell basic structure.

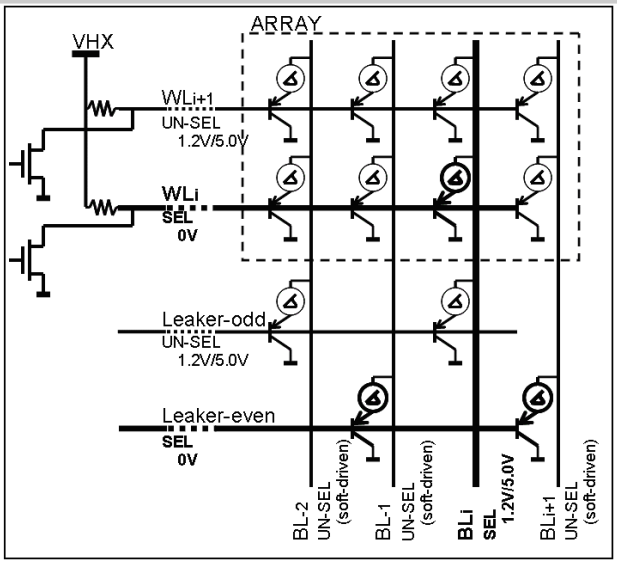

Figure 14.8.2: WL and BL biasing.

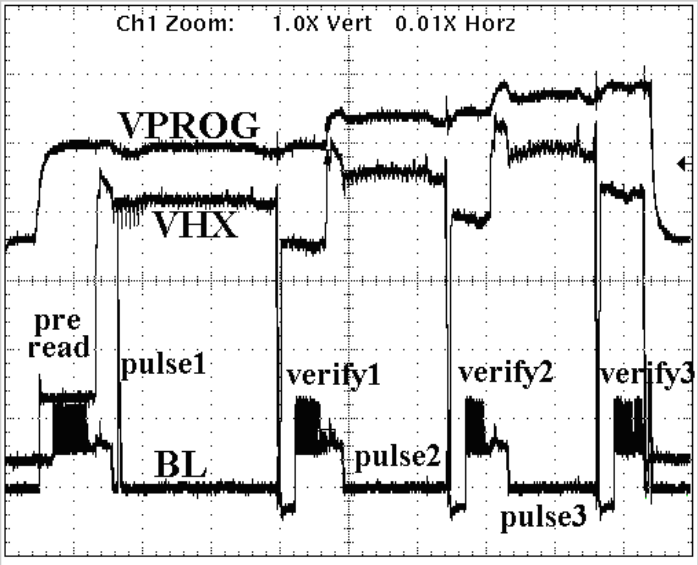

Figure 14.8.3: Program measure.

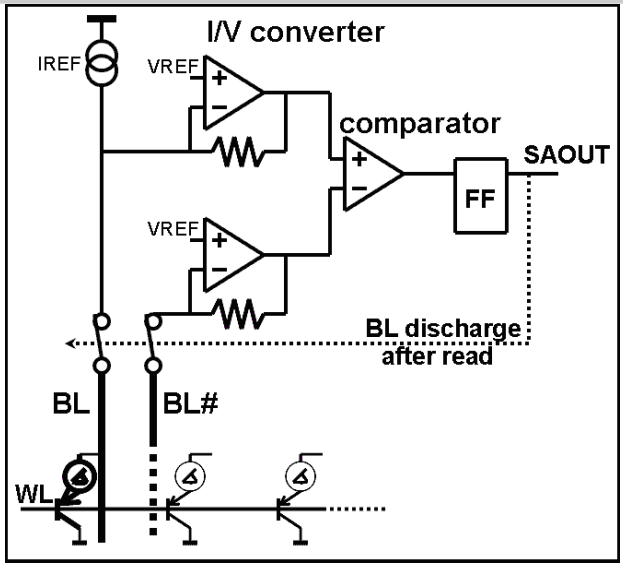

Figure 14.8.4: Sense ampflifier block scheme.

Process Technology	45nm PCM
Storage Material	Ge_2-Sb_2-Te_5
Interconnects	Triple Cu + Alucap
Cell selector	Vertical pnp bjt
Chip area	37.5mm2
Power supply range	1.7V, 2.0V
Stand by consumption	40µA
Temperature range	-40C, +85C
Initial Access Speed	85ns
Max Read Throughput	266MB/s
Max Prog Throughput	9MB/s
Bit Alterable Writes	Yes

Figure 14.8.5: Summary of key features.

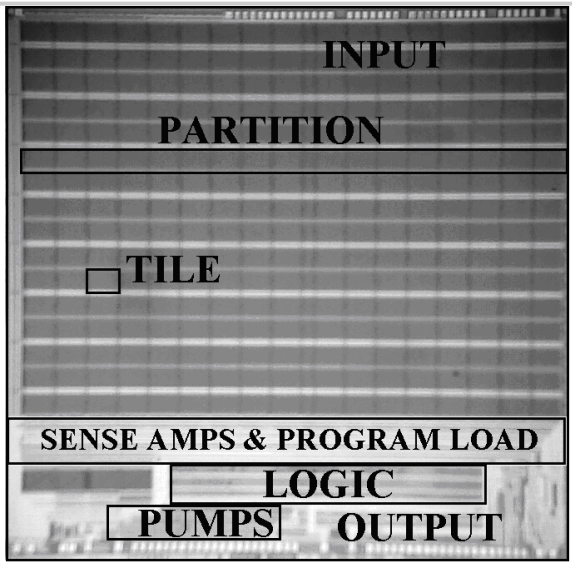

Figure 14.8.6: Die micrograph.

14