12.5 A 128Gb 3b/cell NAND Flash Design Using 20nm Planar-Cell Technology

G. Naso[1], L. Botticchio[1], M. Castelli[1], C. Cerafogli[1], M. Cichocki[1], P. Conenna[1], A. D'Alessandro[1], L. De Santis[1], D. Di Cicco[1], W. Di Francesco[1], M.L. Gallese[1], G. Gallo[2], M. Incarnati[1], C. Lattaro[1], A. Macerola[1], G. Marotta[1], V. Moschiano[1], D. Orlandi[1], F. Paolini[1], S. Perugini[1], L. Pilolli[1], P. Pistilli[2], G. Rizzo[2], F. Rori[1], M. Rossini[1], G. Santin[1], E. Sirizotti[1], A. Smaniotto[2], U. Siciliani[2], M. Tiburzi[1], R. Meyer[3], A. Goda[3], B. Filipiak[3], T. Vali[1], M. Helm[4], R. Ghodsi[4]

[1]Micron, Avezzano, Italy,
[2]Micron, Padua, Italy,
[3]Micron, Boise, ID, [4]Micron, San Jose, CA

We develop a 128Gb 3b/cell NAND Flash memory based on 20nm fully planar cell process technology. The planar cell allows the memory cell to be scaled in both the wordline (WL) and bitline (BL) directions, resulting in a small 3b/cell memory device. The sensing scheme is based on a ramping technique that allows the detection of hard and soft states in a single operation.

The data cache structure is designed to efficiently manage features that optimize the read-window budget (RWB), including pre- and post-compensation to mitigate floating-gate-to-floating-gate (FG-FG) interference. The design also includes read algorithms to minimize the bit-error rate by tracking eventual distribution shifts due to cycling, retention, and charge loss. The I/O interface allows ONFI 2/3 protocols for data-in/-out with a 6ns cycle. Array access in the 3b/cell configuration allows a sustained write throughput of 4MB/s and a 100MB/s read speed.

In recent years, applications such as MP3 players, SSDs, and digital cameras have driven the development of increasingly higher-density NAND Flash memory. The current array minimum approaches values below 20nm, requiring management of effects like FG-FG interference and distribution shifts due to cycling, retention, and quick charge loss. These effects deteriorate the RWB needed to allocate the eight distributions associated to the 3b/cell device and requires action from the process and design point of view to manage non-overlapping margins between distributions. From the process point of view, the technology developed so far is adequate down to the 25nm technology node [1]. Our 3b/cell NAND Flash memory is developed and manufactured using this technology [2].

These scaling challenges, particularly after scaling the memory cell in both directions to around 20nm, necessitate the introduction of a number of approaches in design and process technology. This paper describes the key improvements required to overcome these challenges. A new process, based on planar gate cell (eliminating control-gate-floating-gate wrap-around) uses high-Δ and a metal gate [1]. The NAND Flash presented in this article is a 3b/cell NAND Flash that uses this planar 20nm process.

The cross-section of the planar 20nm cell is shown in Fig. 12.5.1. Key features of the process are: thin-poly floating gates to reduce cell-to-cell interference; a metal control gate; air-gap isolation for both cell gates and metal BLs to reduce coupling capacitance; and high-k inter-gate dielectric metal gates. In Fig. 12.5.2, a plot of total FG-FG is reported for various WRAP and PLANAR technology. The device presented in this article is a 128Gb, 3b/cell NAND Flash memory in the 20nm technology node (Fig. 12.5.3). It is organized into two planes, each having 1368 blocks. Each block is composed of a NAND string having 128 physical WLs corresponding to 768 pages (lower, middle, and upper page). Each page is 8KB long and is organized into even/odd couples with a single sense amplifier for each pair of BLs.

The read mechanism is based on a ramped WL sensing circuit and multi-latch page buffers [3]. The ramped WL sensing technique consists of a counter having its output WLDAC connected to a DAC circuit generating a ramped WL waveform during read operation. At every step of the WL ramp, a sensing operation is performed. Only when the WL voltage is greater than cell V_t, the selected cell is turned on and PBDAC digital value is loaded into corresponding pagebuffer. The I/O interface of the device has been designed to be consistent with ONFI 2/ONFI 3 at clock cycle of 6ns DDR.

The voltage for the selected WL during the read and verify operations is generated by a dedicated analog sub-system (see Fig. 12.5.4), designed to shape a digitally controlled voltage ramp. This hardware allows control of the slope of the voltage ramp and the ability to perform temperature compensation [4]. A dedicated resistor ladder is used to generate the voltage levels of the voltage ramp between the initial and final values. The controller performs the ramp generation through this sub-system, driving the ramp generation register (REGISTER_0) with an appropriate counter.

Pages are programmed in this order: first the even, then the odd on the same WL, and then higher-order WLs. During the program operation, the even pages can suffer interference due to the subsequent program operation of the odd in the same WL or to the subsequent program operation of the higher-order WLs. The final distributions achieved are shown in Fig. 12.5.5.

New features have been developed to compensate for the FG-FG in both write and read operations. The pre-compensation technique is shown in Fig. 12.5.5. This technique compensates for the FG-FG interference based on an internal algorithm, evaluating the aggression level from adjacent cells and appropriately positioning the target cell.

The corrective read technique is activated by the external controller and is performed internally in the memory. The purpose of this feature is to perform data correction due to FG-FG interference during the read operation after both the victims and the aggressors have been programmed. The corrective read technique detects if the surrounding cells have been programmed to be aggressors and adjusts the read level of the target cell accordingly. The channel calibration technique uses an algorithm executed on the NAND die to detect the read level that minimizes the bit error rate.

A 3b/cell NAND memory device that uses a planar memory cell scaled in both the BL and WL directions is developed. The memory device uses a ramp WL sense and read technique that allows simultaneous read of hard stored bits and soft bits, indicating the confidence level associated with the stored level. Fig.12.5.7 shows a micrograph of the 3b/cell device.

Acknowledgment:
The authors thank all other team members who contributed to this project, from the design, layout, product engineering, and technology teams.

References:
[1] K. Prall, et. al. IEDM 2010, 5.2.1
[2] M. Goldman et. al. IMW 2011
[3] Sarin et. al. USPTO patent 7,948,802 (2011)
[4] G.G Marotta et al. "A 3b/cell 32Gb NAND flash memory at 34nm with 6MB/s program throughput and with a dynamic 2b/cell blocks configuration mode for a program throughput increase up to 13Mb/s," *ISSCC Dig. Tech Papers*, pp. 444-445, Feb. 2010.

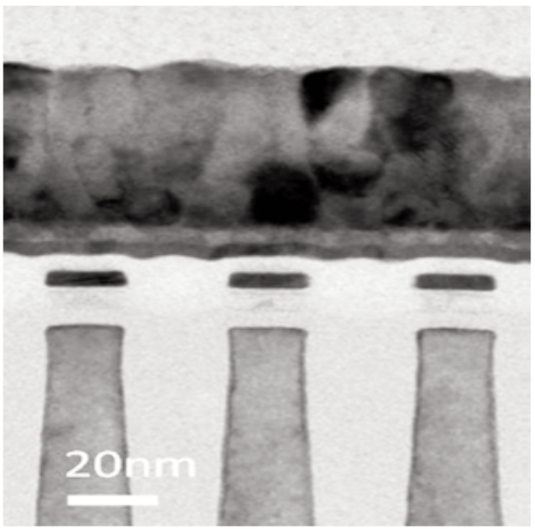

Figure 12.5.1: Planar NAND cell cross-section.

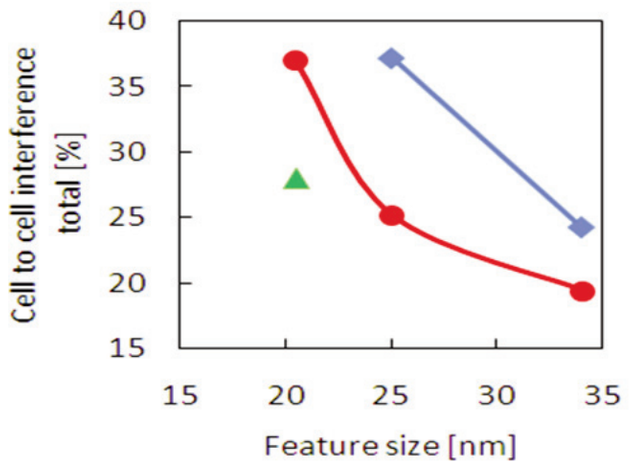

Figure 12.5.2: FG-to-FG coupling at different technology nodes.

Technology	20 nm
Chip size	146.5 mm2
Chip capability	128 Gb at 3bits/cell
Organization	8192 bytes per page x 2 (even/odd) x 384 pages x 1368 blocks x 2 planes x 8 IO
tCLK DDR	6ns
Read time	90 us
Average Program time	2.8ms

Figure 12.5.3: Basic device features.

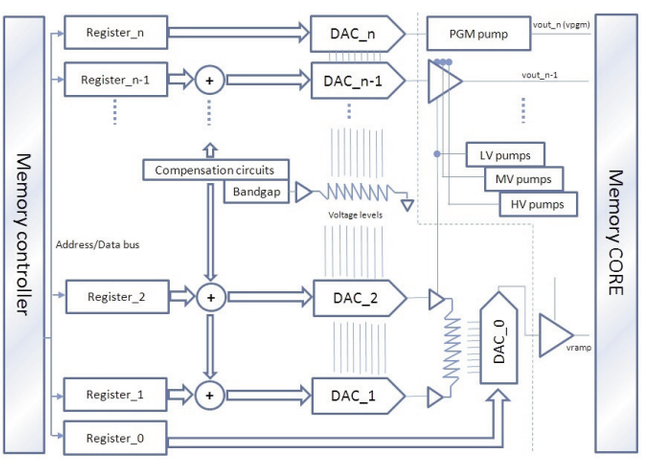

Figure 12.5.4: Ramp generation.

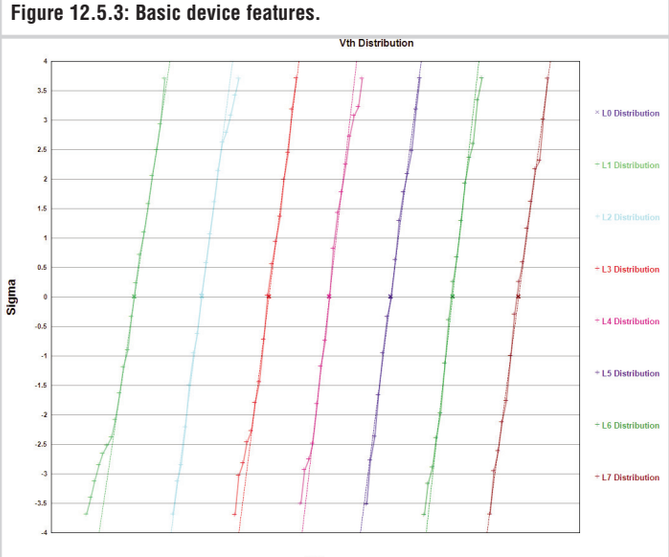

Figure 12.5.5: Silicon data of 3b/cell distributions.

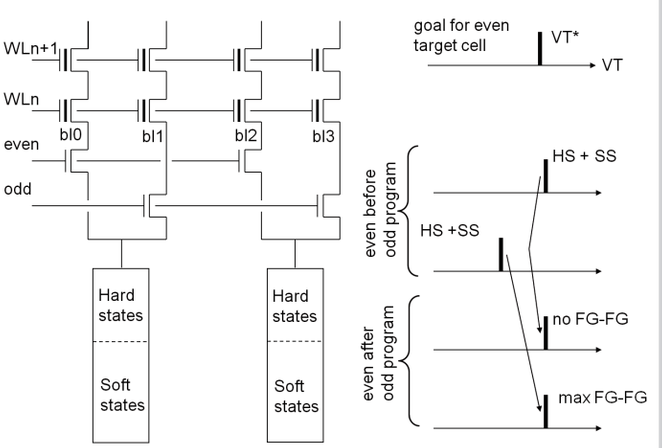

Figure 12.5.6: Pre-compensation.

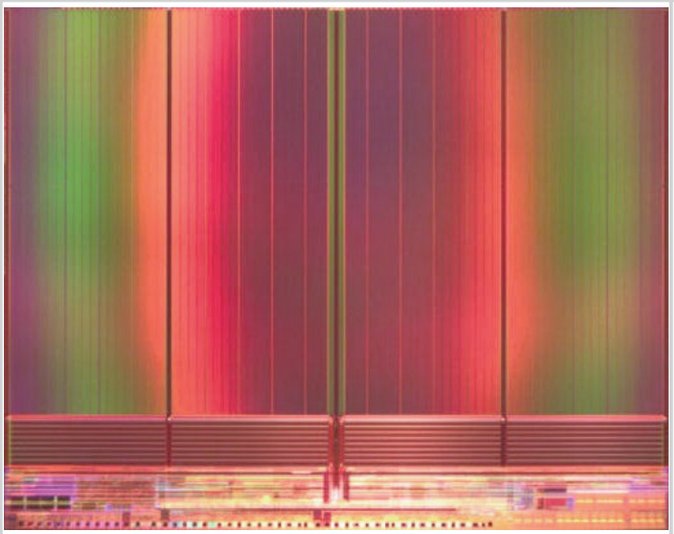

Figure 12.5.7: Device micrograph.