24.7 A 3bit/Cell 32Gb NAND Flash Memory at 34nm with 6MB/s Program Throughput and with Dynamic 2b/Cell Blocks Configuration Mode for a Program Throughput Increase up to 13MB/s

G.G. Marotta[1], A. Macerola[1], A. D'Alessandro[1], A. Torsi[1], C. Cerafogli[1], C. Lattaro[1], C. Musilli[1], D. Rivers[2], E. Sirizotti[1], F. Paolini[1], G. Imondi[1], G. Naso[1], G. Santin[1], L. Botticchio[1], L. De Santis[1], L. Pilolli[1], M.L. Gallese[1], M. Incarnati[1], M. Tiburzi[1], P. Conenna[1], S. Perugini[1], V. Moschiano[1], W. Di Francesco[1], M. Goldman[3], C. Haid[3], D. Di Cicco[1], D. Orlandi[1], F. Rori[1], M. Rossini[1], T. Vali[1], R. Ghodsi[4], F. Roohparvar[4]

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

In recent years applications such as mp3 players, SSD, digital cameras and video camcorders have driven the development of increasingly higher density NAND memories. In the presented 3b/cell memory the read and programming throughputs are been enhanced with the adoption of a quad-plane architecture and an industry standard even-odd bitline (BL) decoding scheme. The architecture, while featuring same page size of 16KB as recently disclosed ABL architectures [3,4], avoids the shortcomings such an ABL scheme exhibits in programming mode due to floating-gate-to–floating-gate coupling. The chip features both the newly developed synchronous DDR interface and the standard, asynchronous NAND flash interface. A 66-cell string is adopted to optimize the die size at 126mm^2.

The chip is divided into four 8Gb arrays. Each array consists of 684 blocks and each block consists of 384 pages. The page size is 4KB that can be extended to 16KB in quad-plane operation. The chips main features are summarized in Fig. 24.7.1.

A technique called source-bias read/verify, consisting in raising the source voltage during read, is been implemented to widen the V_T allocation window. A new sensing technique is adopted to ensure that the BL-to-source-voltage difference remains constant as the source voltage is raised during read. The LMU technique, consisting in programming lower (L), middle (M) and upper (U) pages in consecutive order, as shown in Fig. 24.7.2, is adopted to improve the precision of distribution placement by mitigating the floating-gate interference (FG-FG) [5]. The first pulse ramp technique [6-7], consisting in shaping the first program and erase pulse with an initial voltage ramp, is implemented to better preserve oxide integrity and, therefore, data retention. The final distributions achieved with all the above-mentioned techniques are shown in Fig 24.7.3.

The chip supports asynchronous data I/O at 20ns cycle time and synchronous DDR access at an I/O rate of 6ns/byte. This read performance is achieved by using the data path architecture shown in Fig. 24.7.4. The architecture is based on the following key design points: four bytes simultaneous sensing in a plane, three stages pipeline and byte mux at I/O pads level. The first stage pipe is to account for RE# and column-address counter propagation; the second stage pipe is to account for address decoder, cache selection and second sensing (SSA) propagation; the third stage pipe is to account for plane mux and data propagation. Multi-plane operations are supported up to a maximum of four planes.

The analog system features 10-bit resolution. A digital thermometer is developed for on-chip temperature compensation of array operation. Most of the 3b/c related features are implemented by enhancing the controllability of the analog block by the chip controller via dedicated control registers and by performing all accurate analog operations in precision low-voltage low-noise analog-signal-level (ASL) circuitry built around a 10-bit resistor ladder, as shown in Fig. 24.7.5. In this approach, the charge-pump section, which is the largest-area analog block, does not require major changes with respect to the 2b/c HV, MV and LV pump implementation. The ASL circuitry is composed of a voltage-reference generator, some fine adjustment and buffering circuits, a digital thermometer, some compensation circuitry and a bank of digital to analog converters (DACs). The chip controller shapes the voltage waveforms applied to the memory core by writing digital patterns to the DAC input registers as determined by read, write and erase algorithms. Each analog voltage, except the selected wordline program voltage, is provided by a linear regulator that amplifies the DAC low-voltage output by a fixed gain factor. The linear regulator supply voltages are generated by charge pumps for regulated voltage values higher than the chip V_{CC}. The programming voltage is provided directly by the output of a charge pump. All DACs share the same resistor divider string, while each DAC consists of an analog multiplexer selecting the resistor-divider tap addressed by the digital code written to the input register.

The analog system performs temperature sensing and conversion to a 7-bit digital code for the purpose of compensating temperature-variation effects. Several output voltages can be compensated independently, as schematically shown in Fig 24.7.6. The compensation scheme works by changing the input code of the DACs according to chip temperature. Compensation data, stored in an auxiliary table, are added by dedicated hardware to DAC inputs to generate the compensated code. The chip temperature is sensed through a comparison of the bandgap reference output with PTAT voltage, followed by analog-to-digital conversion. The digital thermometer output can be read by the controller and used by the internal algorithms.

The control scheme of the NAND device is based on a 32-bit-instruction 16-bit-data RISC processor running a custom instruction set dedicated to NAND Flash requirements. The microcode size for 3b/c algorithms is about 8K instructions and is stored in a metal2-programmable ROM. The processor is composed of a control unit that decodes the instruction and manages all other blocks, a 64-word register file for temporary-variable storage and an arithmetic-logic unit (ALU) performing addition, subtraction, comparisons and basic logic operations (OR, AND, NOT, XOR). The die micrograph is shown in Fig. 24.7.7.

[1] Y. Li, S.Lee et alt; "A 16Gb 3b/Cell NAND Flash Memory in 56nm with 8MB/s Write Rate." *ISSCC Dig. Tech Papers*, pp.506-507, Feb., 2008.
[2] Nobunaga et al., "A 50nm 8Gb NAND Flash Memory with 100MB/s Program Throughput and 200MB/s DDR Interface," *ISSCC Dig. Tech Papers,* pp. 426- 625, Feb., 200.8
[3] Y. Li et al., "A 16Gb 3b/Cell NAND Flash Memory in 56nm with 8MB/s Write Rate," *ISSCC Dig. Tech Papers,* Feb., 2008.
[4] T. Futatsuyama et al., "A 113 mm2 32Gb 3b/cell NAND Flash Memory," *ISSCC Dig. Tech Papers,* Feb., 2009.
[5] M. Park, K. Kim, J. Park, J. Choi, "Direct Field Effect of Neighboring Cell Transistor on Cell-to Cell Interference of NAND Flash Cell Array," *IEEE EDL*, Pg. 174, Feb., 2009.
[6] M. Smayling, "The Impact of Vpp Waveform Upon EEPROM Write/Erase endurance," Texas Instruments Technical Journal, Pg. 77, Jan.-Feb., 1991.
[7] Michael Smayling, Dual Slope, Feedback controlled, EEPROM programming, P.N. 4628487, Dec. 9, 1986.

 978-1-4244-6034-2/10/$26.00

Technology	34nm triple-well 3 metals
Cell size	0.0054 um2 (select gates included) 64 WLs per string
Chip size	126 mm2
Organization	4096 Bytes per page x 384 pages x 684 blocks x 4 planes x 8 IO
ECC per 4KB page	224 Bytes
Array read time	60 us typ – 100 us max
Program time	2 ms typ – 10 ms max
Erase time	10 ms typ – 30 ms max
Clock cycle time	12 ns
Supported multiple bit/cell	1 b/c – 2 b/c – 3 b/c

Figure 24.7.1: Chip basic features.

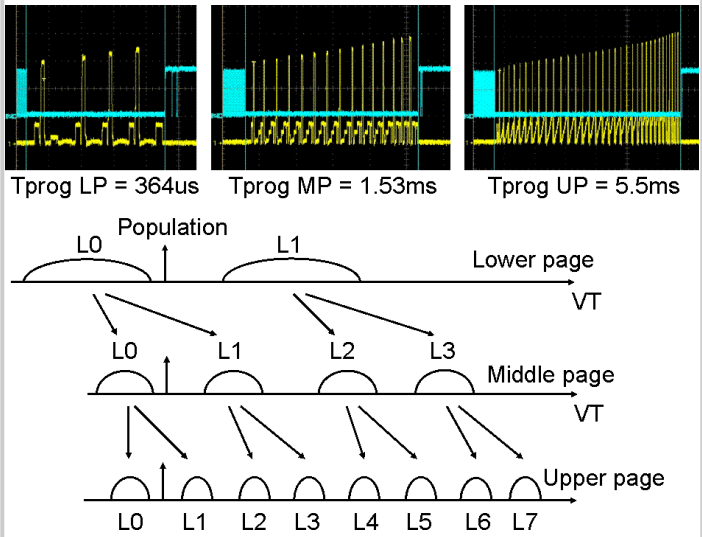

Figure 24.7.2: LMU technique.

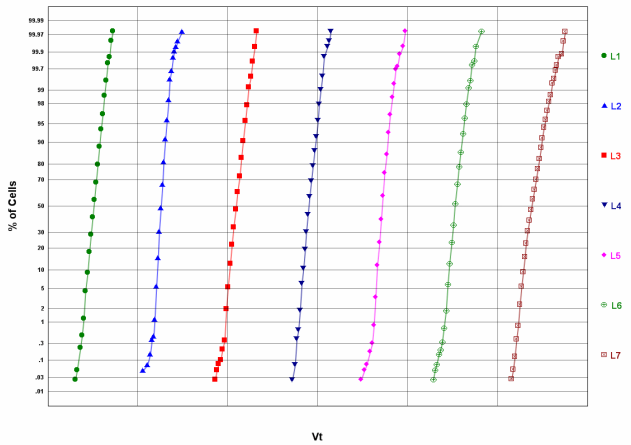

Figure 24.7.3: Silicon data of 3b/cell distributions.

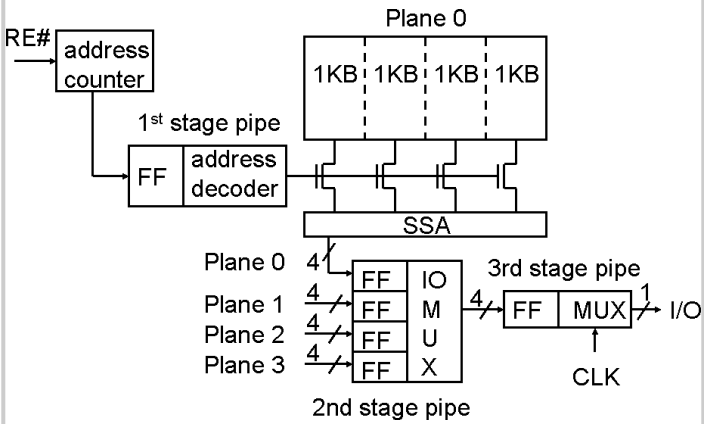

Figure 24.7.4: Read data-path architecture.

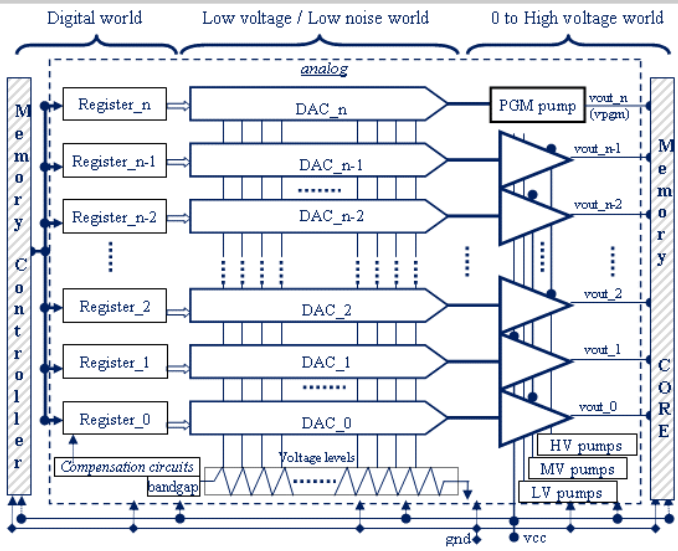

Figure 24.7.5: Analog system.

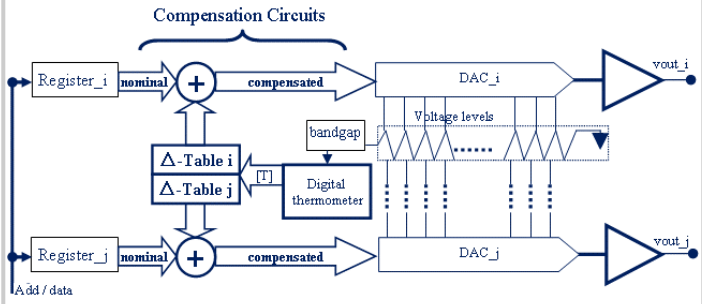

Figure 24.7.6: Digital temperature compensation.

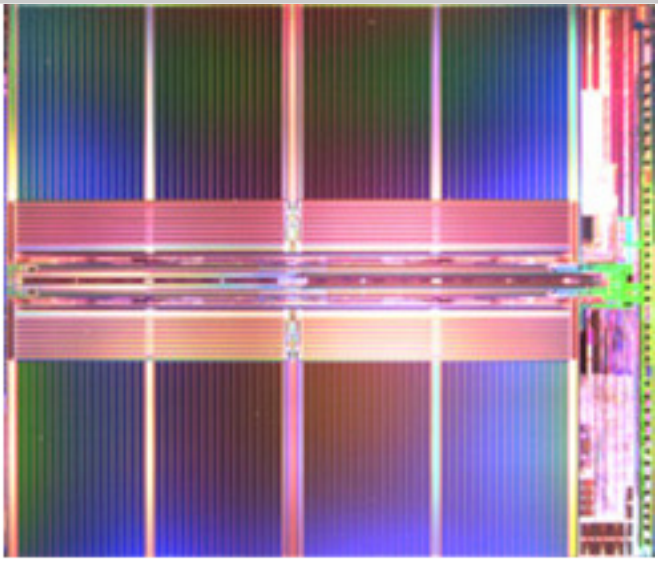

Figure 24.7.7: Chip micrograph.

978-1-4244-6034-2/10/$26.00