First Detection of Single-Electron Charging of the Floating Gate of NAND Flash Memory Cells

Christian Monzio Compagnoni, *Senior Member, IEEE*, Giovanni M. Paolucci, Carmine Miccoli, Alessandro S. Spinelli, *Senior Member, IEEE*, Andrea L. Lacaita, *Fellow, IEEE*, Angelo Visconti, *Member, IEEE*, and Akira Goda

I. Introduction

THE miniaturization of cell dimensions following the uninterrupted scaling of the Flash technology since its introduction in the 80's [1], [2] has largely increased the impact that single electrons in the gate stack have on device threshold voltage (V_T) [3]. This, first, opened the possibility to clearly detect the capture and release of single electrons in the cell tunnel-oxide giving rise to random telegraph noise (RTN) [4]–[7] and post-cycling charge detrapping [8]–[10]. The closest proximity of electrons in the tunnel oxide to the channel surface along with their localized nature in the presence of a percolative source-to-drain current conduction [11], [12] allowed, in fact, to easily identify cells in the array where a very high V_T shift (ΔV_T) occurs after a single capture/release event. This was not possible, instead, for electrons stored in the cell floating gate, as the farther and well-defined distance of the electronic charge from the channel surface and, more important, its uniform distribution over the channel area in this case preclude any significant statistical boosting of the ΔV_T arising from single electrons.

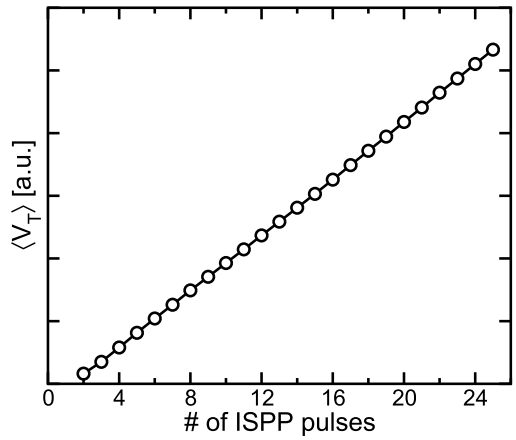

Fig. 1. $\langle V_T \rangle$ of the cells in a page of our NAND array as a function of the number of programming pulses applied during ISPP.

As a consequence, despite a few attempts on unoptimized and exotic cell structures [13], the discrete nature of electron storage in the floating gate of mainstream Flash cells could be detected, so far, only indirectly by monitoring the statistical spread of the ΔV_T resulting from a programming pulse [14]–[16].

In this letter, we prove that experimentally detecting single-electron charging of the floating gate of mainstream Flash cells can be easily accomplished on state-of-the-art technologies. This reveals that the discrete process of electron storage taking place during cell programming can be directly monitored today, allowing a more in-depth investigation of the programming accuracy of future technology generations.

II. Experimental Results

A. Single Read

We investigated the ΔV_T arising from the storage of single electrons in the floating gate on our 16 nm planar NAND Flash technology [17]. To this aim and following our previous works [14]–[16], we adopted the incremental step pulse programming (ISPP) algorithm [18] to make the cells reach a stationary programming regime where their average ΔV_T per pulse is nearly equal to the ISPP step amplitude V_s [16]. This leads to the almost linear increase of the average V_T ($\langle V_T \rangle$) of the cells in a page of the memory array with the number of applied pulses shown in Fig. 1, representing a mandatory condition for accurate program operations [15].

Manuscript received November 13, 2014; revised November 24, 2014; accepted December 1, 2014. Date of publication December 9, 2014; date of current version January 23, 2015. The review of this letter was arranged by Editor T. Wang.

C. Monzio Compagnoni, G. M. Paolucci, and A. S. Spinelli are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy (e-mail: monzio@elet.polimi.it).

C. Miccoli and A. Goda are with Process Research and Development, Micron Technology Inc., Boise, ID 83716 USA.

A. L. Lacaita is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan 20133, Italy, and also with the Consiglio Nazionale delle Ricerche, Istituto di Fotonica e Nanotecnologie, Milan 20133, Italy.

A. Visconti is with Process Research and Development, Micron Technology Inc., Agrate Brianza 20864, Italy.

Color versions of one or more of the figures in this letter are available online.

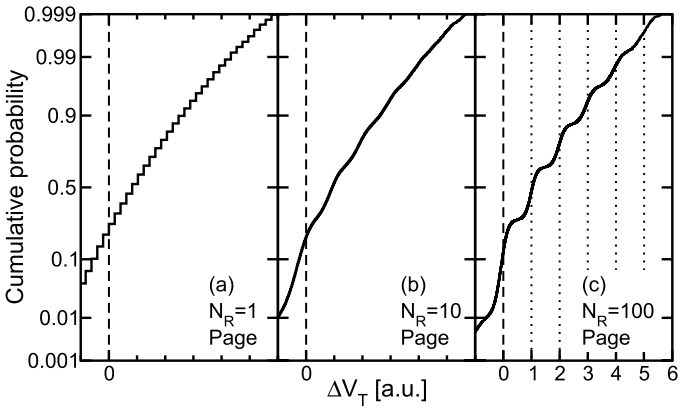

Fig. 2. Statistical distribution of the ΔV_T per pulse among the cells in a page of the NAND array, for increasing number of read operations N_R averaged after each ISPP pulse. On the x axis of (c), the scale shows the number of electrons injected into the floating gate during the programming pulse. The separation of the vertical dashed lines in (c) is in quite good agreement with the calculated ΔV_T corresponding to single-electron storage in the floating gate.

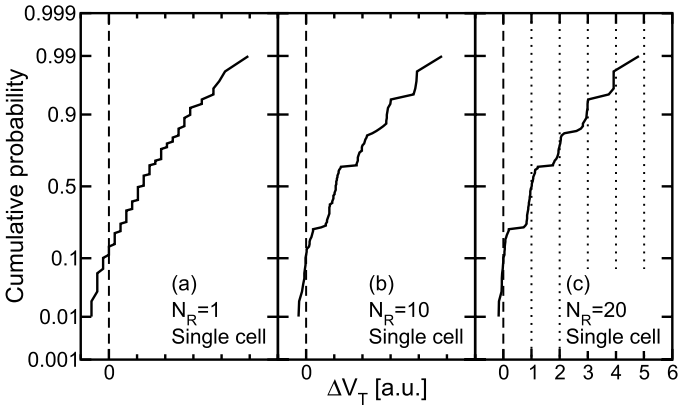

Fig. 3. Same as in Fig. 2, but with statistical results gathered on a single memory cell by repeating many times (100 times) the ISPP operation. On the x axis of (c), the scale shows the number of electrons injected into the floating gate during the programming pulse.

Although the ISPP algorithm minimizes the impact of the unavoidable physical differences among the cells on their V_T shift, Fig. 2a shows that a relevant statistical variability of the ΔV_T per pulse still exists. This variability is the result, first of all, of the statistical nature of the process leading to electron injection into the floating gate during each ISPP pulse (*program noise*) [14]–[16], [19] and, in addition, of fluctuations of V_T due to *read noise*, mainly in the form of RTN and gaussian noise from the sensing circuitry (note that the read noise contribution to ΔV_T is the main reason for the negative tail appearing in Fig. 2a). A similar spread of the ΔV_T per pulse is clearly evident also when gathering statistical results on a single memory cell by repeating many times the ISPP operation, as shown in Fig. 3a. This further confirms that the major component to the ΔV_T spread is not related to cell-to-cell variability but to the intrinsic statistical dispersion of the electron injection process into the cell floating gate [14]–[16] and to read noise. As a final remark, note that both the ΔV_T distribution in Fig. 2a and that in Fig. 3a display a staircase behavior arising from V_T discretization during read. The read operation involves, in fact, increasing the word-line bias by a constant step and defining cell V_T as the first value making cell current greater than a reference level in the tens-of-nA range.

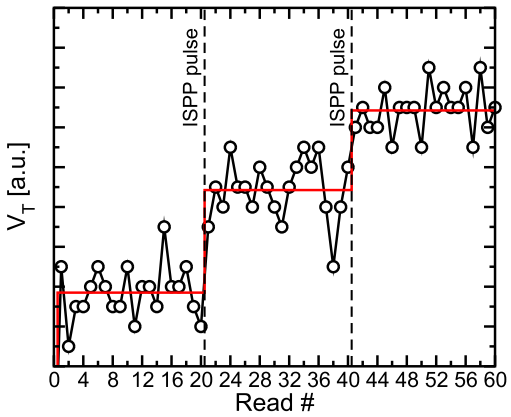

Fig. 4. V_T of a single memory cell resulting from many subsequent read operations after each ISPP pulse. The application of the ISPP pulses is highlighted by the vertical dashed lines. Red line is the result of read averaging.

B. Averaging Many Reads

In order to reduce the impact of read noise on ΔV_T and better highlight the contribution given by the increase of the number of electrons stored in the floating gate during programming, we evaluated V_T as the average of N_R read operations after each ISPP pulse, as shown in Fig. 4 on a single memory cell. In this figure, the relevant fluctuation of the single read operations clearly appears (symbols), which, for the selected V_s, is almost comparable to the V_T increase arising from the application of the ISPP pulses. The effect of averaging $N_R = 10$ and 100 reads on the statistical distribution of the ΔV_T per pulse is shown, respectively, in Figs. 2b and 2c, in the case the statistics is gathered by a single ISPP operation on all the cells in a page. Results highlight that the distribution assumes a staircase behavior for increasing N_R, with the staircase steps equally spaced on the horizontal axis. This is even more evident in Figs. 3b and 3c, where the results achieved with many ISPP operations on a single cell are shown in the case of $N_R = 10$ and 20. In particular, Fig. 3c allows to easily and accurately extract the width of the steps of the ΔV_T staircase (see the vertical dashed lines, reported also in Fig. 2c), which is constant and much larger than that determined by V_T discretization in Figs. 2a and 3a (this represents a mandatory condition setting the precision required of the V_T measurement).

III. Discussion

The staircase behavior of the ΔV_T distribution in Figs. 2c and 3c represents the first direct experimental detection of single-electron charging of the floating gate of mainstream Flash cells. Note, in fact, that the width of the steps in the figures is in very good agreement with the calculated V_T shift corresponding to the storage of a single electron in the cell floating gate, equal to q/C_{pp}, where q is the electronic charge and C_{pp} is the floating

gate to control-gate capacitance (extracted from program noise analyses [16] and numerical simulations [20]). The discretization of ΔV_T giving rise to the staircase behavior of the distributions of Figs. 2c and 3c is, therefore, set by the discrete number of electrons that can be injected in the cell floating gate during ISPP pulses, with the first step above 0 corresponding to 1 electron added to the floating gate, the next one to 2 electrons added to the floating gate and so on. As a remark, note that the smoothing of the distribution that appears at high ΔV_T in Fig. 2c is the result of gathering statistics on many microscopically different cells, with the possibility of a higher RTN in some of them and of small variability in the value of their C_{pp} coming into play.

Finally, it is important to point out that the average number of electrons injected into the cell floating gate during the ISPP pulses is, from Figs. 2c and 3c, nearly equal to 1. This comes from the rather low value of V_s that we used in our experiments, which is lower than those typically adopted even for accurate programming of multi-level devices. However, the result reveals that the number of electrons added to the floating gate of modern Flash cells during each ISPP pulse is quite low and this represents a major spread source for ΔV_T as appearing from Figs. 2c and 3c. To this regard, note that, owing to the very high electric fields in the tunnel oxide during programming, Coulomb-blockade effects [13] are not effective in controlling the number of stored electrons in the floating gate.

IV. Conclusions

In this letter, we have presented the first experimental evidence of single-electron charging of the floating gate of mainstream Flash cells. Results clearly reveal the discretization of the number of electrons that are stored in the cell floating gate during a program pulse, opening the possibility for further studies on the electron injection process.

Acknowledgments

The authors would like to thank S. Beltrami and M. Bertuccio from Micron Technology Inc. for discussions and support.

References

[1] F. Masuoka *et al.*, "A new Flash E^2PROM cell using triple polysilicon technology," in *IEDM Tech. Dig.*, 1984, pp. 464–467.

[2] F. Masuoka *et al.*, "New ultra high density EPROM and Flash EEPROM with NAND structure cell," in *IEDM Tech. Dig.*, 1987, pp. 552–555.

[3] K. Prall and K. Parat, "25 nm 64 Gb MLC NAND technology and scaling challenges," in *IEEE IEDM Tech. Dig.*, Dec. 2010, pp. 5.2.1–5.2.4.

[4] H. Kurata *et al.*, "The impact of random telegraph signals on the scaling of multilevel Flash memories," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2006, pp. 112–113.

[5] N. Tega *et al.*, "Anomalously large threshold voltage fluctuation by complex random telegraph signal in floating gate Flash memory," in *IEDM Tech. Dig.*, Dec. 2006, pp. 1–4.

[6] R. Gusmeroli *et al.*, "Defects spectroscopy in SiO_2 by statistical random telegraph noise analysis," in *IEEE IEDM Tech. Dig.*, Dec. 2006, pp. 1–4.

[7] C. Monzio Compagnoni *et al.*, "Statistical model for random telegraph noise in Flash memories," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 388–395, Jan. 2008.

[8] C. Miccoli *et al.*, "Resolving discrete emission events: A new perspective for detrapping investigation in NAND Flash memories," in *Proc. IEEE IRPS*, Apr. 2013, pp. 3B.1.1–3B.1.6.

[9] G. M. Paolucci *et al.*, "Revisiting charge trapping/detrapping in Flash memories from a discrete and statistical standpoint—Part I: V_T instabilities," *IEEE Trans. Electron Devices*, vol. 61, no. 8, pp. 2802–2810, Aug. 2014.

[10] G. M. Paolucci *et al.*, "Revisiting charge trapping/detrapping in Flash memories from a discrete and statistical standpoint—Part II: On-field operation and distributed-cycling effects," *IEEE Trans. Electron Devices*, vol. 61, no. 8, pp. 2811–2819, Aug. 2014.

[11] A. Ghetti *et al.*, "Physical modeling of single-trap RTS statistical distribution in Flash memories," in *Proc. IEEE IRPS*, Apr./May 2008, pp. 610–615.

[12] A. Ghetti *et al.*, "Comprehensive analysis of random telegraph noise instability and its scaling in deca–nanometer Flash memories," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1746–1752, Aug. 2009.

[13] G. Molas *et al.*, "Manipulation of periodic Coulomb blockade oscillations in ultra-scaled memories by single electron charging of silicon nanocrystal floating gates," *IEEE Trans. Nanotechnol.*, vol. 4, no. 3, pp. 374–379, May 2005.

[14] C. Monzio Compagnoni *et al.*, "First evidence for injection statistics accuracy limitations in NAND Flash constant-current Fowler–Nordheim programming," in *IEEE IEDM Tech. Dig.*, Dec. 2007, pp. 165–168.

[15] C. Monzio Compagnoni *et al.*, "Ultimate accuracy for the NAND Flash program algorithm due to the electron injection statistics," *IEEE Trans. Electron Devices*, vol. 55, no. 10, pp. 2695–2702, Oct. 2008.

[16] C. Monzio Compagnoni *et al.*, "Analytical model for the electron-injection statistics during programming of nanoscale NAND Flash memories," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 3192–3199, Nov. 2008.

[17] M. Helm *et al.*, "A 128 Gb MLC NAND-Flash device using 16 nm planar cell," in *Proc. IEEE ISSCC*, Feb. 2014, pp. 326–327.

[18] G. J. Hemink *et al.*, "Fast and accurate programming method for multi-level NAND EEPROMs," in *Symp. VLSI Technol. Dig. Tech. Papers*, Jun. 1995, pp. 129–130.

[19] C. Monzio Compagnoni *et al.*, "Impact of control-gate and floating-gate design on the electron-injection spread of decananometer NAND Flash memories," *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1196–1198, Nov. 2010.

[20] A. Ghetti, L. Bortesi, and L. Vendrame, "3D simulation study of gate coupling and gate cross-interference in advanced floating gate non-volatile memories," *Solid-State Electron.*, vol. 49, no. 11, pp. 1805–1812, 2005.