Reduction of Bitline to Control Gate leakage for improved Embedded 0.18 μm FLASH Yield and Reliability

A.Cacciato, S. Nelson, M. Diekema, M. Hendriks, L. van Marwijk, C. Deuper, E. Gerritsen, R. Verhaar, and D. Dormans
Philips Semiconductors, MOS4YOU, Gerstweg 2, 6534 AE, Nijmegen, The Netherlands
Antonio.Cacciato@philips.com

Abstract

In this paper we illustrate the modifications to the standard logic CMOS process flow that are necessary to control the bitline to control gate leakage in embedded 0.18 μm FLASH memory arrays. In particular, it will be shown that phase-shift masks for bitline patterning and dual-frequency boarderless nitride deposition are required to guarantee a reliable and manufacturable bitline to control gate isolation.

1. Introduction

Today's trend towards single-chip solutions for digital and wireless electronics forces device manufactures to integrate on the same chip embedded non-volatile memories with high performance logic and analogue functions. This is all but a trivial task. For example, as showed in a previous paper, because of the extra topography introduced by the memory cells, the pre-metal isolation layer (or Inter Level Dielectric) has to be modified for embedded 0.18 μm FLASH technology to avoid bitline-bitline shorts caused by voids in the ILD layer [1].

In addition to the extra topography, another major concern for the integration of non-volatile and logic is the isolation between poly-Si gates and contact plugs, which, unlike in the logic, must withstand in the array the high voltage difference between the bitline contact (BL) and the control gate poly-Si (CG) during the write operation (≈ 15 V in the case of Philip's embedded 0.18 μm FLASH technology [2]).

In this paper we tackle the challenge of achieving a reliable BL-CG isolation for embedded 0.18 μm FLASH technology and we illustrate some modifications in the logic Local Interconnect Layer (LIL) and Inter Level Dielectric (ILD) modules to improve the yield and reliability of the FLASH array.

2. Fabrication of the FLASH cells and leakage measurements

Figure 1 shows a schematic cross section of the BL contact and of the stacked flash cell used in this work. The cell consists of control (CG) and floating (FG) gates separated by an Oxide/Nitride/Oxide layer (effective thickness ≈ 15 nm). The nominal critical dimension (CD) of both the cell length and the BL contact was 240 nm. After cell patterning, spacer formation and salicidation, the ILD stack, consisting of a 80 nm boarderless Si_3N_4 film and a Sub-Atmosperic Chemical Vapour Deposited (SACVD) TEOS layer (350 nm) capped by Plasma-Enhanced CVD TEOS, was deposited. Chemical Mechanical Polishing planarization was then carried out and the LIL mask was used to pattern both BL contacts and LIL source lines. Finally, contacts and lines were etched and filled with Tungsten. After W deposition, wafers were further processed up to metal 5 following the standard logic process flow.

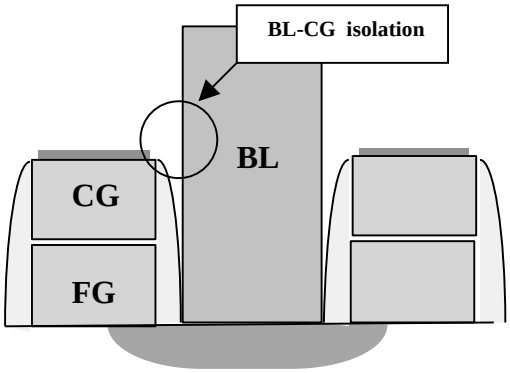

Figure 1. Schematic cross-section of a shared bitline (BL) contact between two Flash cells

BL-CG isolation was tested by performing *a)* leakage measurements between BL and CG (CG voltage = 15 V, BL voltage = 0 V); *b*) ramped Voltage-to-Breakdown (V_{bd}) measurements (CG start voltage = 0 V, ramp rate = 0.25 V/step, BL voltage = 0 V). Breakdown was

considered reached if a current of 50 nA was detected at the CG.

Leakage measurements were performed on 4 Mb and 256 Kb parallel arrays. The BL-CG distance was 120 nm in the 4 Mb arrays, whereas it varied in the range 140 nm – 60 nm in the 256 Kb arrays. Voltage-to-breakdown measurements were carried out on 4 Mb arrays without active.

3. LIL optimization

Figure 2 shows the BL-CG leakage yield (defined as the percentage of 4 Mb arrays with leakage smaller than 10^{-7} A) for several lots as a function of the maximum BL to CG misalignment in the direction perpendicular to the CG lines. Data in the figure reveal that the BL-CG alignment should be controlled within ± 70 nm to keep the BL-CG leakage under control. This alignment spec is 30 nm tighter than that used in logic for poly to contact alignment. The consequences of such tighter control would be an increased rework rate and/or the necessity of single-tool exposure for many Flash-specific critical layers. Both will have a negative impact on cycle time and manufacturability.

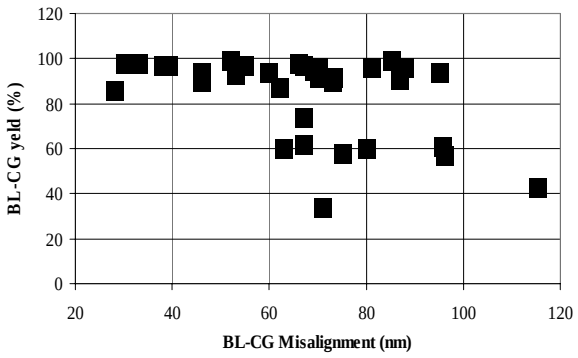

Figure 2. 4 Mb BL-CG leakage yield as a function of the BL-CG misalignment

An effective way to relax the ± 70 nm spec is a better control of the bitline CD. In fact, the BL contacts for the lots in Fig. 2 have been patterned using a binary LIL mask, resulting (see Fig. 3) in an average bitine CD of ≈ 0.28 μm, i.e. 40 nm bigger than the design rule (0.24 μm). The difficulty, however, is that in the LIL technology, both BL contacts and LIL source lines must be exposed at the same time. This makes it difficult to print with a binary mask smaller contacts while still controlling the line width. A possible solution could be provided by the Phase-Shift Mask (PSM) technology. With PSM the chrome layer is replaced by MoSi which transmits around 6% of light, but phase-shifted by 180 °. The image contrast is then improved by destructive interference between the phase-shifted light transmitted by the nominally dark area and the light coming from the clear areas [3]. Indeed, Figure 3 shows that the distribution of the bitline CD could be shifted down by ≈ 35 nm by using a phase-shift mask for LIL exposure.

The effect of the smaller bitline CD's on the BL-CG leakage is illustrated in Fig. 4. The figure shows the leakage distributions measured on 256 Kb arrays with 120 nm (Fig. 4a) and 100 nm (Fig. 4b) BL-CG distance. Hardly any difference is detected between binary and PSM if the BL-CG distance is 120 nm. However, if the distance is reduced by 20 nm, the yield drops to 60 % in the case of the LIL binary mask, whereas it is still ≈100 % in the case of the PSM mask. Further reduction of the BL-CG distance resulted in severe yield loss for both binary and PSM masks, thus indicating that the BL-CG alignment window can be improved by 20 nm by using the PSM mask.

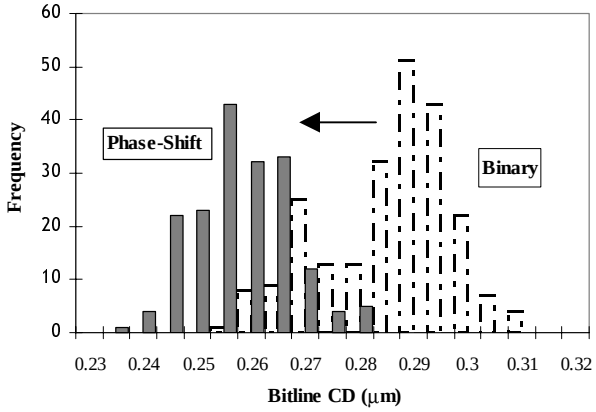

Figure 3. Bitline CD's (after etch) for binary or phase-shift LIL masks

4. ILD optimization

In addition to a better CD control, another possibility to increase the BL-CG alignment window is suggested by the TEM cross-section in Fig. 5a. The TEM picture shows that a crack is present in the nitride film (also called *boarderless* nitride) that is used as stop layer during ILD etch to allow LIL contacts and lines to overlap the STI-active border. Tungsten may diffuse during deposition into the crack if the BL is misaligned to the CG, therefore increasing the probability of BL-CG shorts. To improve conformality, nitride films were deposited using a dual-frequency chamber. In this chamber, a low-frequency rf power is applied to the wafer susceptor in addition to the high-frequency rf power used to ignite the plasma. This enhances the ion bombardment of the as-deposited films, thus improving the stability, density and conformality of the deposited layer [4]. Indeed, TEM cross-sections of nitride layers deposited in a dual-frequency chamber confirmed the absence of cracks at the top CG corners (see Fig. 5b). The cumulative distributions of the BL-CG leakage of wafers with standard or dual-frequency nitride are

compared in Fig. 6. Data show that if the BL-CG distance is reduced from 120 nm (6a) down to 100 nm (6b) the yield drops to 70 % in the case of the standard nitride whereas it is still ≈100 % in the case of the dual-frequency nitride. This demonstrates that the replacement of the standard logic boarderless nitride with a more conformal material would allow a 20 nm wider BL-CG alignment window.

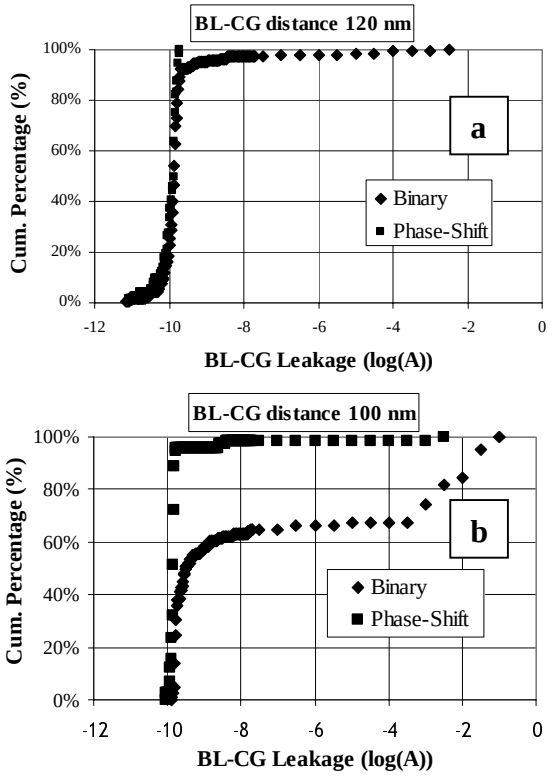

Figure 4. Cumulative distributions of the BL-CG leakage measured on 256 Kb arrays with two BL-CG distances: (a) 120 nm and (b) 100 nm. Binary and phase-shift masks are compared

5. Reliability measurements

Finally, the impact of process optimisation on the reliability of the BL-CG isolation was evaluated by comparing the V_{BD} measurements for lots processed with or without the optimised LIL and ILD modules. Figure 7 compares the cumulative percentage of the breakdown voltage for the two variants. In the case of the non optimised processing (squares), a significant percentage of the 4 Mb arrays have a V_{BD} value lower or slightly higher than the programming voltage (15 V, perpendicular line in Fig. 7). Arrays with V_{BD} lower than 15 V can be screened out by the BL-CG leakage measurements. Instead, arrays with V_{BD} slightly higher than the programming voltage represent a reliability hazard. In fact, their BL-CG isolation will breakdown when submitted to write/erase cycling during the device lifetime. In contrast, the percentage is almost zero if the optimised processing is used (diamonds).

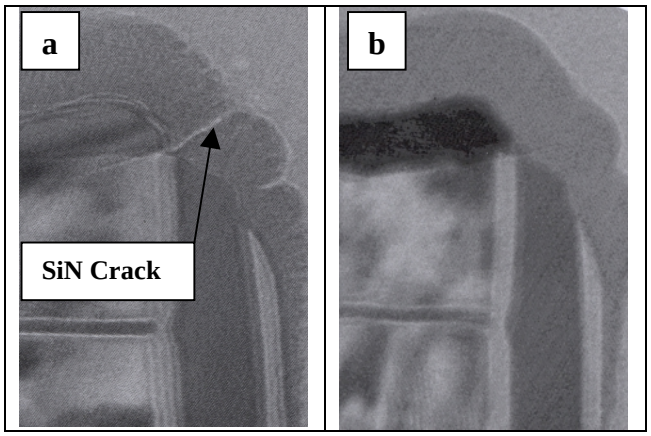

Figure 5. XTEM of memory cell with (a) standard (non-conformal) and (b) dual-frequency (conformal) boarderless nitride.

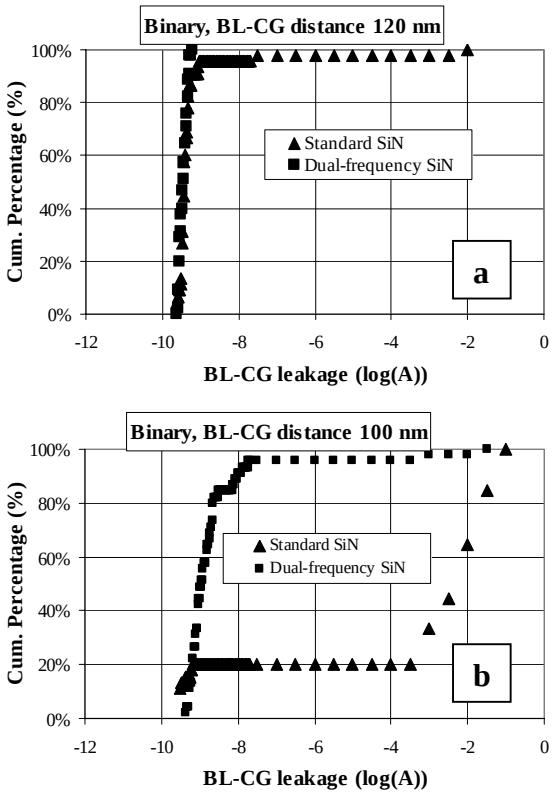

Figure 6. Cumulative distributions of the BL-CG leakage measured on 256 Kb arrays with two BL-CG distances: (a) 100 nm and (b) 80 nm. Conformal and non-conformal boarderless nitrides are compared

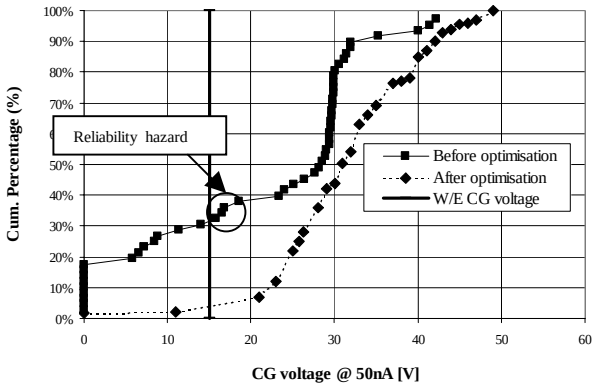

Figure 7. Cumulative probability of the Voltage-to-Breakdown for the BL-CG isolation measured on 4 Mb arrays.

6. Conclusions

We have shown that the integration of embedded 0.18 μm FLASH cells in a standard CMOS logic process implies not only the addition of some FLASH-specific process steps, but also the modification of relevant logic modules. In particular, it has been demonstrated that while for logic a binary mask at LIL and a non-conformal boarderless nitride can be tolerated, a phase-shift mask and a conformal nitride have to be used in the case of FLASH to guarantee a reliable and manufacturable bitline to control gate isolation.

7. References

[1] A. Cacciato *et al.* "ILD development for embedded 0.18μm FLASH technology" Proceedings of the European Solid-State Device Research Conference, Nuremberg, Germany, 11-13 September 2001, pp.435;

[2] D. Dormans *et al.* "Low-Voltage Embedded FLASH-EEPROM in 0.18 μm CMOS", Extended Abstract of the 2001 International Conference on Solid-State Devices and Materials, Tokyo, pp. 540-541

[3] M. Quirk, J. Serda "Semiconductor Manufacturing Technology" Prentice Hall Inc., p. 398;

[4] C.F. Lin *et al.* Jpn. J. Appl. Phys. **37** (1998), 6364