



HAL
open science

Flash gate optimized process and integration for electrical performances requirement on advanced embedded memory

El Amine Agharben, M. Bileci, A. Roussy, Marc Bocquet

► To cite this version:

El Amine Agharben, M. Bileci, A. Roussy, Marc Bocquet. Flash gate optimized process and integration for electrical performances requirement on advanced embedded memory. 2016 International Symposium on Semiconductor Manufacturing (ISSM), Dec 2016, Tokyo, Japan. 10.1109/ISSM.2016.7934533 . hal-01737950

HAL Id: hal-01737950

<https://hal.science/hal-01737950v1>

Submitted on 20 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Flash gate optimized process and integration for electrical performances requirement on advanced embedded memory

El Amine Agharben^{1,2,3}, M.Bileci¹, A.Roussy², M.Bocquet³
¹STMICROELECTRONICS, ²EMSE-CMP, ³IM2NP
Marseille, France
elamine.agharben@st.com

Abstract – In this paper a correlation between specific inline and electrical parameters related to erase threshold voltage has been deeply investigated. This study shows the impact of the WL roughness on a NOR flash memory performances. A process recipe optimization and Run to Run feed-forward controller have been evaluated to keep the process in control.

I. INTRODUCTION

NOR Flash memory is a type of Electronically Erasable Programmable Read Only Memory. This device is one of the most important media to store any type of data. With feature size decreasing rapidly in ultra-large scale integration (ULSI) circuits, chip performances have become increasingly affected by the process variation.

The device we studied consists of a select transistor for two word-lines (Fig.1). The cells are connected in parallel, with gates connected through the Word Line (WL), while the drains are shared along the Bit Line (BL). Programming and erasing are done respectively by Channel Hot Electron and Fowler-Nordheim mechanisms.

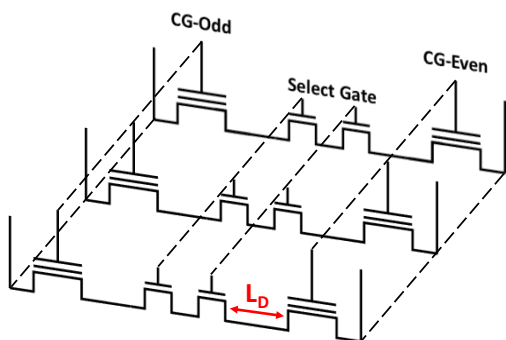


Figure 1. Device description

Several tests are performed to evaluate the device performances. The first reliability tests show a high spread of the erased cell threshold voltage V_{TE} (Fig.2). Given the large number of parameters impacting the transistor memory performances and the process complexity, a statistical analysis was performed on the “process manufacturing” in order to identify the main process contributors [1]. These V_{TE} variations were found to be primarily due to two components: the flash gate etch process and the process integration defining the

distance between the select transistor and the memory stack (L_D).

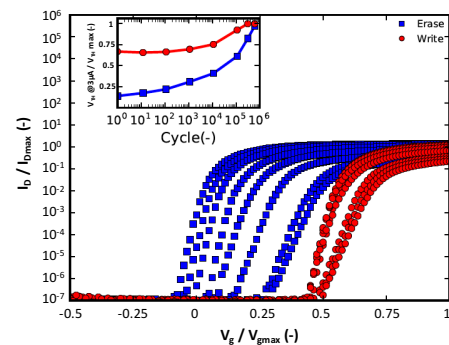


Figure 2. $I_D(V_G)$ distribution during cycling test

In this paper, the correlation between these specific inline and electrical parameters was deeply investigated. The impact of the WL roughness on the device architecture was underlined. To recover these marginalities a process recipe optimization and Run-to-Run control were developed to ensure reliable performances.

II. PROCESS IMPROVEMENT

In Flash process technology, the memory gate stack etch is one of the most challenging process steps. In our case the etch must go through an anti-reflective layer (ARL), a thick layer of polysilicon (Control Gate), an Oxide/Nitride/Oxide (ONO) dielectric and a second layer of polysilicon (Floating Gate), eventually stopping on a thin tunnel oxide.

Several inline measurements were performed in order to understand the impact of the flash gate etch process on the electrical performances. These measurements consist mostly on analyzing the flash gate Critical Dimension (CD) variability. A deep analysis helped to highlight a 10nm spread of the memory gate CD on the same WL. This spread can be assimilated to a Line Width Roughness (LWR) issue [2], [3].

In this device, the flash memory S/D regions are defined by the distance between the memory stack and the select transistor. It previously was proved that a fluctuation of S/D dose had a strong and direct effect on the V_{TE} during cycling tests. So it seemed likely that any variation of the distance L_D should directly affect the spread of the V_{TE} .

In order to improve the LWR of the memory gate, a new etch recipe was developed. This new recipe consists of using a different gas to etch the anti-reflective layer in order to minimize the line edge roughness generated by the photoresist pattern and intrinsic to this specific lithography process.

With this new etch recipe we were able to improve the memory gate CD spread on the same WL by 76% compared to the old recipe (Fig.3). Finally, the device performances were enhanced because of the V_{TE} spread improvement, by 87% in the worst cases (Fig.4).

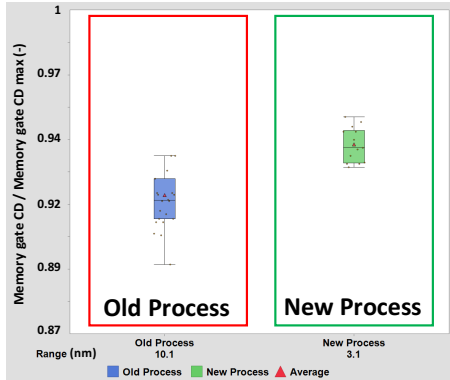


Figure 3. Memory gate CD spread on the same WL between old and new process

As mentioned above, this device performances are sensitive to the distance between the select transistor and the memory stack. Unfortunately, advanced products in embedded NOR Flash technology are affected by a large gate CD lot-to-lot variation. Traditional methods of CD control include a feedback component where the scanner energy dose is adjusted over time, based on the measured CD of the previous lot. In our case the inline measurement shows a lot-to-lot variation of the select gate CD.

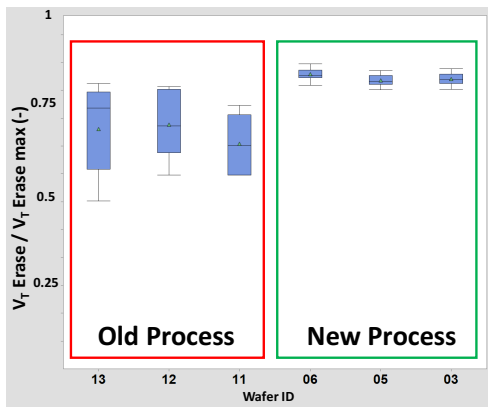


Figure 4. VTE spread between old and new process

The aim of this part is to propose a solution to maintain the distance between the memory stack and select transistor constant. In this context, the use of Advanced Process Control (APC) is mandatory to improve the variability induced by the processes and equipment limitation. Among all the APC components that have been considered, we chose a Run-to-Run

feedforward control scheme from select gate etching to memory gate etching [4]. Using feedforward control, the measured deviation of the post-etch select gate linewidth is automatically compensated by adjusting the gate CD. Although Run-to-Run (R2R) control is by far, accordingly to literature, the most effective technique to keep the process in control, this paper includes a novel application and explores new constraints.

III. RUN-TO-RUN DESCRIPTION

Considering the system defined in Fig.5, once a lot is measured, after select gate etching, the wafer proceeds to the memory cell etch step. The cell gate CD target is then adjusted to compensate the Select Gate CD variation. The controller requires a predictive model to adjust the memory gate etch recipe.

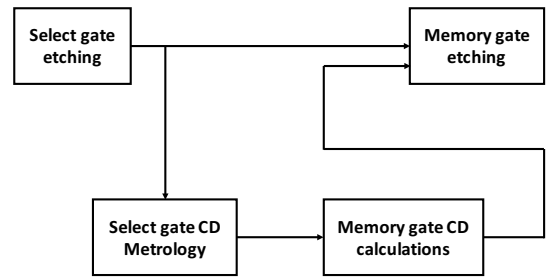


Figure 5. Run-to-Run description

A. The predictive Model

The model links the response variable Y (gate CD) to the gate etch parameters (pressure, coil RF power, time). By using a design of experiments, we worked out a response surface, that is a low order polynomial to be used as a single model [5]. Fig.6 shows the results of this analysis.

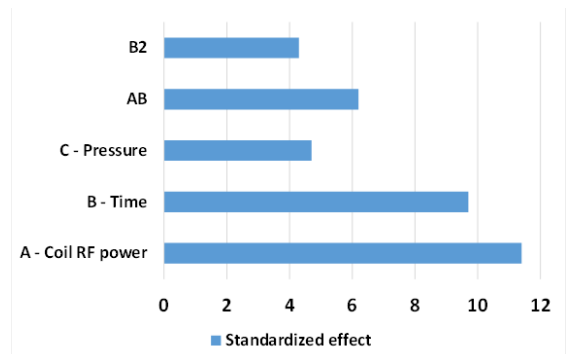


Figure 6. Standardized Pareto chart relative to the memory gate CD

A quadratic model was created for the memory gate CD:

$$Y = \varphi + \alpha.A + \beta.B + \gamma.C + \omega.AB + \varepsilon.B^2$$

where φ is the model centering

α, β, γ are coefficients of Coil RF power, Pressure and time

These coefficients depend on the etcher equipment and the model error is induced by different sources, like:

- Uncertainties in the model coefficients
- Mismatching between different equipment
- Mean time between failures

In order to calculate the actual parameters, we leveraged data from a 6 lots sample. This model reliability was verified (tested) for an actual production environment. The results are reported on Table.1:

Std Deviation	0.38	R-Squared	0.95
Adj R-square	0.94	Pred R-Square	0.87

Table 1. Model results

B. Run to Run implementation

As written before the new R2R controller impacts both inline and electrical parameters. Since this technology product is still in development mode, minimal trials were possible to check the capability of R2R controller. In this case five lots were used to simulate the worst cases of the select gate CD lot-to-lot variation.

At the initial implementation stage, we use a single equipment approach for each process step. All CD measurements were done with a scanning electron microscope. In case of positive results, the implementation of the R2R loop regulation will be considered for a larger number of equipment.

Based upon these criteria we defined two groups: the first (G1) consists of lots whose cell gate CD has been adjusted, while the second (G2) corresponds to lots processed with the standard conditions. When the CD of select gate is above a defined limit, the lot is manually managed because it mustn't be adjusted using the R2R controller. Moreover, in case of an automation fault, this action could be very useful.

Fig.7 shows the effect for each group with and without R2R controller, on the L_D parameter with respect to the select gate CD. We can observe a clear decrease of L_D variation when using the R2R controller. On top of this, the V_{TE} trend, plotted in Fig.8, shows a 60% reduction of the V_{TE} lot-to-lot variation.

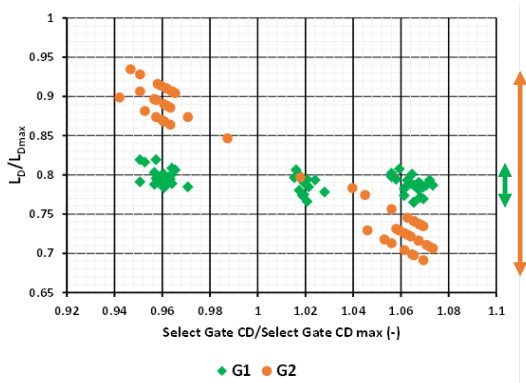


Figure 7. L_D parameter variation before and after using R2R controller

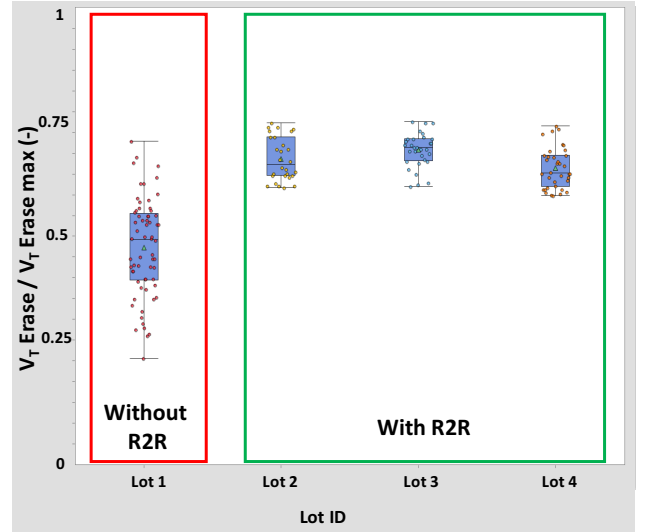


Figure 8. V_{TE} spread before and after R2R implementation

IV. RESULTS AND DISCUSSION

In this paper a new device has been characterized in order to evaluate the impact of inline parameters on electrical performances. The aim of this study is to propose solutions to better control the process and improve product reliability.

The inline analysis shows a high spread of the gate CD on the same WL. To minimize this effect, a new memory gate etch process was developed, the new recipe improved the WL roughness by 76% and enhanced the electrical performances. The inline analysis also reveals a lot-to-lot variation of the select gate CD. To recover these variations a R2R controller was proposed to compensate the select gate CD deviation by adjusting the memory gate CD.

On top of the electrical analysis already done, specific reliability tests are in progress to deeply evaluate the impact of the R2R controller on the device performances. Beside the positive results obtained, additional actions will be evaluated to improve the reliability of this R2R controller.

Several papers underlined the limitations of SEM equipment. It provides measurements with high spread mainly due to image focus limitations, while optical scatterometry offers higher accuracy and better repeatability. Replacing the SEM equipment by optical scatterometry will provide the same kind of information as SEM cross-section and several geometrical parameters. Like line thickness, sidewall angle, etc.

The purpose here is to use the inputs given by the scatterometry to create a filtering module for the R2R controller. In this case, we will use the Goodness Of Fit (GOF) as indicator for this module. It will filter out measurements with low GOF, and segregates lots in different groups defined by same bloc variables.

V. CONCLUSION

In the work here presented, we succeeded in decorrelate and quantify the impact of the WL roughness and L_D parameter variation on electrical cell performances in a new NOR flash memory. Besides, a R2R feed-forward controller between select gate etching and memory gate etching was proposed to keep the process in control. Therefore, additional investigation will be studied to further improve process control to get a stable electrical response.

VI. ACKNOWLEDGEMENTS

This research is part of the project MAGE, financed by the “Programme d’Investissement d’Avenir” of the French Government.

VII. REFERENCES

- [1] E. A. Agharben, A. Roussy, M. Bocquet, M. Bileci, S. Begouin, and A. Marchadier, “Critical sensitivity of flash gate dimension spread on electrical performances for advanced embedded memory,” in *2015 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 2015, pp. 401–404.
- [2] X. Jiang, R. Wang, T. Yu, J. Chen, and R. Huang, “Investigations on line-edge roughness (LER) and line-width roughness (LWR) in nanoscale cmos technology: Part i-modeling and simulation method,” *IEEE Trans. Electron Devices*, vol. 60, no. 11, pp. 3669–3675, 2013.
- [3] R. Wang, X. Jiang, T. Yu, J. Fan, J. Chen, D. Z. Pan, and R. Huang, “Investigations on Line-Edge Roughness (LER) and Line-Width Roughness (LWR) in Nanoscale CMOS Technology: Part II—Experimental Results and Impacts on Device Variability,” *IEEE Trans. Electron Devices*, vol. 60, no. 11, pp. 3676–3682, Nov. 2013.
- [4] S. Ruegsegger, A. Wagner, J. S. Freudenberg, and D. S. Grimard, “Feedforward control for reduced run-to-run variation in microelectronics manufacturing,” *IEEE Trans. Semicond. Manuf.*, vol. 12, no. 4, pp. 493–502, 1999.
- [5] D. C. Montgomery, *Design and analysis of experiments*. John Wiley & Sons, Inc, 2013.