Solid-State Electronics 199 (2023) 108498

Contents lists available at ScienceDirect

Solid State Electronics

journal homepage: www.elsevier.com/locate/sse

An inner gate as enabler for vertical pitch scaling in macaroni channel gate-all-around 3-D NAND flash memory[☆]

D. Verreck [*], A. Arreghini, G. Van den bosch, M. Rosmeulen

imec, 3001 Leuven, Belgium

ARTICLE INFO

Keywords:
Memory devices
CMOS-based memories
SRAM
DRAM
Flash

ABSTRACT

Scaling the vertical cell pitch to increase bit density in 3-D NAND flash memories degrades both the cell transistor characteristics and the memory operation. Here, we therefore investigate an inner gate to mitigate the scaling impact in macaroni channel devices. We evaluate several scenarios with varying complexity using calibrated TCAD simulations: from keeping the inner gate voltage grounded to coupling it to the read gate. We find a trade-off between improved cell transfer characteristics and program voltage determined by the inner gate to read gate coupling ratio.

1. Introduction

NAND flash memories are currently in an era of "happy scaling" thanks to the transition from 2-D arrays to 3-D vertical strings [1]. In a vertical architecture, all of the cells on a memory string are fabricated at the same time by the deposition of a layered stack, followed by etching the memory hole and filling it up with the memory stack and channel materials. The number of cells on a string, and thus bit density, can be increased by adding layers to the stack, while keeping the vertical cell pitch relaxed to avoid short channel effects. As the aspect ratio of the memory hole grows, however, the etch becomes so challenging that vertical pitch scaling is again required to limit the total stack height [2]. Unfortunately, the loss of gate control associated with the reduction of the cell dimensions results in undesirable effects such as an increasingly negative cell threshold voltage (V_{TH}) and an increase in programming voltage (V_{PGM}) [3]. In planar structures, the addition of a separate back gate for each cell has been proposed to mitigate these scaling effects [4]. The required alignment of both gates complicates the fabrication, however. One study suggested to simplify the structure to a single back gate, but without a detailed evaluation of the resulting device performance [5].

Here, we study with calibrated TCAD simulations whether a similar back gating principle can be applied to 3-D NAND macaroni channel strings by replacing the filler oxide inside the channel with an inner gate (IG). First, we consider the scaling behavior of cell characteristics and V_{PGM} of a reference device without IG, looking specifically at the impact of neighbor-induced barrier lowering (NIBL). Next, we compare this to a structure with an IG at a fixed voltage for varying equivalent oxide thickness (EOT_{IG}). Finally, we evaluate the coupling of the voltage on the IG to the read gate in varying ratios.

2. Structure and simulation flow

The three-gate structure under study is shown in Fig. 1(a) with a cross-section of the memory hole in Fig. 1(b) and simulation parameters in Fig. 1(c). This macaroni channel structure mimics our experimental test vehicles, with the addition of an IG contact and corresponding dielectric in the center of the memory hole. The IG covers the entire backside of all three gates, but is not connected to the source or drain contact. The vertical pitch is scaled by varying the gate lengths (L_G) and intergate spacings (L_{IGS}).

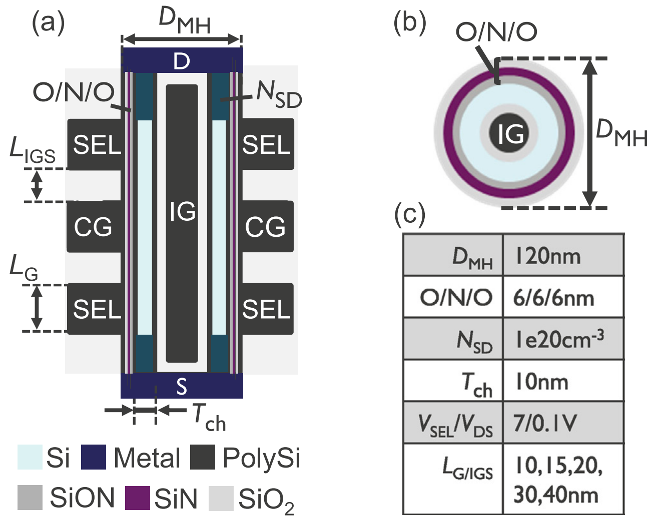
The simulations are performed with the Minimos tool in the Global TCAD Solutions software package [6], which incorporates a jointly developed charge trap layer (CTL) memory operation model [3]. During program, carriers are injected with a Wentzel–Kramers–Brillouin tunneling approach and then distributed over the CTL according to a Gaussian profile. The carriers are captured in the CTL through a Shockley–Read–Hall mechanism with a constant trap density throughout the layer. The parameters related to the tunneling, CTL transport and trapping models have been calibrated to our in-house experimental data [3]. For the read operation, a standard drift-diffusion model is employed in the channel layer, combined with the solution of Poisson's equation. To limit computational time, all simulations utilize the cylindrical symmetry of the device structure. Since we did not include the channel grains or traps, the parameters related to the channel (V_{TH}, I_{ON}) should be interpreted qualitatively.

[☆] The review of this paper was arranged by Francisco Gamiz.

[*] Corresponding author.
E-mail address: devin.verreck@imec.be (D. Verreck).

https://doi.org/10.1016/j.sse.2022.108498

Available online 29 October 2022
0038-1101/© 2022 Elsevier Ltd. All rights reserved.



D_{MH}	120nm
O/N/O	6/6/6nm
N_{SD}	1e20cm^{-3}
T_{ch}	10nm
V_{SEL}/V_{DS}	7/0.1V
$L_{G/IGS}$	10,15,20, 30,40nm

Fig. 1. (a) Simulated three-gate macaroni channel 3-D NAND flash string with a center cell under study (CG) and two neighboring cells (SEL) (b) memory hole cross-section along a cutline through the center of the CG and (c) simulation parameters.

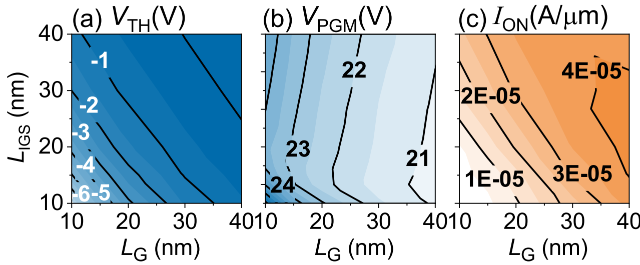


Fig. 2. Impact of scaling L_G, L_{IGS} for the reference device without an IG on (a) cell V_{TH} of CG extracted at I_{OFF} of 1e-9 A/µm, (b) V_{PGM} required to reach ΔV_{TH} of 5 V on CG and (c) device current at V_{CG} = V_{TH}+2V.

3. Reference case simulation results

We first simulate the scaling behavior for the reference case without an IG and find a strongly negative V_{TH} shift for smaller vertical pitch dimensions (Fig. 2(a)), combined with an increased V_{PGM} (Fig. 2(b)) and decreased on-current (I_{ON}) (Fig. 2(c)). Scaling either L_G or L_{IGS} has a similar effect. For the smallest pitches, V_{TH} shifts down to −6 V, which would require the generation of large negative voltages during the operation of the memory string, which is generally undesired. Such a magnitude in V_{TH} shift would also be challenging to correct for with gate work function engineering.

Fig. 3 explains the origin of the degraded cell characteristics to be a loss in gate control for the center gate (CG). Fig. 3(a) shows the increasing influence of the side gates on the electrostatic potential underneath the CG for smaller L_{IGS}. This corresponds to a loss of gate control for the CG on the channel region underneath, resulting in a reduced energy barrier in the subthreshold regime (Fig. 3(b)) and therefore a negatively shifted V_{TH}. This effect, which we refer to as neighbor-induced barrier lowering (NIBL), has a relatively larger impact at shorter L_G. For the depicted L_G of 10 nm, the NIBL is already significant for an L_{IGS} of 30 nm. The loss in gate control for the CG is also at the origin of the degradation in V_{PGM} and I_{ON} [3].

4. Inner gate simulation results

Next, we introduce the IG with an EOT_{IG} equal to the EOT of the ONO stack (11.5 nm) and ground it to 0 V, which could be implemented

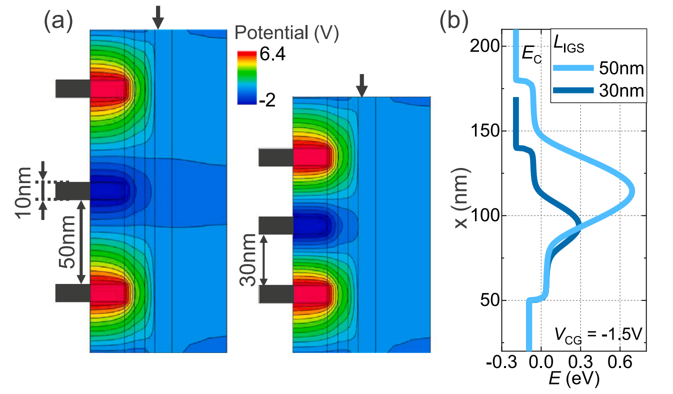


Fig. 3. (a) Electrostatic potential contour plots of the reference case without inner gate during the read operation for varying L_{IGS} at fixed L_G. (b) Conduction band profile along the length of the channel at 1 nm from the tunnel oxide interface, as indicated by the arrows in (a).

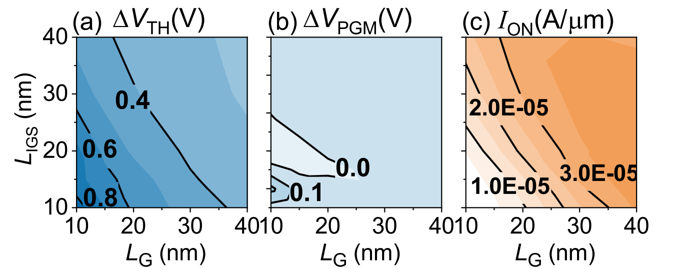


Fig. 4. L_G, L_{IGS} scaling for device with IG with V_{IG} = 0 V and EOT_{IG} = 11.5 nm. (a) V_{TH} difference of CG relative to non-IG case (b) V_{PGM} difference relative to non-IG case and (c) device current at V_{CG} = V_{TH}+2V.

e.g. through a connection with the source contact. We find that a V_{IG} of 0 V has a small positive impact on the cell V_{TH} relative to the reference case (Fig. 4(a)), while V_{PGM} is almost unaffected (Fig. 4(b)). I_{ON} is slightly degraded for the larger pitches (Fig. 4(c)). Reducing EOT_{IG} to 5 nm increases the positive ΔV_{TH} (Fig. 5(a)), while keeping the impact on V_{PGM} and I_{ON} limited (Figs. 5(b–c)). The positive shift in V_{TH} is a straightforward consequence of the increasing influence of the inner gate on the CG channel potential, which becomes stronger as EOT_{IG} is scaled. However, since the IG covers also the side gate channels and the IGS regions, it counters the pass gate potential induced there and decreases the carrier concentration. This results in an overall increase in string resistance and corresponding reduction of I_{ON}. For scaled vertical pitch, however, this effect is less severe as the string is shorter.

The band diagram along the CG center in Fig. 6 explains that V_{PGM} is insensitive to V_{IG} because the channel screens the IG potential, leaving the electric field over the ONO stack quasi unaffected. This remains true even for significantly negative V_{IG}. The screening effect of the channel therefore also prevents the IG from being used to improve programming behavior, and we therefore keep V_{IG} at 0 V during program for subsequent cases.

In the next scenario, we connect IG to CG during the read operation, which results in a significant improvement of V_{TH} and I_{ON} (Figs. 7(a) and (c)). The improvement in V_{TH} goes up to 4 V for the most scaled cases. This is thanks to the double gate action of IG and CG, strongly improving the gate control over the channel and counteracting the NIBL from the side gates. V_{PGM} is strongly increased, however (Fig. 7(b)). During the read operation, the channel is now turned on by both CG and IG and since charge is only programmed in the CTL on one side of the channel, its control on V_{TH} is strongly reduced compared to the reference case without the IG.

Finally, we couple the voltage on IG and CG with a varying ratio (V_{IG} = V_{CG}/D), which results in a trade-off between V_{TH} and I_{ON} on

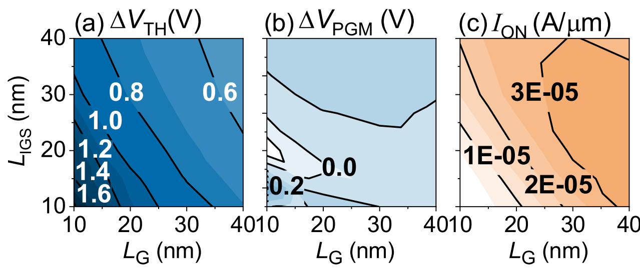


Fig. 5. L_G, L_{IGS} scaling for device with IG with V_{IG} = 0 V and EOT_{IG} = 5 nm. (a) V_{TH} difference of CG relative to non-IG case (b) V_{PGM} difference relative to non-IG case and (c) device current at $V_{CG} = V_{TH}$+2V.

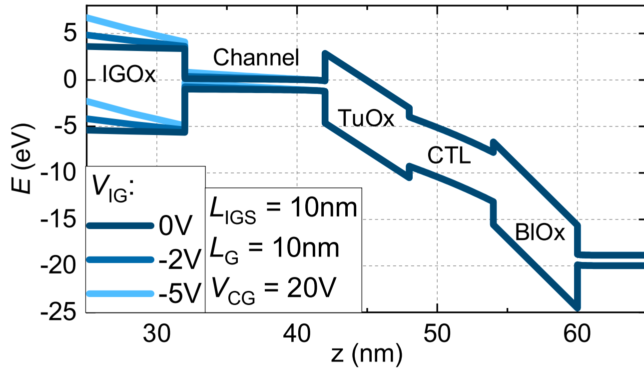


Fig. 6. Energy band diagram along a cutline through the center of CG, orthogonal to the channel, for varying V_{IG} during the program operation. z = 0 nm corresponds to the center of the memory hole.

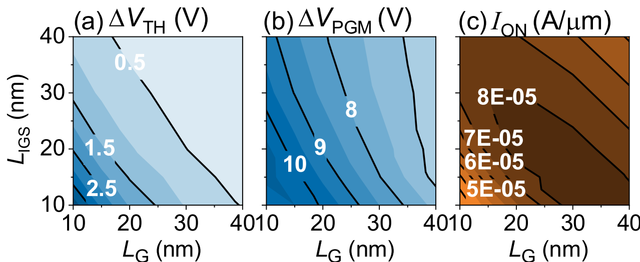


Fig. 7. L_G, L_{IGS} scaling for device with IG with $V_{IG} = V_{CG}$ during the read operation and EOT_{IG} = 11.5 nm. (a) V_{TH} difference of CG relative to non-IG case (b) V_{PGM} difference relative to non-IG case and (c) device current at $V_{CG} = V_{TH}$+2V.

one hand and V_{PGM} on the other (Figs. 8(a–c)). The results interpolate between the V_{IG} = 0 V and coupled $V_{IG} = V_{CG}$ cases discussed above. Depending on the application, a compromise can be found, e.g. for D = 2, a 2.5 V V_{TH} improvement is achieved at the cost of a 17% V_{PGM} increase at the smallest pitch of $L_G = L_{IGS}$ = 10 nm.

5. Conclusion

We conclude that an inner gate can be used to mitigate negative V_{TH} shifts and improve I_{ON} in vertically scaled 3-D NAND cells at the expense of required programming voltage. We found that the trade-off depends on the coupling ratio of the inner gate to the read gate. Additionally, we showed that an inner gate has very limited impact on the program operation due to the screening of the inner gate potential by free carriers in the channel.

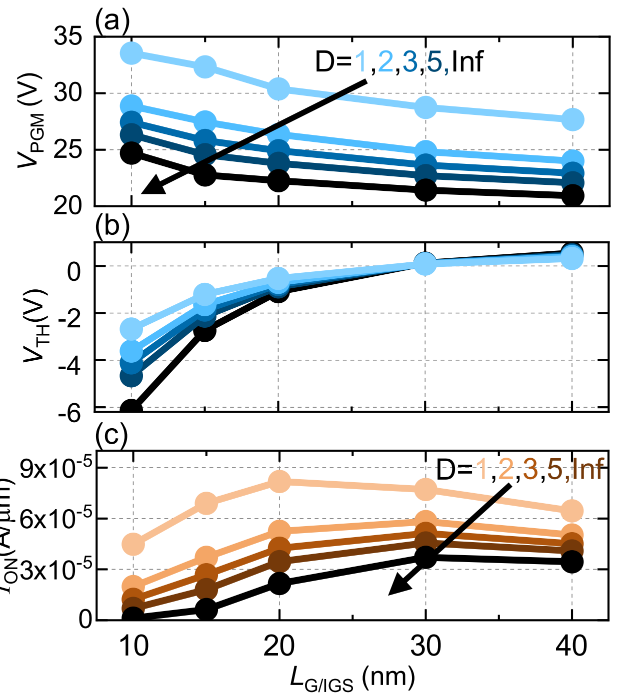


Fig. 8. Vertical pitch scaling impact with $L_G = L_{IGS}$ for a connection between IG and CG during the read operation with varying ratio $V_{IG} = V_{CG}$/D and EOT_{IG} = 11.5 nm. (a) V_{PGM} for ΔV_{TH} of 5 V, (b) V_{TH} of CG at an I_{OFF} of 1e-9 A/µm and (c) I_{ON} at $V_{CG} = V_{TH}$+2V. D = Inf corresponds to V_{IG} = 0 V.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by imec's Industrial Affiliation Program for storage memory.

References

[1] Alsmeier J, Higashitani M, Paak SS, Sivaram S. Past and Future of 3D Flash. In: 2020 IEEE International Electron Devices Meeting. IEEE; 2020, p. 6.1.1–4.
[2] Yanagihara Y, Miyaji K, Takeuchi K. Control gate length, spacing and stacked layer number design for 3D-stackable NAND flash memory. In: 2012 4th IEEE International Memory Workshop. IEEE; 2012, p. 1–4.
[3] Verreck D, Arreghini A, Schanovsky F, Rzepa G, Stanojevic Z, Mitterbauer F, et al. Understanding the ISPP Slope in Charge Trap Flash Memory and its Impact on 3-D NAND Scaling. In: 2021 IEEE International Electron Devices Meeting. 2021, p. 1–4.
[4] Walker AJ. Sub-50-nm Dual-Gate Thin-Film Transistors for Monolithic 3-D Flash. IEEE Trans Electron Devices 2009;56(11):2703–10.
[5] Lin H-C, Lin Z-M, Chen W-C, Huang T-Y. Read Characteristics of Independent Double-Gate Poly-Si Nanowire SONOS Devices. IEEE Trans Electron Devices 2011;58(11):3771–7.
[6] Global TCAD Solutions Rel.2022.03, URL http://www.globaltcad.com/.