

Operational Voltage Reduction of Flash Memory Using High- κ Composite Tunnel Barriers

Sarves Verma, *Student Member, IEEE*, Eric Pop, Pawan Kapur, Krishna Parat, and Krishna C. Saraswat, *Fellow, IEEE*

Abstract—We explore the performance of symmetric (low- κ /high- κ /low- κ) and asymmetric (low- κ /high- κ) composite tunnel barriers with conventional Flash constraints of retention, erase, read and program disturbs. Simulations, including five different high- κ materials, were performed under these criteria to minimize the programming voltage V_{prog} . Among all constraints, we find read disturb to be the most restrictive both in terms of lowering V_{prog} and choosing the high- κ materials for such stacks. Furthermore, the symmetric barrier stack is found to be more promising versus the asymmetric barrier stack. For the common read disturb voltages of 2.5 and 3.6 V, the lowest V_{prog} of ~ 4 and 5 V, respectively (relative to the floating gate), are obtained. In addition, the maximum required operating Flash voltage is found to be 30%–40% lower than the prevalent voltages.

Index Terms—Flash memory, Flash operating constraints, high- κ dielectrics, program disturb, read disturb, retention, tunnel barrier engineering.

I. INTRODUCTION

THE LACK of voltage scaling in conventional Flash memory presents a serious bottleneck [1]. Continuous complementary metal–oxide–semiconductor supply voltage (V_{dd}) reduction further increases the discrepancy between V_{dd} (~ 1 V) and Flash operating voltages (~ 15 V). Thus, Flash structural modifications that yield lower operating voltages are imperative. One approach is to replace the SiO_2 tunnel oxide with a composite stack of different materials. High- κ dielectrics with larger thicknesses for the same equivalent oxide thickness (EOT) [2] serve as ideal stack materials. Past work has explored crested composite barriers [3] and VARIOT [4], [5]. However, to date, a Flash-compliant approach to optimize the design space for minimum programming voltage (V_{prog}) and EOT has not been presented. In this letter, we quantify the minimum V_{prog} and EOT for a composite tunnel barrier stack while accounting for the retention, read and program disturb criteria,

Manuscript received November 20, 2007. The review of this letter was arranged by Editor S. Chung.

S. Verma is with the Center for Integrated Systems, Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: sarves@stanford.edu).

E. Pop was with Intel Corporation, Santa Clara, CA 95054-1549 USA. He is now with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA.

P. Kapur and K. C. Saraswat are with the Center for Integrated Systems, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: saraswat@stanford.edu).

K. Parat is with Intel Corporation, Santa Clara, CA 95054-1549 USA.

Color versions of one or more of the figures in this letter are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LED.2007.915376

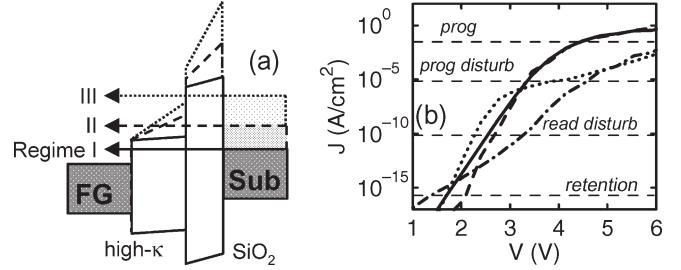


Fig. 1. (a) Electron substrate injection in regimes I–III. (b) J – V characteristics of the asymmetric stack (solid line, $T_{\text{ox}} = 2$ nm), the symmetric stack (dashed line, $T_{\text{ox}} = 2$ nm on either side), and the pure SiO_2 stack (dash-dotted line) with a 6-nm total EOT. The dotted line is for a thicker $T_{\text{ox}} = 3.5$ nm in the asymmetric 6-nm EOT stack. The four horizontal dashed lines are the Flash constraints.

thus also alluding to the best material combination and the geometry of the stack.

II. TUNNEL BARRIER ENGINEERING

In practice, the crested barrier approach [3] is difficult as it requires an abrupt high- κ interface with silicon. Thus, we focus on the alternate approach with a low- κ /high- κ /low- κ (symmetric) or low- κ /high- κ (asymmetric) stack. Such composite barriers can yield steeper current–voltage (J – V) characteristics, resulting in a lower current at a low voltage (retention) and a higher current at a higher voltage (program). Thus, V_{prog} can be reduced while meeting retention.

Fig. 1(a) elucidates the operation of the asymmetric composite barrier. With an increasing gate bias, three distinct regimes for substrate injection of electrons are observed. Regime I (low gate bias) corresponds to direct tunneling (DT) through the entire stack. In regime II, electrons tunnel into a high- κ conduction band [fast Fowler–Nordheim (FN) transport]. Regime III consists of DT only through the low- κ barrier. Fig. 1(b) shows that the simulated composite barrier yields higher nonlinearity in J – V , as compared to the SiO_2 alone, because in regime II, the current increases due to both higher E-field and high- κ FN barrier lowering. By comparison, the barrier height of the SiO_2 tunnel dielectric is fixed.

For a given EOT of an asymmetric stack, a thicker SiO_2 layer (T_{ox}) yields steeper J – V 's in regime II, but the transition to regime III occurs at a lower voltage [Fig. 1(b)]. Thus, as long as the programming current (J_{prog}) is in regime II, an increase in T_{ox} reduces V_{prog} . However, at a higher T_{ox} , J_{prog} transfers to regime III, where the J – V slope is lower. This results in an optimum T_{ox} , minimizing V_{prog} [Fig. 2(a)].

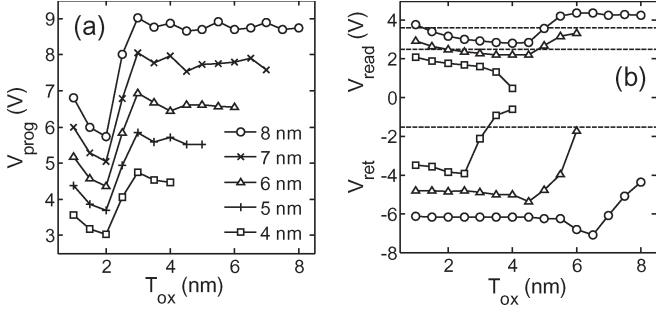


Fig. 2. (a) V_{prog} scaling with T_{ox} for the different asymmetric EOTs with HfO_2 (note the minima). (b) Read disturb (top) and retention voltage (bottom) scaling with T_{ox} for the 4-, 6-, and 8-nm EOT stacks. The top horizontal dashed lines correspond to the 2.5- and 3.6-V read disturb voltages. The bottom one corresponds to the -1.5-V retention. Symbols for the different EOTs are consistent in both figures.

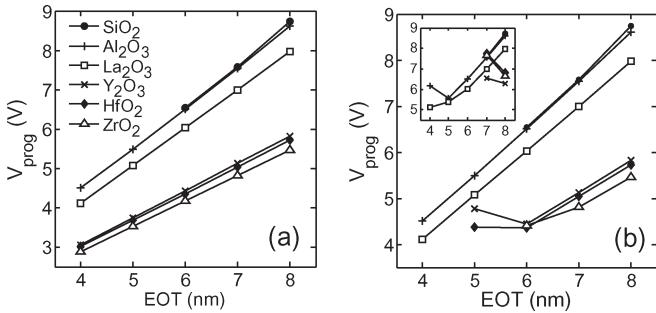


Fig. 3. Optimal V_{prog} at each EOT for all asymmetric stacks and high- κ materials. (a) Imposes only the retention constraint. (b) Adds in the $V_{\text{read}} = 2.5$ -V read disturb and considers the more restrictive 3.6 V in the inset. Symbols are consistently used in the figures.

III. METHODOLOGY

For a given high- κ material stack, we first fix the total EOT and vary T_{ox} . For each T_{ox} , we calculate the J - V curve using the transfer matrix formulation [6]. From this, we obtain V_{prog} at J_{prog} ($\sim 3 \times 10^{-2} \text{ A/cm}^2$, NAND Flash). By repeating for different EOTs, we get a family of V_{prog} versus T_{ox} curves; each curve exhibits an EOT-specific minimum [Fig. 2(a)]. Next, we impose the retention and read [Fig. 2(b)] and program disturb (not shown) constraints such that at gate voltages corresponding to each condition, J does not exceed a maximum amount. This narrows the allowed T_{ox} range for each EOT. Thus, the minimum V_{prog} for each EOT is taken to be the lowest voltage [in Fig. 2(a)] only in the restricted T_{ox} domain (Fig. 3). The analysis assumes that the intrinsic current dominates retention and other constraints. In practice, trap-assisted tunneling can be important. Thus, this Letter represents the best-case scenario and yields correct qualitative trends. The methodology also identifies the minimum EOT below which no associated T_{ox} satisfies the Flash constraints. We also consider several high- κ materials to get the lowest possible V_{prog} and, in the process, to also identify the best material set, the lowest EOT, and the optimum T_{ox} for that EOT. A minimum erase voltage V_{erase} ($J_{\text{erase}} > 7 \times 10^{-3} \text{ A/cm}^2$) was also obtained for all material sets. Ultimately, the maximum operating voltage of the Flash cell [on the floating gate (FG)], which is the maximum of V_{prog} and V_{erase} , was minimized across different material sets, EOT, and T_{ox} values.

Both asymmetric and symmetric composite barriers were examined with EOTs from 4 to 8 nm. For each EOT, T_{ox} was varied from 1 nm up to EOT (asymmetric) and EOT/2 (symmetric barrier), and a high- κ thickness was correspondingly adjusted. High- κ simulation parameters were taken from [7].

IV. RESULTS AND DISCUSSION

Six-nanometer EOT $\text{HfO}_2/\text{SiO}_2$ symmetric and asymmetric J - V curves reveal higher nonlinearity compared to pure SiO_2 with the same EOT [Fig. 1(b)]. In addition, the symmetric curve is shifted to the right, compared to the asymmetric curve because of later FN tunneling onset. Thus, although, for a given J_{prog} , the asymmetric stack is programmed faster, it is more susceptible to disturbs, making the symmetric/asymmetric choice highly constraint dependent.

For the read constraint, we consider two possible voltages on the FG during read (V_{read}): 1) 3.6 V (current industry standard) and 2) 2.5 V (corresponds to International Roadmap for Semiconductors projections [8]). The maximum tolerable substrate-injected J to prevent state contamination during a single read is $7 \times 10^{-11} \text{ A/cm}^2$. During programming, the maximum tolerable J to prevent program disturb is $7 \times 10^{-6} \text{ A/cm}^2$ and should be met at $V_{\text{prog}}/2$. Finally, during retention, the maximum allowable gate-injected current density is $2 \times 10^{-16} \text{ A/cm}^2$ for a ten-year retention, and the typical voltage on the FG during retention V_{ret} is about 1.5 V.

Fig. 2(b) depicts the maximum allowable retention and read voltages on the FG as a function of T_{ox} (only HfO_2 asymmetric is shown) so as not to exceed the aforementioned retention and read disturb J 's. These curves are directly obtained using such J - V characteristics as in Fig. 1(b). The Flash constraints are only met if these voltages (magnitude) are higher than the aforementioned read disturb and retention voltages [dashed lines in Fig. 2(b)]. These constraints are met only for a subset of the T_{ox} domain. For every EOT, we then transfer the selected domain back to Fig. 2(a) to choose the minimum V_{prog} in that domain.

Fig. 3 shows the minimum possible V_{prog} (all considered high- κ 's) as a function of EOT. Fig. 3(a) only uses the retention criterion, whereas Fig. 3(b) uses all constraints, including read and program disturbs. In Fig. 3(a), the HfO_2 , ZrO_2 , and Y_2O_3 stacks yield the lowest V_{prog} (~ 2.9 V) and an EOT as low as 4 nm. The above V_{prog} must be divided by the gate coupling ratio (~ 0.6) to obtain the bias at the control gate. In addition, La_2O_3 has the highest dielectric constant but still fails to provide a low V_{prog} due to its higher band offset.

Adding read disturb ($V_{\text{read}} = 2.5$ V) further restricts the allowed T_{ox} domain [Fig. 3(b)]. For HfO_2 , ZrO_2 , and Y_2O_3 , this increases the minimum EOT and V_{prog} [the global minimum of Fig. 2(a) no longer falls in the allowed domain]. Whereas for La_2O_3 , adding read disturb does not affect V_{prog} because it exhibits a less steep J - V (higher band offset). This results in a lower J at V_{read} , allowing it to easily satisfy read disturb. For the more stringent read voltage condition of 3.6 V [inset Fig. 3(b)], HfO_2 , ZrO_2 , and Y_2O_3 satisfy read disturb only down to a 7-nm EOT. La_2O_3 yields the best results with EOTs as low as 4 nm and a corresponding V_{prog} of ~ 5.1 V on

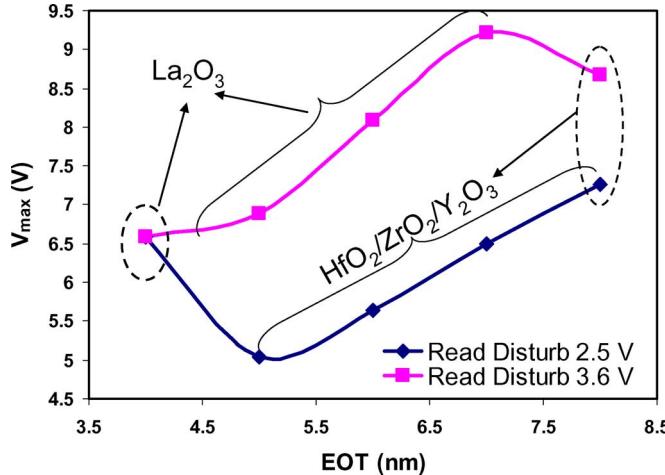


Fig. 4. Global performance optimization for all stacks in consideration, irrespective of the high- κ material considered. Furthermore, all constraints (retention, erase, and read and program disturbs) have been taken into account to find the maximum required operating voltage (V_{\max}) for Flash memory. Note that only La_2O_3 sustains the stringent read disturb criterion of 3.6 V.

the FG. Similar results were obtained for symmetric stacks (not shown). We also find that the program disturb constraint is less stringent and does not modify these conclusions.

Finally, Fig. 4 shows the maximum possible operational voltage V_{\max} (maximum of V_{prog} and V_{erase} on the FG) versus EOT. The optimization runs across all considered high- κ materials. We find that La_2O_3 performs best for a strict read disturb criterion of 3.6 V, whereas HfO_2 outperforms other high- κ materials for a 2.5-V read disturb. HfO_2 enables the lowest V_{prog} (~ 3.95 V) at an EOT of 5 nm, whereas La_2O_3 gives the lowest V_{prog} (~ 5.1 V) at an EOT of 4 nm. Global optimization yields the largest operational voltage (V_{\max}) of $\sim 5\text{--}7$ V, constituting a $\sim 30\text{--}40\%$ voltage reduction over the conventional SiO_2 -based Flash cells. We observe that read disturb is more constraining than the retention criterion and results in a higher possible operating voltage. Furthermore, the

asymmetric stack (compared to the symmetric stack) is found to have a higher V_{erase} (not shown). Thus, the symmetric stacks are more promising under these conditions.

V. CONCLUSION

We examined composite tunnel barrier stacks for obtaining minimum V_{prog} and V_{erase} under typical Flash constraints of retention and read and program disturbs. These voltages were obtained across five different high- κ materials, symmetric and asymmetric stack configurations, different EOTs, and T_{ox} . We found that the read disturb is the most constraining factor, partly restricting the advantage offered by the composite barrier. A stringent read disturb limits the use of HfO_2 , ZrO_2 , and Y_2O_3 in such composite barriers, whereas the La_2O_3 stack is found to be more scalable. We also found that symmetric barrier stacks are more promising (versus asymmetric barrier stacks) and that a V_{prog} of $\approx 4\text{--}5$ V on the FG is ideally achievable.

REFERENCES

- [1] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to Flash memory," *Proc. IEEE*, vol. 91, no. 4, pp. 489–502, Apr. 2003.
- [2] J. Robertson, "High dielectric constant gate oxides for metal oxide Si transistors," *Rep. Prog. Phys.*, vol. 69, no. 2, pp. 327–396, Feb. 2006.
- [3] K. K. Likharev, "Layered tunnel barriers for nonvolatile memory devices," *Appl. Phys. Lett.*, vol. 73, no. 15, pp. 2137–2139, Oct. 1998.
- [4] B. Govoreanu, P. Blomme, M. Rosmeulen, J. Van Houdt, and K. De Meyer, "VARIOT: A novel multilayer tunnel barrier concept for low-voltage non-volatile memory devices," *IEEE Electron Device Lett.*, vol. 24, no. 2, pp. 99–101, Feb. 2003.
- [5] J. Van Houdt, "High- κ materials for nonvolatile memory applications," in *Proc. 43rd Annu. IEEE IRPS*, San Jose, CA, 2005, pp. 234–239.
- [6] Y. Ando and T. Itoh, "Calculation of transmission tunneling current across arbitrary potential barriers," *J. Appl. Phys.*, vol. 61, no. 4, pp. 1497–1502, Feb. 1987.
- [7] C. L. Hinkle, C. Fulton, R. J. Nemanich, and G. Lucovsky, "A novel approach for determining the effective tunneling mass of electrons in HfO_2 and other high- κ alternative gate dielectrics for advanced CMOS devices," *Microelectron. Eng.*, vol. 72, no. 1–4, pp. 257–262, May 2004.
- [8] International Technology Roadmap for Semiconductors (ITRS), *Emerging Research Devices*, 2005. [Online]. Available: <http://public.itrs.net>