Feasibility Study of Composite Dielectric Tunnel Barriers for Flash Memory

Sarves Verma¹, Eric Pop², Pawan Kapur³, Prashant Majhi³, Krishna Parat², Krishna C. Saraswat¹

1 Center for Integrated Systems, Stanford University, 20 Via Palou, MS 4075, Stanford, CA 94305
2. Intel Corporation, Phone: 650-725-3670, email: sarves@stanford.edu

Lack of voltage reduction presents a serious impediment in future flash memory scaling. Replacing conventional SiO₂ tunnel dielectric with a composite dielectric material (combination of high-κ and SiO₂ layers) stack (Fig.1) potentially yields a powerful approach to achieve voltage reduction. The method relies on obtaining non-linearity in gate current vs. gate voltage (Vgs) characteristics, such that, low voltages (corresponding to retention and read disturb (Vprog)) render a lower current; whereas, high voltages (corresponding to program and erase) result in a higher current. Related past work involves using a composite stack with a smaller barrier material (high-κ) between two large barrier materials (SiO₂) [1] and a crested barrier approach [2]. However, in these systems, to-date, a top-down approach to optimize the design space for minimum programming voltage (Vprog) and equivalent oxide thickness (EOT), while meeting flash retention, read and program-disturb constrains, has not been pursued. Such an approach identifies the best high-κ option by comparisons, yielding lowest possible flash operational voltages.

Towards this end, we explore both the symmetric (low-κ/high-κ/low-κ) and the asymmetric (low-κ/high-κ) composite tunnel barriers using several high-κ materials. There are three possible tunneling mechanisms through a composite barrier depending on the Vgs (Fig. 2a). A corresponding simulated current density (J-Vgs) (Fig. 2b) indeed demonstrates a higher non-linearity for composite barrier structure compared to the conventional SiO₂ dielectric, thus, yielding a lower Vprog. The higher non-linearity stems from a lowering of both the high-κ barrier (modulated by voltage across low-κ) and the increased E-field with Vgs. By comparison, for conventional SiO₂ tunnel dielectric, the barrier height is fixed and only one factor- increase in E-field is responsible for the increase in current with Vgs.

Five different high-κ materials (HfO₂, La₂O₃, Y₂O₃, ZrO₂, Al₂O₃) were explored using an in-house simulator based on the transfer matrix approach [3] for current calculations. We first choose the high-κ material. Next, we fix the total EOT (high-κ + SiO₂ layers) and vary SiO₂ thickness (Tox). For each tox, we simulate the J-Vgs curves, and obtain the Vprog at the required programming current density (Jprog = 3 × 10¹² A/cm², typical for Nand Flash). By repeating this for different EOTs, we get a family of Vprog vs. tox curves, each exhibiting a minimum Vprog (Fig. 3a). Similarly, Fig. 3b shows the voltages corresponding to the maximum allowed retention (Jret < 2 × 10¹⁵ A/cm²) and read-disturb (Jread < 7 × 10¹¹ A/cm²) current densities (shown as horizontal dashed lines, also see Fig. 2b), along with the actual voltages encountered during these conditions. In order to meet these constrains, the voltages obtained should be larger (in absolute value) than the horizontal dashed lines. Thus, these constrains manifest themselves in limiting the allowed tox range for certain EOTs, resulting in a domain down-selection in Fig. 3a (for Vprog). Combining the tox domain down-selection with the Vprog vs. tox plot, we obtain the optimum Vprog for each EOT (Fig. 4). The curve also reveals the minimum possible EOT below which no tox satisfies the Flash constrains. We repeat the process for different high-κ materials to get the lowest possible Vprog along with the 1) best material set, 2) the lowest EOT, and 3) the optimum tox for that EOT (Fig. 4a,b). The effect of adding constrains can be seen in Fig. 5. In general, constrains increase Vprog when compared to the minimum obtained in Fig. 3a, except for program disturb (corresponding to Jprog < 7 × 10¹⁵ A/cm² refer Fig. 2b) which is always satisfied. Further, upon extraction of erase voltage (Jerase > 7 × 10¹⁶ A/cm²) (Fig. 6), the asymmetric stack was found to have a larger Verase than the symmetric one leading to an important conclusion that symmetric stacks are more promising. Finally, Fig. 7 shows the maximum possible operational voltage Vmax (maximum of Vprog and Verase on the floating gate) vs. EOT. The optimization is across all considered high-κ materials. We find that La₂O₃ performs best for a strict read disturb criterion of 3.6 V while HfO₂ outperforms other high-κ materials for a 2.5 V read disturb. It yields the largest operational voltage of ~5–7 V constituting a ~30%-40% voltage reduction over the conventional SiO₂ based Flash cells.

Next, we experimentally corroborate these results using MOS capacitors. A J-Vgs comparison between pure SiO₂ tunnel dielectric and an asymmetric tunnel stack of SiO₂/HfSiON (~6.4 nm EOT), reveals a higher non-linearity for the composite stack (Fig. 8). A good agreement between simulations and experiments can also be observed in Fig. 8. Further, C-V measurements (Fig. 9) with minimal hysteresis confirmed a high quality composite dielectric stack. Finally, to ensure that tunneling was indeed Fowler-Nordheim (F-N) dominated, F-N slope was calculated (Fig. 10). An excellent linear fit was obtained from which both barrier height and tunneling mass were estimated.

In conclusion, we have developed a novel optimization methodology for deriving the minimum operational voltage of a composite tunnel dielectric stack for Flash memory. The methodology accounts for normal Flash operation constraints: retention, read and program disturb, and reveals that the higher J-Vgs, non-linearity of these stacks can result in up to 40% voltage reduction. The symmetric stack is found to be more efficient than asymmetric in reducing voltage. The optimum materials along with their thickness are also identified. Experiments confirm the non-linearity in J-Vgs using HfSiON, a high-κ which for the first time is used in the context of Flash memory.

Fig. 1. Schematic of a Flash memory showing the conventional tunnel oxide. (A) Shows replacement of the tunnel stack by an asymmetric barrier while (B) Shows that by a symmetric one. The thicker layer represents high-$\kappa$ material while thinner represents SiO$_2$.

Fig. 3a. $V_{prog}$ scaling with $T_{ox}$ for different asymmetric EOTs with HfO$_2$ (note the minima). Fig. 3b shows the 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 2.5 and 3.6 V read disturb voltage; the bottom one corresponds to -1.5V retention. Symbols for different EOTs are consistent across both figures.

Fig. 5. Change in $V_{prog}$ for asymmetric barriers with different constraints: retention only (no read disturb), and with different read disturb criteria $V_{read} = 2.5$ and 3.6 V. The values plotted represent global minima, across all high-$\kappa$ materials considered here.

Fig. 6. Erase simulations for a representative case of SiO$_x$/HfO$_y$/SiO$_2$, symmetric stack for different oxide thicknesses (within the same EOT) and with varying EOT.

Fig. 8. Experimental J-V curves for a control oxide (SiO$_2$) of 5.6 nm thickness and for an asymmetric SiO$_x$/HfSiON$_y$ stack of 6.4 nm EOT. The dotted lines represent experimental results while the solid lines are simulated curves for pure SiO$_2$ stacks of same EOT. Note that the asymmetric stack has a higher nonlinearity when compared to simulated pure SiO$_2$ of the same EOT. Also note that the simulations do not account for the breakdown in oxides. All samples are made on n-type substrates and for each sample three measurements were taken accounting for different areas.

Fig. 9. C-V measurements for the same stack as in Fig. 8. Note that the high-$\kappa$ /SiO$_2$ asymmetric stack shows no hysteresis implying a good interface with the Si substrate. Both samples were prepared under different conditions and have different doping levels. For each sample, C-V measurements were taken for 10 KHz, 100 KHz and 1 MHz frequencies.

Fig. 10. Plot of Fowler-Nordheim slope for SiO$_2$/HfSiON$_y$, asymmetric stack (the J-V of which has been shown in Fig.7). Barrier Height (phi) and effective mass (m*) have hence been derived.