Gate Dielectric Process Technology for the Sub-1 nm Equivalent Oxide Thickness (EOT) Era

by L. Colombo, J. J. Chambers, and H. Niimi

The semiconductor industry is now in its third generation of gate dielectrics. The first generation was the silicon dioxide (SiO_2) era from the early 1960s to about the mid-1990s. The benefits of SiO_2 noted in the earlier articles in this issue of *Interface* included utilization as: (a) passivation of surface dangling bonds and p-n junction interfaces, (b) pattern/diffusion masking ability, and (c) insulator supporting medium for aluminum interconnects between sections of the integrated circuit. In addition, the amorphous SiO_2 has a large energy gap (~9 eV) and a dielectric strength sufficient to support electric field strengths of several megavolts/cm, (10^6 V/cm); the latter is especially critical for the required metal-oxide-semiconductor field effect transistor (MOSFET) operation. These latter benefits have enabled the semiconductor industry to scale transistors down to about the 180 nm technology node* , corresponding to an SiO_2 thickness of ~3 nm. At about this oxide thickness, direct tunneling leakage currents rather than source-drain or substrate leakage currents, reached levels that were a significant portion of the allowable device leakage (~ 33%). In addition, at the sub 3 nm film thicknesses regime, boron dopant penetration from the boron-doped p^+ poly-silicon electrode through the SiO_2 into the channel region was a serious issue for pMOSFETs. An extensive literature has been published on the increasing SiO_2 gate leakage and boron penetration[1,2] with decreasing oxide thickness and their impact on continued device scaling as enunciated in Moore's law.

The continuance of scaling methodologies, however, has been achieved by the incorporation of nitrogen (~10 atomic %) in the SiO_2, whereby silicon oxynitride (SiON) thicknesses ranging from ~3.0 nm down to ~1.0 nm or so were achieved. The introduction of nitrogen in SiO_2 to form SiON increases the effective dielectric constant, κ. The higher effective dielectric constant leads to a lower effective gate dielectric thickness that is referred to as the equivalent oxide thickness (EOT) described in **general** by Eq. 1.

$$EOT = t_{SiO_2} + \sum_{i}^{n} \frac{3.9}{\kappa_i} (t_{SiON})_i \qquad (1)$$

where t_{SiO_2} is the physical thickness of the interface layer, potentially without nitrogen, t_{SiON} is the physical thickness of the SiON film, κ is the dielectric constant of the nitrogen containing SiO_2 film, and the summation over "*i*" is for layers that do not have a uniform nitrogen profile.

Today, the era of SiON films with an equivalent oxide thickness (EOT) in the range of ~1.0-1.2 nm have been in mass production for nearly a decade, meeting demanding transistor and reliability requirements. For SiON thickness less than ~1.0 nm, however, it was again found that direct tunneling leakage currents became excessive through the SiON. Scaling below ~1 nm for higher-performance devices and lower than ~1.5 nm for lower-power devices also became limited by poly-silicon depletion as well as gate dielectric leakage. Initial solutions will require either higher content nitrogen in SiON than currently in production or higher-κ dielectric constant gate materials (κ ~ 10-20).

Serious attention for alternate, higher dielectric constant materials began in about 1995 and have resulted in an extensive literature.[3-5] Hafnium-based gate dielectrics have emerged as the broad industrial choice for the high-κ gate dielectric material. The principal motivation for moving to high-κ gate dielectrics was the need to reduce the direct tunneling gate leakage currents. This was achieved by increasing the gate dielectric's physical thickness inasmuch as the direct tunneling leakage current is drastically reduced due to its exponential dependence on the physical gate dielectric thickness. Concurrently, the gate dielectric constant, κ, is increased. These two concurrent changes result in the approximate constancy of the MOSFET's capacitance, a significant parameter controlling the speed of the device. This is achieved by appropriately adjusting the ratio of the selected gate dielectric constant, κ, to the dielectric's physical thickness. As a result the MOSFET electrically behaves as though its gate dielectric thickness is smaller than its physical thickness (to reduce direct tunneling leakage current) by the ratio of the dielectric constant of SiON (generally slightly more than SiO_2's value of ~3.9 but significantly less than the dielectric constant of Si_3N_4, ~7.5) to the higher dielectric constant of the new gate dielectric (κ ≈ 15-20 for the HfO_2 and ~4-24 for Hf based HfSiON. This third regime of gate dielectric technology has also required the replacement of the doped poly-silicon gate electrodes by metallic gate electrodes to both reduce the poly-silicon depletion effect and to improve the work function match between the high-κ and gate electrode materials, itself a significant area of research.

The methods utilized for the fabrication of SiON for the second era will be discussed in the first section, the status of HfSiON gate dielectric in presented in the second section, and the appropriate metal gate issues will briefly be summarized in the third section.

Silicon Oxynitride

SiON dielectrics were first introduced at an EOT of about 3 nm and at that time required less than about 10 atomic percent nitrogen to minimize boron penetration and gate leakage. Controlled incorporation of nitrogen, especially concentration and depth profile, was critical to the introduction of SiON. The key process that enabled the industry to introduce the first reliable SiON gate dielectric was plasma nitridation of SiO_2.[6]

The advantage of plasma nitridation of SiO_2 is its ability to (a.) control the dielectric layer thickness, (b.) precisely control the content and/or location of the nitrogen, and (c.) improve reliability. Plasma nitrided SiO_2 can be created by (i.) top surface nitridation, (ii.) homogeneous SiON formation, where the nitrogen is distributed uniformly through the oxide, and (iii.) incorporation of nitrogen at the Si-SiO_2 interface. Simulations have shown that top surface nitrogen profiles have better boron blocking characteristics than other profiles. The bulk nitrogen concentration in SiON has to be critically controlled, but even more critical is the concentration of N atoms at the Si-dielectric interface. The concentration of nitrogen at the interface has to follow the bonding constraint theory[7] in order to minimize defects. According to the bonding constraint theory, whose

* The technology node refers to the metal line-to-line spacing—½ pitch—for a particular DRAM generation. The physical channel length for a logic chip consonant with the DRAM generation is about 45% of the technology node as discussed in the *International Technology Roadmap for Semiconductors*, ITRS.

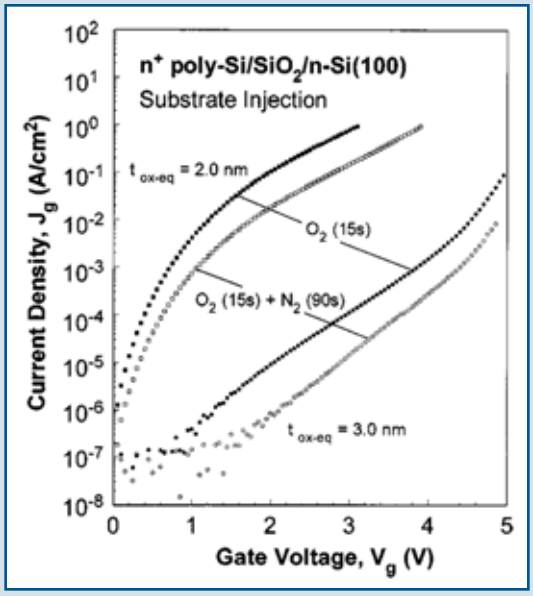

Fig. 1. *Gate leakage current density as a function of gate voltage for 2.0 and 3.0 nm EOT thick gate tunneling currents compared to SiO_2 (filled symbols).*[9]

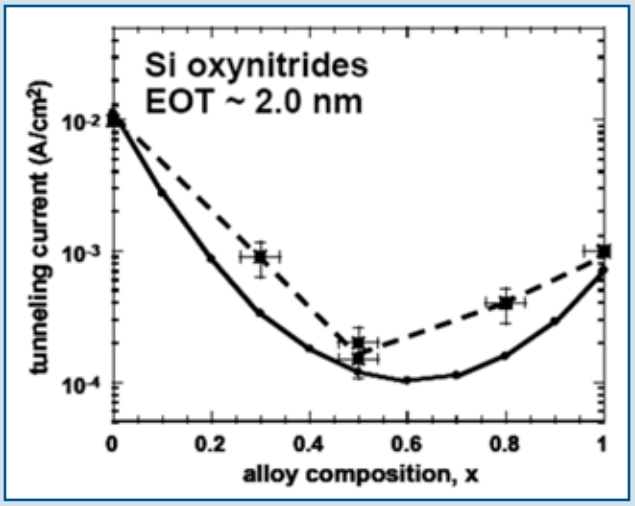

Fig. 2. *Effect of nitrogen on the tunneling current of $(SiO_2)(1-x)\bullet(Si_3N_4)(x)$ showing that the gate leakage can be reduced by about a factor of 100 by optimizing the nitrogen concentration (continuous curve is theory).*[11]

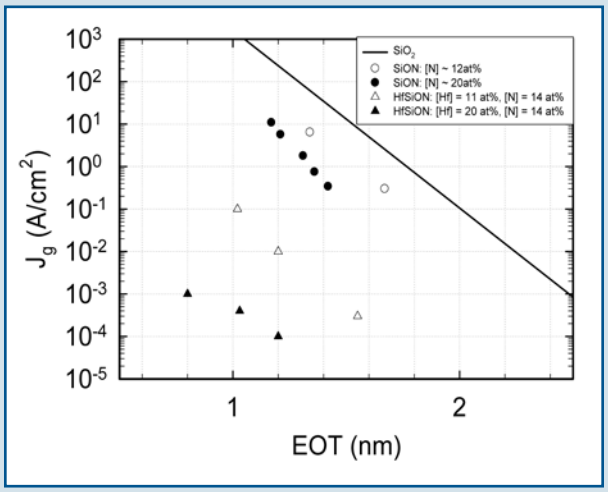

Fig. 3. *Gate leakage of SiON and HfSiON as a function of EOT and composition.*

criteria were originally established for the formation of low defect density glasses, the maximum average number of bonds (N_{av}) of ~3 is the limit for a low defect density dielectric. If a thin SiO_2 layer (~0.5nm, about one molecular layer) is formed between the Si substrate and Si_3N_4, the average bonding coordination is reduced to 3.0, and is essentially the same as for the Si-SiO_2 interfaces leading to significantly improved electrical performance. Monolayer-level nitrogen incorporation (~7 x 10^{14} cm^{-2}) to form Si_3-N bonding configurations in the vicinity of the Si-SiO_2 interface[8] reduces gate leakage by approximately tenfold compared to pure SiO_2 as shown in Fig. 1.[9] The incorporation of nitrogen near the Si-SiO_2 interface leads to reduced interfacial sub-oxide bonding which defines the transition region between the Si substrate and bulk SiO_2. The modified interface structure results in the reduction of gate tunneling currents by increasing the tunneling barrier height. Although monolayer level incorporation of nitrogen at the interface improves leakage characteristics, further scaling will require higher bulk nitrogen levels.

Increasing the nitrogen concentration in the bulk of the film by either direct nitridation or nitridation of the SiO_2 base oxide reduces gate leakage. In the case of $(SiO_2)_{1-x}(Si_3N_4)_x$ solution, the gate leakage can be reduced by as much as a factor of about 100.[10] Figure 2 shows leakage current as a function of silicon nitride fraction in SiON.[11] The leakage reduction is clearly demonstrated for the solid solution in the vicinity of the 50-50% composition point. Figure 3 further shows that the leakage current of SiON can indeed be decreased by increasing the nitrogen concentration in SiON to ~20 at.% as an example. Nitrogen concentrations in excess of ~20 at.%, however, results in large shifts in the threshold voltage which can be observed in both nMOSFET and pMOSFET devices with the pMOSFET shift being larger. In addition, the leakage reduction of SiON with high nitrogen is not sufficient for scaling dielectrics beyond the 45 nm technology node and high-κ is necessary to meet the scaling requirements.[12]

High-κ Dielectric Constant Materials

The semiconductor community has been searching for high-κ gate dielectrics since about 1995. It appears that, after investigating a multitude of materials, Hf-based gate dielectrics have emerged as the "final" choice for high-κ material. The principal motivation for moving to high-κ gate dielectrics was the need to reduce direct tunneling leakage currents as discussed earlier. Figure 3 shows a comparison of the gate leakage of SiON and hafnium silicon oxynitride (HfSiON) as a function of composition indicating that HfSiON is scalable beyond 1 nm EOT with much lower gate leakage than currently achieved with SiON. Lower leakage is clearly a benefit of the physically thicker high-κ gate dielectrics, but only to first order because many other properties must be satisfied before adopting any specific high-κ material.

One of the key properties of SiO_2 and SiON is their amorphous state even after annealing at temperatures well above the typical complementary metal-oxide-semiconductor (CMOS) processing temperatures (~1050°C). It is generally agreed that amorphous gate dielectrics are preferred over crystalline materials. For example, hafnium oxide (HfO_2) and hafnium silicon oxide (HfSiO) were key gate dielectrics studied in the early stages of high-κ development, but both have relatively low crystallization temperatures. Figure 4a shows a cross-sectional transmission electron micrograph (TEM) of HfO_2 that clearly shows ordering, and indication of crystallization in the film after poly silicon deposition at about 700°C. HfSiO also crystallizes at temperatures lower than the maximum CMOS processing temperatures and, in addition, also suffers from phase separation that results in crystalline HfO_2 phases in a matrix of HfSiO with a lower Hf concentration than the original as deposited uniform HfSiO.[13] These results are less than desirable because polycrystalline films contain point defects and grain boundaries that place significant limitations on the device reliability. Therefore, more thermally stable dielectrics were preferable.

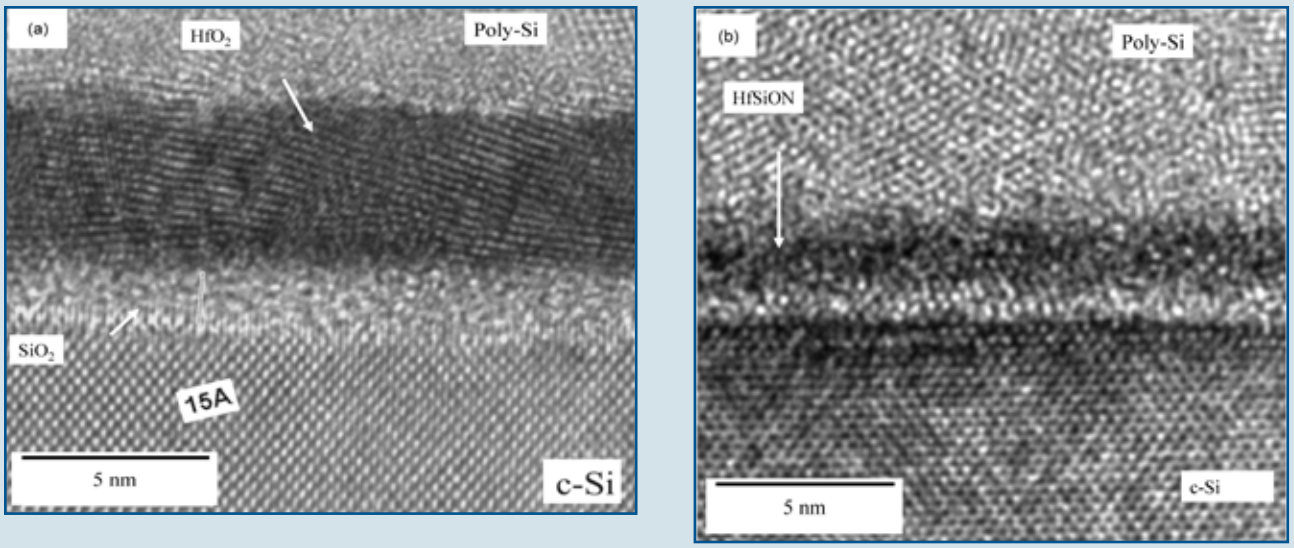

Fig. 4. *TEM micrographs of (a) Si/SiO_2/HfO_2/Poly Si where the maximum temperature was 700°C and (b) Si/SiO_2/HfSiON/Poly Si after 1050°C annealing.*

A few research groups have reported that HfSiON is structurally stable, has a high dielectric constant, has low charge trapping, and high electron and hole mobility, typically ~90% and > 90% of SiO_2 universal mobility curve at an electric field of 1 MV/cm, for HfSiON having Hf and N concentrations of ~11 at.% and 14 at.%, respectively. The addition of ~14 at.% nitrogen to HfSiO (Hf ~11 at%) to form HfSiON was found to stabilize the structure of the films such that the material remains amorphous up to 1100°C for low Hf concentrations.[14] Figure 4b shows a TEM micrograph of poly-Si/HfSiON/Si after annealing at 1050°C where no lattice fringes are observed due to the amorphous nature of the films. X-ray diffraction and Fourier Transform Infrared Spectroscopy data also support the amorphous nature of HfSiON.[14,15] Thus, in general, the idea of creating HfSiON is that hafnium is added to silicon dioxide to increase the dielectric constant and nitrogen is added to stabilize the material structure. An additional challenge observed with HfO_2 deposited by atomic layer deposition was large C-V hysterisis for n-channel devices as a result of a high density of trapped charge in the bulk of the dielectric film.[16] HfSiON on the other hand has shown significantly lower trapped charge than HfO_2 as shown in Figure 5.[17] Subsequently, improvements in charge trapping were observed for very thin hafnium oxynitride (HfON) on thin SiO_2, (EOT ≤ 1.1 nm) with the improvement being attributed to a lower number of total defects in HfON and the formation of a "silicate" at the interface.[18] The improvement in charge trapping for the thinner HfON film was also followed by electron mobility improvements (approaching ~ 90% of the SiO_2 electron mobility at an electric field of ~ 1 MV/cm). The presence of a thin SiO_2 or SiON interface and the addition of Si to HfON have in general been found to improve mobility as shown in Figure 6 and decrease charge trapping.[18] The industry recognized the many benefits of HfSiON as a high-κ gate dielectric and devoted considerable resources in order to introduce poly-Si/HfSiON gate stack into the conventional flow. However, pMOSFET with poly-Si gate electrode devices suffered from a high threshold voltage (V_T), a phenomenon that became known as pMOSFET V_T offset; threshold voltages offset as high as 0.5V have been observed

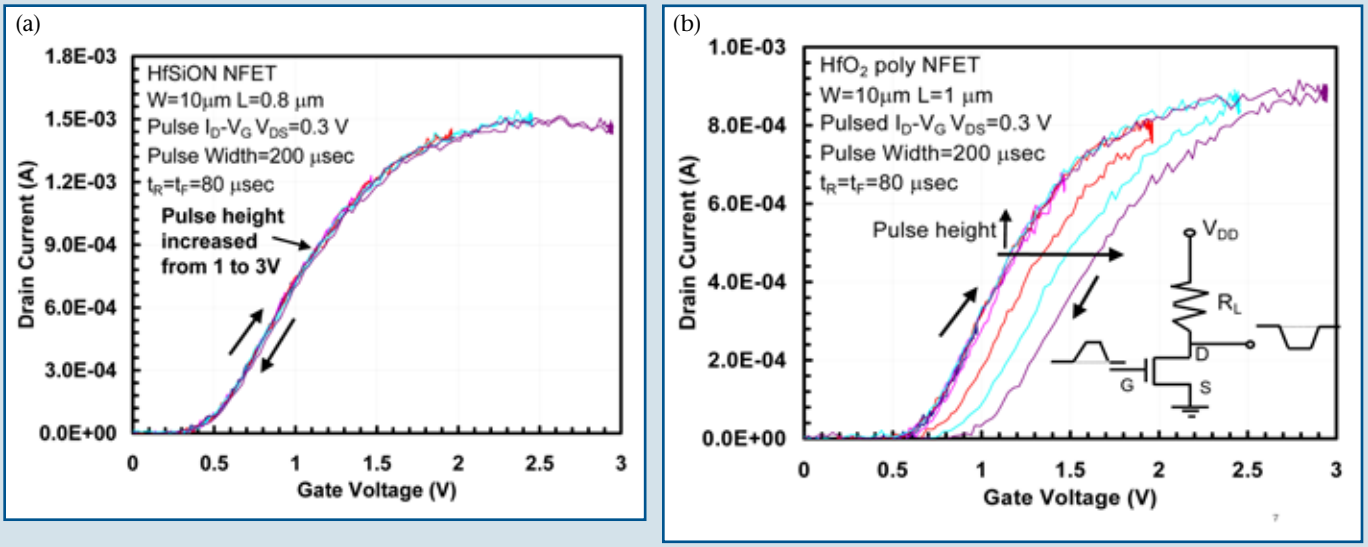

Fig. 5. *Pulsed I_D-V_G measurements on nMOSFETs. Measurement involves applying a pulse of voltage at the gate terminal and calculating drain current by measuring voltage across the drain terminal. (a) HfO_2 and (b) HfSiON gate dielectrics.*[17]

(continued from previous page)

for pMOSFET devices. The pMOSFET V_T offset quickly became a showstopper for producing high performance poly-Si/HfSiON devices, which required lower V_T.

Metal Gate Electrodes

As a result of the pMOSFET V_T offset found in poly-Si/high-k gate stacks, the industry initiated a significant effort to the selection of nMOSFET and pMOSFET "metal" gate electrodes in order to obtain low threshold voltages (~ 0.25 eV) and to minimize the depletion effect found in poly-Si electrodes. Researchers have investigated many metal gate electrode candidates, but the industry has not converged on one set of materials at the present time. Three basic methodologies have been pursued to set the work function: (i.) selection of a metal with appropriate vacuum work function, (ii.) modification of the metal/dielectric interface, and (iii.) modification of the gate dielectric using non-band edge metals.

The discrepancy between the vacuum work function and that observed in transistor devices is a result of poor control of the material composition at the interface between the gate dielectric and the metal gate electrode interface. The placement of low work function metals or high work function metals at the metal/dielectric interface have been found to change the effective work function of the MOSFET and thus V_T of the device but this is also dependent on the thermal budget.[19-21] The reactions between the dielectric and metals and the charges in the dielectrics have made setting the work function at the band edges difficult. The most effective method to date of setting the nMOSFET work function has been modification of the gate dielectric by the introduction of lanthanide metals, with a non-band edge metal electrode.[22] In this case the lanthanide metal or metal oxide reacts with the gate dielectric thus changing the charge state of the dielectric. Work functions of 4.05 to 4.1 eV have been reported with the expected device threshold voltage of about 0.25 eV for nMOSFETs. Incorporation of Al in the gate dielectric has also been used to shift the work function of pMOSFET devices toward the band edge but mobility degradation has been observed as result of Al diffusion into the substrate.[23] The data to date are clear in one respect; that is, metal gates reduce gate depletion, but it is challenging to obtain the correct work functions that result in low device V_T. Setting the pMOSFET work function near the band edge has proven to be more difficult, as compared to nMOSFET, at technologically relevant EOTs of ~ 1 nm. As the EOT decreases toward 1 nm, the work function decreases to unacceptable values; this effect is referred to as pMOSFET V_T roll-off, which was found to increase with increasing process temperature and/or increasing metal (Hf) concentration in the dielectric.

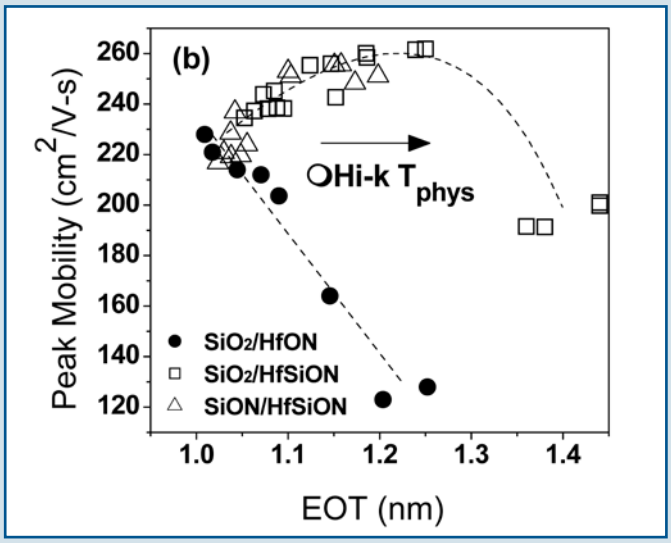

Fig. 6. *HfON and HfSiON electron mobility as a function of EOT for Si/SiO_2/HfON, Si/SiO_2/HfSiON, and Si/SiON/HfSiON.*[18]

The data on the effect of process temperature on work function have been clear for a few years now, although the advantage of the conventional transistor flow has kept focus on gate first approaches. However, given the need to reduce gate dielectric leakage, reduce metal gate depletion and the nature of the metal/dielectric interface chemistry, the benefits of a lower thermal budget process needs to be implemented in order to integrate the high-κ/metal gate stack into the transistor flow.

One such approach is referred to as the replacement gate process flow, which enables decoupling the junction formation from the dielectric/metal gate formation process.[24] In this flow, a sacrificial gate electrode (e.g. poly Si) and a sacrificial gate dielectric, e.g., SiO_2, **may be** used and then removed after all the junctions are formed. Subsequently, the gate dielectric and/or the metal were deposited to form the gate electrode stack. A related process that has been used to form the metal gate electrode on high-κ is the NiSi Full Silicidation process (FUSI).[25] The FUSI process has shown some very encouraging results by using lanthanide metals for nMOSFET and noble metals for pMOSFET at the gate dielectric/FUSI interface. Because of its low thermal budget, the replacement process flow may be the most likely integration scheme that will enable the introduction of metal gates, in conjunction with the high-κ gate dielectric, that can provide devices with the desired threshold voltages.[26] Indeed, a plethora of modifications to the gate first or replacement gate methodologies have been and continue to be discussed in the scientific literature.

Summary

As we approach the limits of SiON, HfSiON high-κ gate dielectrics with metal gate electrodes are being incorporated into some devices at the 45 nm technology node. Hafnium-based high-κ dielectrics will usher in the third generation of CMOS gate dielectrics and will enable the sub-1 nm EOT era. Metal gates will replace doped poly silicon electrodes to scale the inversion dielectric thickness to levels not achievable by polysilicon. The sub-1 nm EOT era would not have been possible without the extensive knowledge gained perfecting the previous generations of SiO_2 and SiON dielectrics.

About the Authors

Luigi Colombo is a TI Fellow in the Silicon Technology Development Group at Texas Instruments in Dallas, TX. His areas of expertise and interest include deposition and/or growth of infrared materials such as II-VI compounds, multi-component oxides for capacitor gate dielectric applications, ferromagnetic semiconductors such as chalcogenide spinels, and nitrided SiO_2. He is the author and co-author of over 125 publications in journals and conference proceedings, 3 book chapters, and holds over 50 U.S. patents. He may be reached at colombo@ti.com.

Jim Chambers is a Member of the Technical Staff at Texas Instruments in Dallas, TX. His current duties include integrating high-k and metal gate electrode for the 45 nm node. Jim is the author of over 30 publications and one book chapter, and holds over 25 U.S. patents in the area of SiON, high-k gate dielectrics, and metal gate electrodes. He is a member of the technical committee for the Semiconductor Interface Specialists Conference. He may be reached at chambers@ti.com.

Hiroaki "Hiro" Niimi is a Group Member of Technical Staff at Texas Instruments. He is the author or co-author of over 80 refereed publication in journals and conference proceedings in the field of silicon dioxide based gate dielectrics, one book chapter, and over 10 U.S. patents in area of plasma nitridation for gate dielectrics. He may be reached at h-niimi@ti.com.

References

1. P. A. Packan, Pushing the Limits, *Science*, **285**, 2079 (1999).
2. D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, *Nature*, **399**, 759 (1999).
3. *High Dielectric Constant Materials: VLSI MOSFET Applications (Springer Series in Advanced Microelectronics)*, H. R. Huff and D. C. Gilmer, Editors, Berlin-Heidelberg-New York (2005).
4. *Materials Fundamentals of Gate Dielectrics*, A. A. Demkov and A. Navrotsky, Editors, Springer Dordretch, Berlin, Heidelberg, New York (2005).
5. *High-κ Gate Dielectrics*, M. Houssa, Editor, Institute of Physics Publishing, Bristol, UK (2004).
6. S. V. Hattangady, H. Niimi, and G. Lucovsky, *Appl. Phys. Lett.*, **66**, 3495 (1995).
7. G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, *Appl. Phys. Lett.*, **74**, 2005 (1999).
8. H. Niimi and G. Lucovsky, *J. Vac. Sci. Tech. A*, **17**, 3185 (1999).
9. H. Niimi and G. Lucovsky, *J. Vac. Sci. Tech. B: Microelectronics and Nanometer Structures*, **17**, 2610 (1999).
10. H. Yang and G. Lucovsky, in *Technical Digest of IEEE International Electron Device Meeting*, The Institute of Electrical and Electronics Engineers, p. 245, Washington, DC (1999).
11. G. Lucovsky and J. L. Whitten, in *High Dielectric Constant Materials: VLSI MOSFET Applications*, H. R. Huff and D. C. Gilmer, Editors, **11**, p. 315 Springer-Verlag, Berlin Heidelberg (2005).
12. International Technology Roadmap of Semiconductors, in *http://www.itrs.net/Links/2007Summer/2007_Public_SF.htm (2007)*.
13. S. Ramanathan, P. C. McIntyre, J. Luning, P. S. Lysaght, Y. Yang, Z. Chen, and S. Stemmer, *J. Electrochem. Soc.*, **150**, F173 (2003).
14. M. R. Visokay, J. J. Chambers, A. L. P. Rotondaro, A. Shanware, and L. Colombo, *Appl. Phys. Lett.*, **80**, 3183 (2002).
15. M. A. Quevedo-Lopez, J. J. Chambers, M. R. Visokay, A. Shanware, and L. Colombo, *Appl. Phys. Lett.*, **87**, 012902 (2005).
16. A. Kerber, E. Cartier, L. Pantisano, M. Rosmeulen, R. Degraeve, T. Kauerauf, G. Groeseneken, H. E. Maes, and U. Schwalke, in *Reliability Physics Symposium Proceedings, 2003, 41st Annual. 2003 IEEE International*, p. 41 (2003).
17. A. Shanware, M. R. Visokay, J. J. Chambers, A. L. P. Rotondaro, J. McPherson, and L. Colombo, in *Electron Devices Meeting, 2003, IEDM '03 Technical Digest, IEEE International*, p. 38.6.1 (2003).
18. P. D. Kirsch, M. A. Quevedo-Lopez, S. A. Krishnan, B. H. Lee, G. Pant, M. J. Kim, R. M. Wallace, and B. E. Gnade, *Appl. Phys. Lett.*, **89**, 242909 (2006).
19. S. Park, L. Colombo, Y. Nishi, and K. Cho, *Appl. Phys. Lett.*, **86**, 073118-3 (2005).
20. C.-H. Lu, G. M. T. Wong, M. D. Deal, W. Tsai, P. Majhi, C. O. Chui, M. R. Visokay, J. J. Chambers, L. Colombo, B. M. Clemens, and Y. Nishi, *IEEE Electron Device Lett.*, **26**, 445 (2005).
21. H. Y. Yu, C. Ren, Y.-C. Yeo, J. F. Kang, X. P. Wang, H. H. H. Ma, M.-F. Li, D. S. H. Chan, and D.-L. Kwong, *IEEE Electron Device Lett.*, **25**, 337 (2004).
22. H. Alshareef, H. Harris, H. Wen, C. Park, C. Huffman, K. Choi, H. Luan, P. Majhi, B. Lee, R. Jammy, D. Lichtenwalner, J. Jur, and A. Kingon, in *VLSI Technology, 2006, Digest of Technical Papers, 2006 Symposium*, p. 7 (2006).
23. H. N. Alshareef, H. F. Luan, K. Choi, H. R. Harris, H. C. Wen, M. A. Quevedo-Lopez, P. Majhi, and B. H. Lee, *Appl. Phys. Lett.*, **88**, 112114 (2006).
24. A. Chatterjee, R. A. Chapman, G. Dixit, J. Kuehne, S. Hattangady, H. Yang, G. A. Brown, R. Aggarwal, U. Erdogan, Q. He, M. Hanratty, D. Rogers, S. Murtaza, S. J. Fang, R. Kraft, A. L. P. Rotondaro, J. C. Hu, M. Terry, W.Lee, C. Fernando, A. Konecni, G. Wells, D. Frystak, C. Bowen, M. Rodder, and I.-C. Chen, in *Technical Digest of IEEE International Electron Device Meeting*, p. 821 (1997).
25. W. P. Maszara, Z. Krivokapic, P. King, J.-S. Goo, and M.-R. Lin, in *Electron Devices Meeting*, 2002, IEDM '02, Digest. International, p. 367 (2002).
26. S. Yamaguchi, K. Tai, T. Hirano, T. Ando, S. Hiyama, J. Wang, Y. Hagimoto, Y. Nagahama, T. Kato, K. Nagano, M. Yamanaka, S. Terauchi, S. Kanda, R. Yamamoto, Y. Tateshita, Y. Tagawa, H. Iwamoto, M. Saito, N. Nagashima, and S. Kadomura, in *VLSI Technology, 2006, Digest of Technical Papers, 2006 Symposium*, p. 152 (2006).