Intrinsic limitations on device performance and reliability from bond-constraint induced transition regions at interfaces of stacked dielectrics

G. Lucovsky,[a] H. Yang, H. Niimi, J. W. Keister, and J. E. Rowe
Departments of Physics, and Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695-8202

M. F. Thorpe
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

J. C. Phillips
Lucent Bell Laboratories, Murray Hill, New Jersey 07974

(Received 25 February 2000; accepted 28 February 2000)

The substitution of deposited alternative gate dielectrics for thermally grown SiO_2 in aggressively scaled complementary metal–oxide–semiconductor devices requires separate and independent processing steps for (i) the oxidation of the Si substrate to form the Si-dielectric interface and (ii) the deposition of thin film dielectric. Ultrathin plasma-oxidized Si–SiO_2 interface layers which contribute approximately 0.3–0.4 nm to the overall electrical oxide thickness have been integrated into devices with Si nitride, Si oxynitride, and Ta_2O_5 alternative dielectrics. This article proposes an analogy between (i) microscopically inhomogeneous bulk glass alloys such as $GeSe_x$ with $1<x<2$, and (ii) interfaces included in these composite gate dielectric-semiconductor structures including, for examples, the Si–SiO_2 and internal dielectric SiO_2–Si_3N_4 interfaces. Scaling relationships for bond defect states applied initially to microscopically inhomogeneous glasses and thin films are applied here to interfaces in stacked gate dielectrics. © *2000 American Vacuum Society.* [S0734-211X(00)10403-2]

I. INTRODUCTION

The projected scaling of complementary metal–oxide–semiconductor (CMOS) devices into the thickness regime <1.5 nm, in which direct tunneling through thermally grown SiO_2 becomes much greater than 1 $A\,cm^{-2}$, requires introduction of alternative deposited gate dielectrics. An obvious approach is to substitute noncrystalline insulators with dielectric constants, k, greater than SiO_2 ($k_{ox}\sim 3.8$), to obtain physically thicker films with the same capacitance as thinner SiO_2. The increased physical thickness for a direct substitution is given by the ratio of the dielectrics, k/k_{ox}. Since tunneling current scales exponentially with thickness, alternative gate dielectric materials with $k>(2-3)k_{ox}$ are anticipated to produce significant reductions in tunneling current. However, the tunneling transmission probability also depends on (i) the height of the tunneling barrier, E_b, e.g., the conduction band offset energy between the Si substrate and the alternative dielectric, and (ii) the effective mass of tunneling electrons, m_e^*, in the alternative insulator.[1] Since band gaps and dielectric constants generally scale inversely with each other, increased physical thickness alone may not be sufficient to insure the many order of magnitude decreases in tunneling current required for the performance and reliability of advanced CMOS devices with electrical oxide thickness (EOT) ~1 nm or less. Direct deposition of high-k oxides and silicates is frequently accompanied by substrate interactions which decrease the gate dielectric capacitance and/or introduce electrically active defects. One approach to this problem is to use composite dielectrics with an interfacial layer that suppresses substrate interactions and maintains interfacial electrical properties as, for example, nitrided SiO_2. The challenge is even greater for these composite dielectrics since they have an additional band offset step at the internal dielectric interface.[1] Because of decreased band gaps in the high-k dielectrics, the internal potential steps are generally *down hill* going from an SiO_2 interfacial layer to the high-k dielectric. Model calculations have shown this type of energy band profile introduces inherent symmetries between substrate and gate electron injection.[1]

This article focuses on a different aspect of stacked gate dielectrics, mechanical bonding strain at the Si-dielectric and internal dielectric interfaces. This is addressed by combining bond counting on the atomic level with theoretical considerations for network constraints that have previously been applied to bulk glasses and thin films.[2,3] Recent studies have shown that the prediction of a connectivity induced *rigidity* transition in network glasses made over two decades ago,[2] has been realized experimentally in chalcogenide glass alloys.[4] This transition occurs in two steps in the Ge–Se alloys, specifically as second-order and first-order transitions at alloy compositions of approximately 20 and 23 at. % Ge, close to the predicted first-order transition at 20 at. % Ge.[2] Recent refinements in the rigidity theory have predicted two phases transitions in the Ge–Se alloy system at compositions corresponding to an average number of bonds/atom, N_{av}, of 2.375 and 2.389, respectively.[5] These are very close to the

[a]Electronic mail: gerry_lucovsky@ncsu.edu

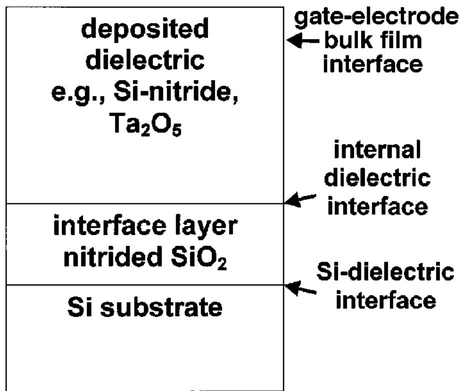

FIG. 1. Schematic representation of a stacked gate dielectric, indicating the Si-dielectric interface and the internal dielectric interface as well.

experimentally reported values of 2.4 and 2.46. For $N_{av} < 2.4$, the Ge–Se alloys are floppy in the sense that the average number of constraints per atom, C_{av}, is less than 3, so that there are zero frequency, or so-called floppy modes.[2,3,5] This concept of a ''rigid'' to ''floppy'' transition as a function of alloy composition in a mean-field theory has recently been extended to a spatially rigid to floppy transition that can occur at interfaces between crystalline semiconductors such as Si, and gate dielectrics such as SiO_2 and Si_3N_4.[6,7] It will be shown below that this extension of mechanical constraint theory to these interfaces helps to explain (i) the many orders of magnitude differences in electrically active defects at Si–SiO_2 and Si–Si_3N_4 interfaces,[6,8] and (ii) the existence of transition regions with suboxide bonding between crystalline Si and SiO_2.[7,9]

II. EXPERIMENTAL RESULTS

Figure 1 is a schematic representation of a gate dielectric structure that includes (i) a Si substrate, (ii) a dielectric comprised of (a) an ultrathin interfacial layer, e.g., nitrided SiO_2 with a physical thickness of 0.5–0.6 nm;[10,11] and (b) a thicker bulk dielectric, e.g., Si_3N_4 or an alternative high-*k* oxide such as Ta_2O_5; and (iii) a gate electrode, e.g., polycrystalline Si, or a simple or compound metal. The gate stack in Fig. 1 includes three interfaces that contribute to electrical performance and reliability. These are (i) the Si-dielectric interface, (ii) the internal interface between the two dielectrics, and (iii) the interface between the bulk dielectric and the gate electrode. This article will not address the gate electrode-dielectric interface; however, it is important to note that issues relative to chemical bonding, electronic energy, and defects are equally as challenging and important as at the other two interfaces addressed below.

A. Transition regions with suboxide bonding at Si–SiO_2 interfaces

X-ray photoemission spectroscopy (XPS) studies have demonstrated that in addition to the Si substrate, SiO_2 and ideal interface bonding core shifts in Si $2p$ states that are expected at an ideal chemically abrupt interface, there are additional XPS features indicative of an interfacial transition region with excess suboxide bonding arrangements.[12] Two recent studies by Keister and co-workers[13,14] have focused on three aspects of the suboxide bonding in these transition regions for interfaces prepared by remote plasma processing.[10,11] These are (i) changes in the excess suboxide bonding in the spectral regime between the Si substrate and the SiO_2 as a function of annealing, (ii) increases in integrated excess suboxide bonding at nitrided interfaces, and (iii) the spectral position and width of the interfacial nitrogen feature as function of nitrogen concentration. Figures 2 and 3 summarize some of these results. There is about a 25% reduction in the areal density of Si atoms in suboxide bonding arrangements for interfaces formed on H-terminated Si(111) and Si(100) surfaces by remote plasma-assisted oxidation at 300 °C, and then subjected to annealing in inert ambients at temperatures to 900 °C.[13] These additional spectral features fall into three spectral bands that have been designated as Si^{1+}, Si^{2+}, and Si^{2+}, based on an assumption that they represent bonding arrangements with one, two, and three oxygen atom neighbors, respectively.[12] Consistent with this notation, the substrate Si feature is Si^0, and the SiO_2 feature is Si^{4+}. After a 900 °C anneal, there is a little less than one monolayer of Si atoms in suboxide bonding arrangements in excess of what is required at an ideal abrupt interface. As shown in Fig. 2, the largest decrease in the XPS spectra for a Si(111)–SiO_2 interface is in the Si^{2+} oxidation state corresponding to a Si atom with two oxygen neighbors. This bonding configuration is not *native* to a Si(111) surface unless there are deviations in planarity such as surface steps. If the excess areal density of Si atoms is converted into an equivalent thickness corresponding to an SiO interfacial composition, then the physical thickness of this transition region is approximately 0.27 nm, see Ref. 15 for details.

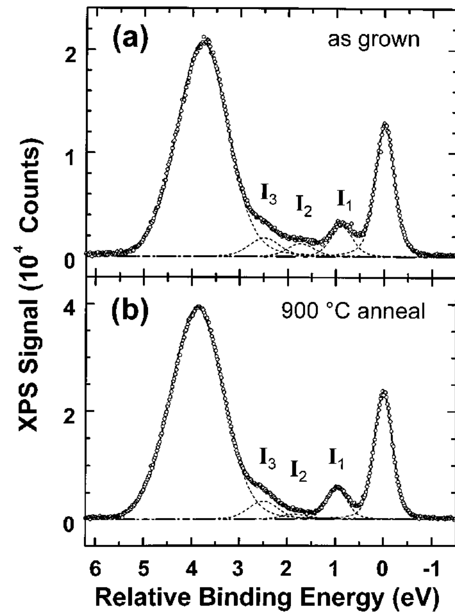

FIG. 2. Synchrotron XPS data in the Si $2p$ region for RPAO interfaces on Si(111) (a) as-grown at 300 °C, and (b) after being subjected to a 900 °C RTA in an inert ambient (e.g., He) (see Ref. 13).

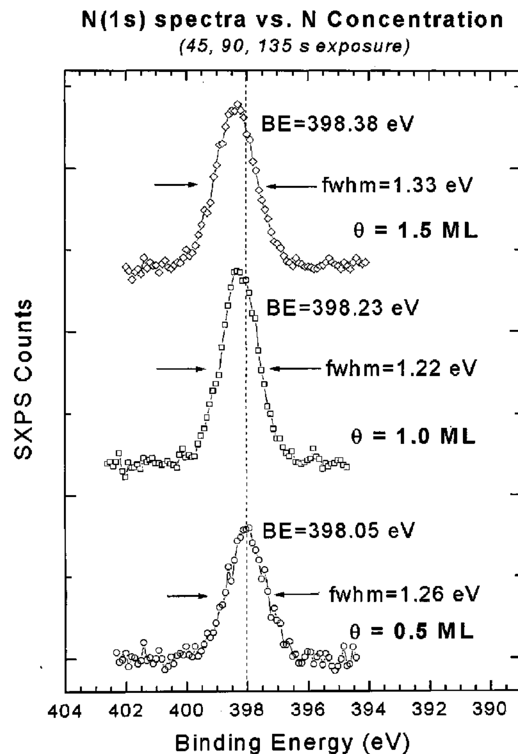

FIG. 3. Soft x-ray photoelectron spectroscopy (SXPS) data in the N 1*s* region for RPAO interfaces with three different interface nitridation processing times, giving nitrogen atom concentrations of approximately 0.5, 1.0, and 1.5 monolayers. These devices have been subjected to a 900 °C RTA in an inert ambient (e.g., He) (see Ref. 14).

Controlled incorporation of nitrogen atoms in a concentration range up to about 2 monolayers has recently been accomplished by a 300 °C remote plasma-assisted, post-oxidation, interface nitridation process.[10,11] Quantification of interfacial nitrogen was done by secondary ion mass spectrometry (SIMS) and nuclear reaction analysis (NRA).[11] For a nitrogen plasma processing time of 90 s, a concentration of $7.5\pm1\times10^{14}$ nitrogen atoms cm^{-2}, or equivalently one nitrogen atom/per interfacial silicon atom on Si(100), was obtained. Additionally, the nitrogen concentration was found by analysis of SIMS data to be approximately linear in processing time up to about a 1.5 monolayer coverage. Analysis of XPS data on Si(100) surfaces with and without interface nitridation after a 900 °C anneal, indicated an increased excess suboxide/subnitride bonding, resulting in a physically thicker interfacial transition region, 0.35 nm as compared to 0.27 nm.[15] The spectra shown in Fig. 3 indicate changes in the character of the N 1*s* feature as a function of increasing interfacial nitrogen. The greatest changes in core level energy and half width occur for a concentration increase from approximately 1–1.5 monolayers. The position of the interfacial nitrogen feature at 398 eV is consistent with nitrogen atoms being bonded directly at the interface to silicon atoms that have silicon as well as oxygen atom second neighbors. This N 1*s* core level position is different from the bonding of nitrogen atoms in interfacial oxynitride alloys prepared by thermal interface nitridation processes and also studied by XPS.[14]

Complementary techniques for probing interfacial bonding and structure have also yielded evidence for transition regions with qualitatively similar suboxide bonding arrangements, and with approximately the same *effective* thickness of 0.3 nm. These include Auger electron spectroscopy (AES),[16] infrared absorption spectroscopy (IRAS),[17] and high-resolution electron energy-loss spectroscopy (HREELS).[18]

B. Internal interfaces in stacked gate dielectrics

Previous experiments have shown that direct deposition of SiO_2 (Refs. 19–21) or Si_3N_4 (Refs. 6 and 8) onto H-terminated Si surfaces yielded interfaces that were inferior to those produced during conventional and rapid thermal oxidation of Si. Significant improvements in device performance and reliability were obtained by performing a remote plasma-assisted oxidation (RPAO) step prior to the deposition of (i) SiO_2 films by remote plasma-enhanced chemical vapor deposition RPECVD[10,11,21] or rapid thermal CVD (RTCVD)[22] and (ii) Si_3N_4 films by RPECVD.[6,8] In a previously reported study, fixed positive charge was found at the two internal dielectric interfaces of stacked oxide–nitride–oxide gate dielectrics.[23] These internal dielectric interfaces were formed by depositing successive layers of SiO_2, Si_3N_4, and SiO_2 by RPECVD at 300 °C,[23] and the Si–SiO_2 interfaces were formed by RPAO. Concentrations of fixed positive charge were extracted from capacitance–voltage $(C-V)$ data were about the same at each of the two interfaces, in the low to mid $10^{12}\,cm^{-2}$ in as-deposited films, decreasing to the mid-to-low $10^{11}\,cm^{-2}$ after a 900 °C anneal in an inert ambient. The physical thicknesses of the films in these stacked structures were 5 nm for the top and bottom SiO_2 films, and from 2 to 20 nm for the middle Si_3N_4.

Figures 4(a) and 4(b) present $C-V$ traces for NMOS and PMOS capacitors (CAPs) with stacked dielectrics containing thin RPAO oxide layers, ~0.6 nm thick, and thicker RPECVD layers of either (i) SiO_2, (ii) an optimized Si oxynitride alloy, $(SiO_2)_{0.5}(Si_3N_4)_{0.5}$, and (iii) Si_3N_4.[24,25] The thickness of each RPECVD layer was adjusted to give essentially the same values for EOT, ~2 nm. Flat band voltages were obtained from a standard analysis of the $C-V$ curves, and are at the inflection points of each trace. Rigid shifts of the $C-V$ traces for the NMOS and PMOS CAPs with oxynitride and nitride dielectrics with respect to the MOS CAPs with the deposited Si oxide dielectrics are essentially the same as the differences in respective flat band voltages. Additionally, the flat band voltages for the NMOS and PMOS CAPs with SiO_2 dielectrics are at voltages that are consistent with substrate doping, $2-5\times10^{17}\,cm^{-3}$, and the respective degenerate polycrystalline gate electrodes, indicating that fixed charge at the Si–SiO_2, and the internal RPAO SiO_2 RPECVD SiO_2 interfaces are $<5\times10^{10}\,cm^{-2}$. The average flat band voltage (or rigid $C-V$) shifts, ΔV_{fb} (or ΔV_{rs}) for the Si oxynitride and Si nitride devices, are ~0.04 and 0.15 V, respectively, with an uncertainty of <0.01 V. Since $CQ_f=\Delta V_{fb}\,(\Delta V_{rs})$, where Q_f is the fixed charge and C is an average capacitance in the transition region of the $C-V$

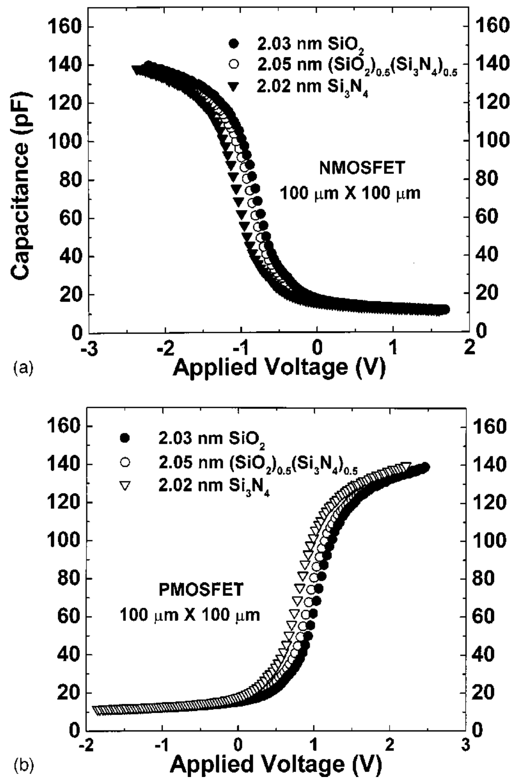

FIG. 4. Capacitance–voltage ($C-V$) curves for (a) NMOS and (b) PMOS devices (EOT ~2 nm) with RPAO interfaces (~0.6 nm) and RPECVD oxide, oxynitride, and nitride bulk dielectric layers. These devices have been subjected to a 900 °C RTA in an inert ambient (e.g., He).

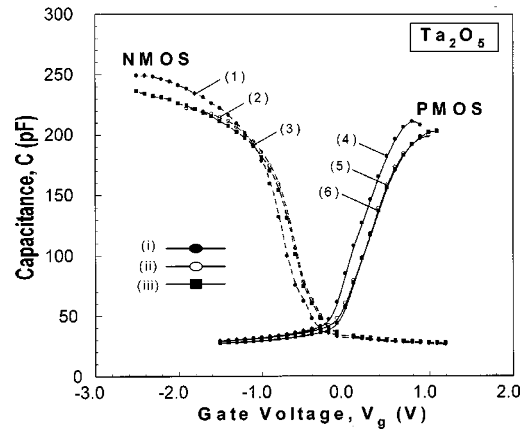

FIG. 5. Capacitance–voltage ($C-V$) plots for devices with stacked gate dielectrics comprised on RPAO SiO_2 or nitrided RPAO SiO_2 and RPECVD Ta_2O_5. For comparison, these plots also include $C-V$ traces for RPECVD Ta_2O_5 directly onto HF-last silicon substrates. These devices have been subjected to a 800 °C RTA in an inert ambient (e.g., He). The solid circles are for deposition on HF-last Si, the open circles for an RPAO SiO_2 layer, and the solid squares are for monolayer nitrided RPAO SiO_2. The EOT values are: (1) 1.09, (2) 1.20, (3) 1.17, (4) 1.19, (5) 1.30, and (6) 1.29 nm.

trace, the effective fixed charge densities are $\sim 2.5\times 10^{12}$ and $7.5\times 10^{11}\,\mathrm{cm}^{-2}$, respectively, for the devices with Si oxynitride and Si nitride stacked dielectrics. Extraction of channel mobilities from the analysis of field effect transistor (FET) current drive from devices on the same wafers as the respective MOS CAPs indicates interfacial defect charged state levels, $N_{it} < 2\times 10^{10}\,\mathrm{cm}^{-2}$.[23] Since the charge densities extracted from the $C-V$ data are much greater than N_{it}, these data, as well as mobility data for other PMOSFETs with different interface layer thickness, establish that the fixed charge densities for the devices in Figs. 4(a) and 4(b) must reside at the internal dielectric interface, and not at the Si-dielectric interface.

The $C-V$ data in Fig. 5 are for MOS CAPs in which RPECVD Ta_2O_5, deposited at 300 °C, and subjected to a rapid thermal anneal (RTA) at 800 °C, replaces the RPECVD SiO_2 layers of the devices in Figs. 4(a) and 4(b). The flat band voltages for Ta_2O_5 devices with RPAO interface layers with and without monolayer interface nitridation on both *n* and *p* substrates respectively the same to <0.002 V, are displaced to more positive values than the respective flat band voltages of devices *without* the RPAO step.[26] Additionally, the flat band voltages for the Ta_2O_5 capacitors on *n*-type and *p*-type substrates RPAO interfaces are the same to <0.005 V as for devices with RPECVD SiO_2 dielectrics fabricated on substrates from the same wafer lots and with equivalent RPAO predeposition processing. This indicates that there is no significant fixed charge at the internal SiO_2–Ta_2O_5 interfaces of either the NMOS or PMOS CAPs of Fig. 5. Based on the high capacitance of these devices, the limiting uncertainty of the $C-V$ measurements and the comparisons with the SiO_2 devices, $Q_f < 6\times 10^{10}\,\mathrm{cm}^{-2}$ at the SiO_2–Ta_2O_5 internal dielectric interfaces.

III. DISCUSSION

The experimental data presented above have established two interfacial properties that can place fundamental limitations on the performance of MOSFET devices with stacked gate dielectrics. These are (i) interfacial transition regions at the Si-dielectric interface, which correspond to about 1 monolayer of Si in suboxide bonding arrangements in excess of what is required at an ideal, abrupt interface, and (ii) fixed positive charge at internal dielectric interfaces of some of the stacked dielectrics so far studied. The constancy of transition region thickness in Si–SiO_2 interfaces prepared in different ways, e.g., plasma, chemical, and thermal oxidation, but subjected to a post-oxidation anneal at about 900 °C is taken as evidence for the intrinsic nature of this transition region. This assumption is supported by the results of Chen and Gibson,[27] which demonstrated that thermal oxidation at 900 °C, gave a higher degree of atomic scale surface roughness than a combination of thermal oxidation, and post-oxidation annealing at the same 900 °C temperature. This means that effects associated with the kinetics of the oxidation process, as for example a nonuniform thermal history for different portions of the film and the interface, can be eliminated by subjecting the interface to an annealing step that subjects the entire gate stack to the same temperature-time profile.

The Si–SiO_2 interface plays a determinant role in several important device properties. For example, differences in atomic-scale structure between nitrided and non-nitrided

plasma-oxidized interfaces as obtained from analysis of XPS data,[15] have accounted for quantitative differences in tunneling current between devices with and without interface nitridation.[11] In addition, it has been established that the time to dielectric breakdown (TDBD) in SiO_2–Ta_2O_5, SiO_2–SiO_2, SiO_2–$[(SiO_2)_{0.5}$–$(Si_3N_4)_{0.5}]$, and SiO_2–Si_3N_4 stacks is determined by the Si–SiO_2 interface layer, independent of the chemical nature of the bulk dielectric film of the stacked structure.[23,28]

Phillips has suggested that the Si–SiO_2 interface should be considered as a material system whose mechanical and bonding properties can be treated by extending concepts of constraint/rigidity theory, previously applied to bulk glasses, to that interface.[7] Within the context of this suggestion, the Si–SiO_2 interface is a boundary between a *rigid* crystalline Si material and an effectively *floppy* and ideal amorphous material, SiO_2. The value of $N_{av}=2.67$ for SiO_2 is higher than the ideal value of 2.4 which corresponds to an average number of bonding constraints per atom, C_{av}, equal to 3.[2] However, the unusually weak bond bonding force constant at the two-hold coordinated oxygen atom sites effectively removes one bonding-bending constraint for each oxygen atom, so that the average number of constraints per atom in SiO_2 is reduced from 3.67 to 3.[3] This places SiO_2 exactly at the boundary between floppy and rigidly constrained dielectrics and/or other materials. The interface between overconstrained and floppy regions has been studied in bulk glasses,[4,5] both theoretically and experimentally. Constraint theory has recently been applied to microscopically inhomogeneous chalcogenide glasses, e.g., Ge–Se alloys than span the range from *floppy* amorphous Se to the *rigid* or overconstrained, chemically ordered compound composition $GeSe_2$, with 33.3 at. % Ge.[5] Application of constraint theory has accounted for two-phase transitions driven by percolation between the floppy Se phase and the rigid $GeSe_2$ composition. Even though N_{av} is the same for SiO_2 and $GeSe_2$, the average number of constraints per atom, C_{av}, in $GeSe_2$ is larger, 3.67, because the bond bending constraint at the twofold-coordinated Se atoms cannot be neglected. Theory has predicted that in the Ge–Se alloys, overconstrained regions containing Ge–Se bonds are separated from floppy regions, regions containing only Se–Se bonds by *encapsulating transition regions* that are (i) self-organized, (ii) not mechanically strained, but (iii) more constrained with higher values of N_{av} than in the floppy regions (see Fig. 6).[5] Experimental results have indicated two percolation ''phase'' transitions which are consistent with this description of the microstructure of these alloys.[4] To a good approximation, the first of these transitions is where overconstrained, but not mechanically strained regions percolate, and the second is where the overconstrained regions percolate.

Extrapolating these results from microscopically inhomogeneous glasses to the Si–SiO_2 interface provides a theoretical framework for describing the intrinsic interfacial transition regions identified above. Constraint theory has previously been applied to semiconductor dielectric interfaces to explain differences in the defect densities between

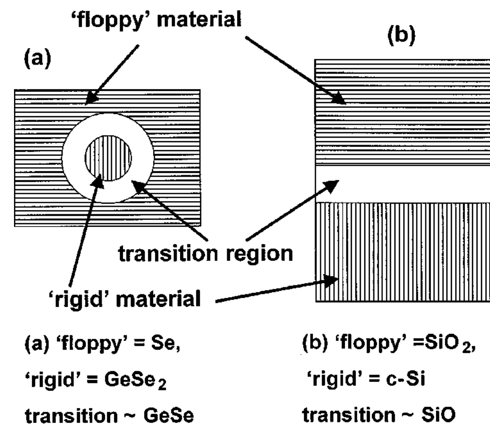

FIG. 6. Schematic representation of transition regions encapsulated between rigid and floppy regions (a) internal to a microscopically inhomogeneous glass alloy, and (b) at a crystalline semiconductor–noncrystalline dielectric interface.

Si–Si_3N_4 and Si–SiO_2 interfaces.[6,8] In this application, mechanical bonding constraints at the interface have been characterized in terms of the average number of bonds per atom in the interfacial region. Following Ref. 6, a *mean-field* interface bonding structure has been *defined* by 0.5 molecular layers of Si (0.5 atoms and two bonds), and 1.5 molecular layers of the dielectric film (SiO_2 or Si_3N_4). Interface nitridation has also been taken into account by inserting one atomic layer of nitrogen between the Si substrate and SiO_2 layer. The mean-field description of the Si–SiO_2 interface is used as a point of reference for mean-field comparisons involving other dielectrics. The Si–SiO_2 interface is characterized by an average number of bonds/atom, $N_{av}^{*}=2.86$. Using a mean-field bonding model, N_{av} equals 3.47 for a Si–Si_3N_4 interface, and 2.89 for a monolayer nitrided Si–SiO_2 interface. As discussed in Ref. 6, the concentration of defects relative to a reference Si–SiO_2 interface is expected to scale as $(N_{av}-N_{av}^{*})^2$. Based on this type of scaling relationship, the defect concentration at an Si–Si_3N_4 interface is expected to be about three orders of magnitude higher than at a monolayer nitrided Si–SiO_2 interface. This is consistent with experimental results which have indicated a defect density of at least 10^{13} donor-like states/cm^2 in the lower half of the interfacial Si band gap.[8] Constraint theory and the associated scaling relationships cannot predict defect concentrations, nor identify the specific defect bonding arrangements or the energy levels of defect states with respect to the band edges of crystalline Si. Instead, constraint theory can provide scaling relationships that provide a quantitative guideline for comparisons of the type discussed above. In the spirit of these scaling relationships, a value of interfacial $N_{av}\sim 3$ has been proposed in Ref. 6 as a demarcation between device-quality interfaces with defect densities $<10^{11}\,cm^{-2}$, and increasingly defective interfaces with defect densities extending to 10^{12}–$10^{13}\,cm^{-2}$ levels. This parallels a similar criterion applied to defects in homogeneous amorphous thin films.[29]

By analogy with the results for the Ge–Se alloy system,

TABLE I. Application of constraint theory to fixed charge at internal dielectric interfaces.

(i) Average number of bonds per atom in dielectric films, N_{av}			
SiO_2	Si_3N_4	oxynitride alloy, $x=0.5$	Ta_2O_5
2.67	3.43	3.05	2.86
(ii) Step in number of bonds/atom at internal interface with SiO_2, ΔN_{av}			
	Si_3N_4	oxynitride alloy, $x=0.5$	Ta_2O_5
	0.76	0.38	0.19
	(a) Scaling factor for defects/cm^2 scale as $[\Delta N_{av}]^2$		
	0.58	0.14	0.04
	(b) Normalized ratio of defects/cm^2		
	16	4	1
	(c) Fixed charge at internal dielectric interface,[a] Q_f (cm^{-2})		
	Si_3N_4	oxynitride alloy, $x=0.5$	Ta_2O_5
	7.5×10^{11}	2×10^{11}	$<0.6\times10^{11}$

[a]Determined from analysis of $C-V$ data.

the transition regions at the $Si-SiO_2$ interface are at the boundary between an overconstrained material on one side of the boundary, crystalline Si, and an ideal glass forming material on the other, SiO_2 (see Fig. 6). Constraint theory would then predict a transition region between these two regions with bonding properties similar to the transition regions between overconstrained and floppy regions within a glassy alloy. This extension of constraint theory is consistent with the experimental observation that interfaces undergo a chemical and structural relaxation after annealing at ~900 °C which results in a transition region whose bonding arrangements and spatial extent are not dependent on the interface formation process. Consistent with these annealing results, these transition regions are assumed (i) to be self-organized, (ii) not to be mechanically strained, and (iii) to be overcoordinated with respect to the bonding within the floppy component of the interface. A quantitative relationship between (i) interfacial bonding statistics, and (ii) interfacial electrical properties is not predictable from constraint theory. The *defective* local bonding arrangements that define the transition region are expected to have localized states within the gap of the Si substrate which can be electrically active as traps and/or fixed charge. In general, these defect states will not be distributed uniformly over the Si band gap, and their charge state will be depend on the Fermi level position and their occupation. For example, a marked asymmetry has been reported for the degradation of the electron and hole transport in NMOS- and PMOSFETs with $Si-Si_3N_4$ interfaces.[8] This is consistent with the interfacial defects being donor-like and lying in the lower half of the Si band gap.

Scaling based on constraint theory can also provide a basis for understanding differences in fixed charge at internal dielectric interfaces between (i) interfacial RPAO SiO_2 layers, and (ii) Si nitride, Si oxynitride, and Ta_2O_5 gate dielectrics [see Figs. 4(a) and 4(b), and 5]. A scaling variable that characterizes interfacial bond counting is the difference in the average number of bonds/atom, ΔN_{av}, between the pair of dielectrics that define the internal dielectric interface. This approach is summarized in Table I which includes: (i) the average number of bonds/atom, N_{av}, as for SiO_2, Si_3N_4, an oxynitride alloy with a composition $(SiO_2)_{0.5}(Si_3N_4)_{0.5}$, and Ta_2O_5, (ii) the step in the average number of bonds/atom, ΔN_{av}, for internal interfaces between SiO_2 and Si_3N_4, $(SiO_2)_{0.5}(Si_3N_4)_{0.5}$, and Ta_2O_5, (iii) absolute and normalized values of $[\Delta N_{av}]^2$, and (iv) values of fixed charge, Q_f, obtained from an analysis of the relevant $C-V$ data of Figs. 4(a) and 4(b), and 5. Values of Q_f are obtained from the relative shifts in $C-V$ curves so that $Q_f=\Delta V/eC$. The limiting value of Q_f for the internal $SiO_2-Ta_2O_5$ interface is estimated from the accuracy of the voltage determination in the $C-V$ traces in Fig. 5. As shown in Table I, the relative values of Q_f obtained in this are consistent with quadratic scaling.

This extension of constraint/rigidity theory to silicon-dielectric and internal dielectric interfaces has identified localized bonding properties that place fundamental limitations on the performance and reliability of MOSFET devices with stacked gate dielectrics. To date, the only interfacial layers that have been demonstrated not to degrade MOSFET current drive, or equivalently reduce the effective channel mobilities of both holes and electrons, are either ultrathin SiO_2 or nitrided SiO_2. Based on these observations, this article has focused on stacked gate dielectrics consisting of a Si substrate, an ultrathin SiO_2 or nitrided SiO_2 layer, and higher-*k* bulk dielectric films such as Si_3N_4, a Si-oxynitride alloys. e.g., $(SiO_2)_{0.5}(Si_3N_4)_{0.5}$, and Ta_2O_5. Other research results for devices with stacked gate dielectrics have demonstrated excellent performance and reliability, but only when the devices included an interfacial layer of SiO_2, either grown prior to deposition, or formed during a post deposition anneal.[30,31] Therefore, the conclusions drawn in this article extend to other stacked gate dielectric structures as well, including the devices of Refs. 30 and 31 where the interfacial layers were not prepared by RPAO.

Finally, the incorporation of interfacial nitrided SiO_2 layers places limitations on the gains in capacitance that can be achieved by dielectric substitutions. Since these gains are quantified through reductions in EOT, it is informative to discuss these limitations in the context of EOT that is referenced to the dielectric constant of SiO_2. For a single layer replacement dielectric, the minimum value of EOT, defined here as EOT_{min}, scales with the maximum physical thickness, $t_{die(max)}$, of the high-*k* film and therefore, inversely with the ratio of the respective dielectric constants,

$$EOT_{min}=t_{die(max)}(k_{ox}/k). \quad (1)$$

Since an interfacial layer of SiO_2, or nitrided SiO_2 will generally be required for maintenance of interface properties and reliability, the physical thickness of the high-*k* film for a given EOT can be considerably reduced. For example, if the physical thickness of a nitrided oxide interface layer is 0.5–0.6 mn, then the contribution of this layer to EOT is reduced to approximately 0.35 mn because the combined dielectric constant increases due to an inherent suboxide transition region and interface nitridation. EOT is then given by

$$EOT=EOT_{min}-0.35\ \ nm. \quad (2)$$

The physical thickness, t_{die}, of the alternative dielectric for a given EOT is also reduced from the value given in Eq. (1), but by an even larger amount,

$$t_{die} = t_{die(max)} - (0.35\ \text{nm}) \times (k/k_{ox}), \quad (3)$$

thereby impacting on the anticipated tunneling current reductions. The incorporation of interfacial layers which reduce the thickness of the alternative gate dielectric, combined with the dependence of the direct tunneling current on conduction band offset energy and electron tunneling mass, creates a considerable challenge in reducing EOT to less than about 0.6. The need for interfacial layers between the high-*k* dielectric and the Si substrate places a significant limitation on the ultimate scaling of CMOS devices, as well as on any other device structures that may require extremely high capacitance to offset decreases in lateral device dimensions. On the other hand, it has been suggested that direct deposition of Zr and Hf silicates and oxides directly onto H-terminated Si would yield good device performance and reliability without the need for a SiO_2-like interface layer, simply because Zr and Hf are group IV elements which have a similar bonding coordination to SiO_2.[32] This would mean that the limiting EOT values are determined only by relative dielectric constants [see Eq. (1)]. While experiments to date have shown promising results on capacitor structures,[32,33] there is as yet insufficient data available on NMOS- and PMOSFETs to demonstrate that increases in capacitance are not offset by decreases in either the electron or hole channel mobilities. In this regard, factor two degradations in electron channel mobilities have been reported in NMOSFETs with Ta_2O_5 dielectric layers, while hole mobilities in similar devices were at the same level as in devices with SiO_2 dielectrics.[34]

ACKNOWLEDGMENTS

This research has been supported in part by the ONR, SRC, and SRC/SEMATECH FEP Center.

[1] H. Y. Yang, H. Niimi, and G. Lucovsky, J. Appl. Phys. **83**, 2327 (1998).

[2] J. C. Phillips, J. Non-Cryst. Solids **34**, 153 (1979); **43**, 37 (1981).

[3] G. Lucovsky and J. C. Phillips, J. Non-Cryst. Solids **227**, 1221 (1998).

[4] P. Boolchand, K. Feng, D. Selvanathan, and W. J. Weaver, in *Rigidity Theory and Applications*, edited by M. F. Thorpe and P. M. Druxbury (Kluwer/Plenum, New York, 1999), p. 279.

[5] M. F. Thorpe, D. J. Jacobs, M. V. Chubynsky, and J. C. Phillips, J. Non-Cryst. Solids (in press).

[6] G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, Appl. Phys. Lett. **74**, 2005 (1999).

[7] J. C. Phillips, J. Vac. Sci. Technol. B **17**, 1803 (1999).

[8] V. Misra, Z. Wang, Y. Wu, H. Niimi, G. Lucovsky, J. J. Wortman, and J. R. Hauser, J. Vac. Sci. Technol. B **17**, 1836 (1999).

[9] Y. Yu and J. Tersoff, J. Vac. Sci. Technol. B (to be published).

[10] H. Niimi, H. Y. Yang, and G. Lucovsky, in *Characterization and Metrology for ULSI Technology: 1998 International Conference*, edited by D. G. Seiler, A. C. Diebold, W. M. Bullis, T. J. Shaffner, R. McDonald, and E. J. Walters (The American Institute of Physics, New York, 1998), p. 273.

[11] H. Niimi and G. Lucovsky, J. Vac. Sci. Technol. A **17**, 3185 (1999); J. Vac. Sci. Technol. B **17**, 2610 (1999).

[12] F. J. Himpsel, F. R. McFeely, A. Taleb-Ibrahimi, Y. A. Yarmoff, and G. Hollinger, Phys. Rev. B **38**, 6084 (1988).

[13] J. W. Keister, J. E. Rowe, J. J. Kolodzie, H. Niimi, N.-S. Tao, T. E. Madey, and G. Lucovsky, J. Vac. Sci. Technol. A **17**, 1340 (1999).

[14] J. W. Keister, J. E. Rowe, Y.-M. Lee, H. Niimi, G. Lucovsky, and G. J. Lapeyre, J. Vac. Sci. Technol. B (in press).

[15] H. Yang, H. Niimi, J. W. Keister, G. Lucovsky, and Jack E. Rowe, IEEE Electron Device Lett. **21**, 76 (2000).

[16] G. Lucovsky, A. Banerjee, B. Hinds, B. Claflin, K. Koh, and H. Yang, J. Vac. Sci. Technol. B **15**, 1074 (1997).

[17] M. Weldon, K. T. Queeny, Y. J. Chabal, B. B. Stefanov, and K. Raghavachai, J. Vac. Sci. Technol. B **17**, 1795 (1999).

[18] D. Muller *et al.*, Nature (London) **399**, 758 (1999).

[19] G. Lucovsky, S. S. Kim, D. V. Tsu, G. G. Fountain, and R. J. Markunas, J. Vac. Sci. Technol. B **7**, 861 (1989).

[20] S. S. Kim and G. Lucovsky, J. Vac. Sci. Technol. A **8**, 2039 (1990).

[21] T. Yasuda, Y. Ma, S. Habermehl, and G. Lucovsky, Appl. Phys. Lett. **60**, 434 (1992).

[22] G. Lucovsky, J. J. Wortman, T. Yasuda, X-L Xu, V. Misra, S. V. Hattangady, Y. Ma and B. Hornung, J. Vac. Sci. Technol. B **12**, 2839 (1994).

[23] H. Yang and G. Lucovsky, IEDM Tech. Dig. 245 (1999).

[24] S. V. Hattangady, H. Niimi, and G. Lucovsky, J. Vac. Sci. Technol. A **14**, 3017 (1996).

[25] G. Lucovsky, IBM J. Res. Dev. **43**, 301 (1999), and references therein.

[26] H. Niimi, R. S. Johnson, and G. Lucovsky (unpublished).

[27] X. Chen and J. M. Gibson, Appl. Phys. Lett. **70**, 1462 (1997).

[28] R. Degraeve, B. Kaczer, M. Houssa, G. Groeseneken, M. Hyens, J. S. Jeon, and A. Halliyal, IEDM Tech. Dig. 327 (1999).

[29] G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, J. Vac Sci. Technol. B **17**, 1806 (1999).

[30] S. C. Song, H. F. Luan, Y. Y. Chen, M. Gardner, J. Fulford, M. Allen, and D. L. Kwong, IEDM Tech. Dig. 373 (1998).

[31] T. P. Ma, IEEE Trans. Electron Devices **45**, 680 (1998); X. Guo and T. P. Ma, IEEE Electron Device Lett. **19**, 207 (1998).

[32] G. D. Wilk and R. M. Wallace, Appl. Phys. Lett. **74**, 2854 (1999).

[33] G. D. Wilk, R. M. Wallace, and J. M. Anthony, J. Appl. Phys. **87**, 484 (2000).

[34] A. Chatterjee *et al.*, IEDM Tech. Dig. 777 (1998).