Transition from thermally grown gate dielectrics to deposited gate dielectrics for advanced silicon devices: A classification scheme based on bond ionicity

Gerald Lucovsky[a)]
Departments of Physics, Electrical, and Computer Engineering, and Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina 27695-8202

(Received 2 April 2001; accepted 25 April 2001)

This article discusses the bonding chemistry of alternative high-*k* gate dielectrics that have been considered for advanced complementary metal–oxide–semiconductor devices. The replacement of SiO_2 by alternative gate dielectrics requires a transition from a thermally deposited native oxide to a deposited gate dielectric. A classification scheme based on bond ionicity separates alternative gate dielectric materials into three groups that are differentiated by their amorphous morphology and electronic structure and properties. This scheme establishes trends between bond ionicity and (i) the average bonding coordination of the constituent atoms, (ii) the thermal stability against chemical phase separation and/or crystallization, and (iii) the dielectric constant. It also provides a framework for the evaluation of different criteria that have been proposed for optimization of alternative high-*k* metal and transition metal oxides, and their alloys with SiO_2 and Al_2O_3. Based on technology targets for device and wafer performance and reliability, there is as yet no *ideal* replacement for SiO_2 that increases capacitance while at the same time maintaining low levels of interfacial defects.
 [DOI: 10.1116/1.1379317]

I. INTRODUCTION

The scaling of lateral dimensions for complementary metal–oxide–semiconductor (CMOS) devices to less than 100 nm requires a parallel scaling of the equivalent oxide thickness (EOT) of the gate dielectric to less than 1.5 nm as referenced to the dielectric constant, *k*, of SiO_2.[1] This moves the dielectric into a thickness regime in which direct tunneling for an oxide bias of 1 V begins to exceed 1 $A\,cm^{-2}$. These levels of tunneling leakage have been deemed detrimental to device operation and reliability. One obvious way to reduce tunneling while maintaining the required levels of gate oxide capacitance is to substitute physically thicker alternative dielectrics with higher values of *k* for thermally grown SiO_2 and thermally and/or plasma nitrided thermally grown SiO_2. There are additional performance and reliability requirements for these alternative dielectrics that are as important as the gate dielectric capacitance for field effect transistor (FET) operation. For example, (i) interfacial trap and fixed charge densities must be sufficiently low so that channel transport of electrons and holes is not degraded, (ii) the interfacial band offset energy between the conduction band of crystalline Si and the alternative dielectric must not be reduced to an extent where it (a) compensates for gains in increased thickness, or (b) provides a transport pathway for charge injection and subsequent trapping in the bulk dielectric, and (iii) lower-*k* interfacial layers formed parasitically during deposition, or intentionally grown or deposited prior to deposition must not significantly reduce the gate dielectric capacitance, and the thickness of the high-*k* component of the gate stack that contributes to a given value of capacitance. A large number of oxides and oxide alloys have been proposed as replacements for thermally grown SiO_2, and these have been discussed at length in a recent review article.[2] This article addresses other issues relative to alternative gate dielectrics, and specifically establishes correlations between (i) the chemical bonding and fundamental electronic structure of alternative gate dielectrics and (ii) properties that are important in the CMOS applications. The alternative dielectrics that are addressed include metal oxides such as Al_2O_3, transition metal oxides such as TiO_2, ZrO_2, HfO_2, Y_2O_3, La_2O_3, and Ta_2O_5, as well as alloys of these transition metal oxides with SiO_2 and Al_2O_3, the so-called *silicate* and *aluminate* alloys.[2] This article is further restricted to noncrystalline or amorphous alternative dielectrics, and therefore does not address issues relevant to epitaxial dielectrics such as crystalline strontium and barium titanate, and their barium–strontium titanate alloys.[3,4] Based on fundamental aspects of bonding chemistry and amorphous morphology, there is as yet no *ideal* replacement for SiO_2, that meets the requirements of the Semiconductor Industry Association Roadmap targets for device performance and reliability.

The incorporation of alternative dielectrics other than thermally grown nitrided silicon oxide will require significant changes in gate stack processing and process integration. Instead of being *grown* directly on the Si substrate, the gate dielectric film must be formed by a deposition process. For thermally grown SiO_2, the $Si–SiO_2$ device interface is continuously regenerated as the SiO_2 film growth proceeds, and is not formed at the starting surface of the Si substrate. In contrast, for deposited dielectrics the interface must either be formed at the starting Si surface either prior to film deposition, during the initial stages of deposition by interface reactions with the process gases, or after deposition by a post-

a)Electronic mail: gerry_lucovsky@ncsu.edu

deposition anneal in an oxidizing ambient. Devices employing deposited SiO_2, and/or silicon nitride and oxynitrides have required separate and independent steps for interface formation, low-temperature film deposition, and post-deposition annealing (see Ref. 5 and references therein).

There are additional issues relative to the choice of gate electrode materials; e.g., polycrystalline Si and Si, Ge, elemental metals, and metal oxides, that are beyond the scope of this article, but are crucial to the fabrication and successful operation of the gate stack in aggressively scaled advanced CMOS devices.

This article will not attempt to address in depth issues of processing or process integration, but instead will focus on a classification scheme for metal oxides and their silicate and aluminate alloys that provides insights into properties that determine their performance as gate dielectrics. The classification scheme is based on bond ionicity and is directly correlated with the electronic energy states of the constituent atoms through the atomic electronegativity, X. Bond ionicity, l_b, is determined from atomic electronegativity differences through an empirical relationship first introduced by Pauling.[6] The atomic scale parameters X and l_b separate elemental and binary oxides into three groups with different amorphous morphologies and technologically important properties, and also identify the importance of the oxygen atom bonding coordination in promoting differences in bonding morphology and properties.

II. AMORPHOUS MORPHOLOGY OF NONCRYSTALLINE SOLIDS

Zallen[7] has discussed in some detail the three different atomic scale amorphous morphologies for noncrystalline solids (i) continuous random networks (CRNs) as exemplified by SiO_2 with predominantly covalent bonding between the constituent atom pairs (ii) modified continuous random networks (MCRNs) as exemplified by silicate alloys in which metal atom ionic bonds *disrupt* and modify the covalently bonded CM structure, and (iii) random close packed (RCP) nonperiodic solids comprised entirely of negative and positive ions.

In the CRN, each atom is bonded according to its primary chemical valence, e.g., the Si atoms of SiO_2 are four-fold coordinated and the oxygen atoms are two-fold coordinated. The predominantly covalent bonds generally can display sigma, σ, as well as pi, π, bonding character.[6] The randomness of the SiO_2 network, which is the primary *source* the configurational entropy that promotes the formation of CRNs with low densities of intrinsic bonding defects ($\leqslant 10^{16}\,cm^{-3}$) derives from two sources: (i) a large spread in bond angle at the O-atom sites, $145° \pm 25°$, and (ii) a random distribution of dihedral angles.[7]

Phillips has demonstrated that the perfection of CRNs relative to bonding defects can be correlated with the average number of bonds/atom, N_{av}, and the average number of bonding constraints/atom, C_{av}.[8,9] In a series of seminal papers, it was demonstrated that the determinant criterion for ideal bulk glass formation is obtained by matching C_{av} to the dimensionality of the network, e.g., $C_{av} \sim 3$ for three-dimensional CRNs such as SiO_2. For CRNs in which the atoms are either two-, three-, or four-fold coordinated, and the bonding geometries of the three- and four-fold coordinated atoms are nonplanar, C_{av} is directly proportional to N_{av} by the following relationship,

$$C_{av} = 2.5N_{av} - 3. \tag{1}$$

Equation (1) is derived from the relations between bonding coordination and the number of valence bond-stretching and bond-bending constraints that apply. If m is the coordination of one of the network atoms, and is greater than or equal to two, then the number of stretching constraints per atom is $m/2$, and the number of bending constraints per atom is $2m$-3. Based on Eq. (1), N_{av} for an ideal CRN in which the bonding is nonplanar at three- and four-fold coordinated atomic sites, is 2.4, as for the compound $As_2S(Se)_3$, and the chalcogenide rich Ge–S(Se) alloys, $GeS(Se)_4$.[8,9] If the atomic coordination is equal to, or greater than three, and the atom is in a planar bonding arrangement, e.g., B in B_2O_3, or N in Si_3N_4, then the number of bond-bending constraints per atom is reduced to m-1. Additional considerations apply to terminal atoms with a coordination of one; however these issues are not relevant to the alternative gate dielectrics that are addressed in this article.

For SiO_2, $N_{av} = 2.67$, so that $C_{av} = 3.67$ and is sufficiently high for the creation of bulk defects. However, since the bending force constant at the O-atom site is exceptionally weak, one bond-bending constraint can be neglected, and C_{av} is reduced to three. This reduction of constraints explains the ease of glass formation for bulk SiO_2, and the unusually low density of intrinsic defects that can be obtained in thermally grown and deposited SiO_2 thin films.[10] The concentration of bulk and interfacial defects in overconstrained dielectrics has been shown to increase as the square of the difference between the number of bonds/atom and the number of bonds/atom in an idealized low defect density material, and is a factor in Si_3N_4 and silicon oxynitrides, both in bulk films and in their interfaces with Si and SiO_2.[11,12]

The *second* amorphous morphology includes silicate and aluminate alloys, and the elemental oxides Al_2O_3, TiO_2, and Ta_2O_5, which have modified CRN structures that include metal atom ions. In the silicates dielectrics, the covalently bonded network is disrupted by the introduction of ionic metals such as Na, Ca, Zr, etc. In elemental oxides, the structure is more complicated, and is better characterized by the average bonding coordination of the oxygen atoms. In particular, the bonding coordination of the oxygen atoms in these modified network materials is increased from two in the CRNs up to approximately three in the MCRN dielectrics.[13] As an example, tetrahedrally coordinated Al groups in noncrystalline Al_2O_3 with a net negative charge comprise 75% of the oxide structure, and provide the network bonding component. Octahedral interstitial voids within this network are sites where Al^{3+} ions are incorporated through dative bonds with nonbonding orbitals of the bridging oxygen atoms of the network.[13] This increases the

average coordination of the oxygen atoms to three. As a second example, IR and Raman spectra of Ta_2O_5 noncrystalline thin films display spectroscopically distinct bond-stretching and bond-bending vibrations consistent with a modified network structure with a tantalum atom coordination of six and an average oxygen atom coordination of 2.4;[13] however, comparisons between the infrared spectra of amorphous and crystalline Ta_2O_5 suggest that the same local coordination is present in both phases. Since crystalline Ta_2O_5 contains an equal mix of six-fold and eight-fold coordinated Ta, the average coordination of oxygen is higher, 2.8 instead of 2.4.

Finally, the IR spectra of SiO_2-rich $(ZrO_2)_x(SiO_2)_{1-x}$ alloys, as well as other transition metal silicate alloys, display vibrational spectra that are also consistent with a modified network structure with an oxygen atom bond coordination less than three.[13,14] It has not as yet been demonstrated that the bond-constraint approach of Phillips and Kerner that has been successfully applied to silicate alloys with low concentrations of group I and II alloy additives, e.g., Na and Ca, respectively, and for low concentrations of Al_2O_3 in SiO_2,[15] is also applicable to noncrystalline Al_2O_3, TiO_2, and Ta_2O_5 and to transition metal silicate and aluminate alloys. For example, the effective coordination of Zr(Hf) atoms in low concentration alloys, e.g., up to about $x \sim 0.2$ in $(ZrO_2)_x(SiO_2)_{1-x}$ alloys, is four as determined by extended x-ray fine structure absorption spectroscopy (EXAFS) and then increases to eight at concentrations between $x \sim 0.5$ and 1.0.[16]

The last group of noncrystalline oxides has an amorphous morphology that can be characterized as a random closed packing of ions. It includes transition metal oxides, as well as Zr and Hf silicate alloys in which the ZrO_2 and HfO_2 fractions, x, are greater than 0.75. These transition metal–oxide-rich Zr and Hf silicate alloys display bond-stretching and bond-bending vibrational modes of the silicate ions, SiO_4^{4-}, in the frequency ranges from 850 to 1100 cm^{-1}, and from 400 to 600 cm^{-1}, respectively, as well as Zr–oxygen bond-stretching vibrations at lower frequencies, less than 600 cm^{-1} and extending to at least 300 cm^{-1}.[16] The average oxygen coordination in Hf and Zr oxide, as well as other group III B transition metal oxides, Y_2O_3, La_2O_3, etc., is approximately equal to four.

III. CLASSIFICATION SCHEME BASED ON BOND IONICITY

A classification scheme based on bond ionicity distinguishes between the three different classes of noncrystalline dielectrics with different amorphous morphologies just discussed. There are several different ways to approach bond ionicity. The most rigorous way is through calculations of the electronic structure with the application of an *atomic scale metric* such as atomic radius that is defined in a self-consistent manner and preserves the overall charge neutrality. However, there is an intimate relationship between electronegativity, X, and electronic structure that is reflected in the variation of X across the periodic table, so that X provides a useful and pragmatic metric for defining a bond ionicity that is correlated to the electronic structure of the constituent atoms.[6] A definition of bond ionicity, I_b, as originally introduced by Pauling is the basis for a classification scheme of this article. It should be noted that bond ionicity scales have previously been used for predicting changes in bonding coordination as in the eight electron/atom pair crystalline solids, including Si, Ge, GaAs, InSb, ZnSe, CdTe, CuCl, and AgI, etc.[17] This article represents the first attempt at using this type of scaling to discriminate between different bonding morphologies in noncrystalline elemental oxides and their alloys.

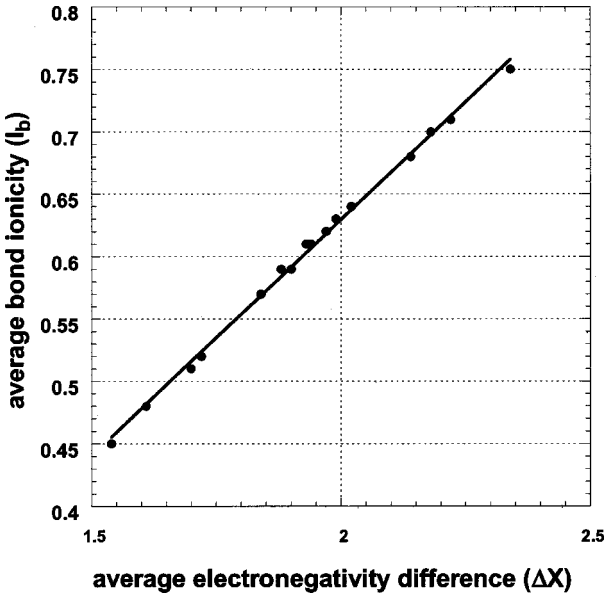

FIG. 1. Plot of an average bond ionicity, I_b, as a function of average electronegativity difference, ΔXr is shown.

If $X(O)$ is the *Pauling electronegativity* of oxygen, 3.44, and $X(\mathrm{Si})$ is the corresponding electronegativity of silicon, 1.90, then the electronegativity difference between these atoms, ΔX, is 1.54. Applying Pauling's empirical definition of bond ionicity, I_b,[6]

$$I_b = 1 - \exp(-0.25(\Delta X)^2), \tag{2}$$

yields a value of I_b for Si–O bonds of ~45%. The range of ΔX values of interest in this article ranges from about from 1.5 to 2.4, corresponding to a bond ionicity range from ~45% to 76%. For this range of ΔX, I_b is essentially a linear function of ΔX as shown in Fig. 1, so that ΔX and I_b are *equivalent* scaling variables. It is important to understand that charge localization on the silicon and oxygen atoms, i.e., effective ionic charges, cannot be determined directly from these values of bond ionicity. In the SiO_2, there are π-bonding interactions involving silicon atom states with d-like symmetries that are involved in a back donation of charge from oxygen to silicon.[18,19] For a detailed discussion of this contribution to the chemical bonding in SiO_2, and a relationship between bond ionicity and charge localization that is unique to SiO_2, the reader is referred to Ref. 19. The electronegativity scale of Sanderson in Ref. 20 provides a

TABLE I. Classification of dielectrics, including amorphous morphology, average electronegativity difference, ΔX, average bond ionicity, I_b, and metal and oxygen atom coordinations.

Dielectric	ΔX	I_b	Coordination	Coordination
Continuous random networks			metal/silicon	oxygen
SiO_2	1.54	0.45	4	2.0
CRNs with network modifiers				
Al_2O_3	1.84	0.57	4 and 6 (3:1 ratio)	3.0
Ta_2O_5	1.94	0.61	6 and 8 (1:1 ratio)	2.8
TiO_2	1.90	0.59	6	3.0
$(ZrO_2)_{0.1}(SiO_2)_{0.9}$	1.61	0.48	8 and 4	2.2
$(ZrO_2)_{0.23}(SiO_2)_{0.77}$	1.70	0.51	8 and 4	2.46
$(ZrO_2)_{0.5}(SiO_2)_{0.5}$	1.88	0.59	8 and 4	3.0
$(TiO_2)_{0.5}(SiO_2)_{0.5}$	1.72	052	6 and 4	2.5
$(Y_2O_3)_1(SiO_2)_2$	1.88	0.59	6 and 4	2.86
$(Y_2O_3)_2(SiO_2)_3$	1.93	0.61	6 and 4	3.0
$(Y_2O_3)_1(SiO_2)_1$	1.99	0.63	6 and 4	3.11
$(Al_2O_3)_4(ZrO_2)_1$	2.02	0.64	4 and 8	3.0
$(Al_2O_3)_3(Y_2O_3)_1$	1.97	0.62	4 and 6	3.0
Random close packed ions				
HfO_2	2.14	0.68	8	4.0
ZrO_2	2.22	0.71	8	4.0
$(La_2O_3)_2(SiO_2)_1$	2.18	0.70	6 and 4	3.5
Y_2O_3	2.22	0.71	6	4.0
La_2O_3	2.34	0.75	6	4.0

different measure of charge localization; however, a comparison between these different electronegativity scales is beyond the scope of this article. The bonds in other good glass formers including oxides such as B_2O_3, P_2O_5, GeO_2, and As_2O_3, and chalcogenides such as $As_2S(Se)_3$ and $GeS(Se)_4$, are generally less ionic in character than SiO_2. For pseudobinary oxide and chalcogenide alloys, e.g., $(SiO_2)_x(B_2O_3)_{1-x}$ and $(GeS_2)_x(As_2S_3)_{1-x}$, respectively, compositionally averaged values of ΔX must be used.

The oxides and chalcogenides previously identified form covalently bonded network structures in which each of the constituent atoms has a coordination that reflects its primary chemical valence; e.g., two for O, S, and Se, four for Si and Ge, three for N and As, etc. The oxides can also contain terminal O atoms that make π bonds with the nonmetal constituent as in P_2O_5. It is important to note that π-bond formation of this type requires the *d* orbitals of the P atoms.[17,18] As just discussed, these covalent-bonded networks have been described as CRNs, where the randomness derives primarily from a spread of bond and dihedral angles (see Ref. 7 and references therein). Many of these materials have low intrinsic defect densities and have been used in electronic applications, e.g., As_2Se_3 is used for electrophotography, and SiO_2 is the gate dielectric for CMOS. This class of dielectrics is differentiated from other oxide dielectrics by a bond ionicity of less than about 47%, or equivalently a ΔX value less than about 1.6. This class of CRN dielectrics also includes Si_3N_4 and silicon oxynitride alloys.

Si_3N_4 and Si oxynitride alloys, $(SiO_2)_{1-x}(Si_3N_4)_x$, also have effective bond ionicities less than 47%. These dielectrics also form CRNs, but because of the three-fold coordination of the *N* have larger values and N_{av} and C_{av}, and hence higher intrinsic defect densities than SiO_2.[10–12] These materials have been studied as alternative gate dielectrics, and fall in the same classification category as SiO_2.[5]

The second class of noncrystalline dielectrics form MCRNs, which include ionic bonding arrangements of metal atoms that disrupt the network structure. This class of dielectrics is characterized by a value of ΔX between about 1.6 and 2.0, or equivalently a bond ionicity between about 47% and 67%. The most extensively studied and characterized oxides in this group in are the metal atom silicate alloys, for example SiO_2 that has been alloyed with Na_2O, CaO, MgO, PbO, etc. and quenched from the melt.[7] This class also includes deposited thin film Al_2O_3, TiO_2, and Ta_2O_5, and transition metal atom silicate alloys such as $(Zr(Hf)O_2)_x(SiO_2)_{1-x}$ in the composition range up to about $x \sim 0.5$.[13] The noncrystalline range of alloy formation in deposited thin films is generally significantly increased with respect to what it is in bulk glasses that are quenched from the melt. The coordination of oxygen atoms in CRNs is typically two, and increases to approximately three in the MCRNs (see Table I). As examples, the coordination of oxygen increases from 2.8 in thin film Ta_2O_5, to 3.0 in Al_2O_3, and increases from 2 to 3 in the group IV B silicate alloys as the ZrO_2 or HfO_2 fraction is increased from very low values, $x \sim 0.05$, to $x = 0.5$.

The third class of noncrystalline oxides are those that have a RCP ionic amorphous morphology.[7] This class of oxides is correlated with $\Delta X > 2$, and a Pauling bond ionicity of greater than about 67%. This group includes transition metal oxides that are deposited by low-temperature techniques including plasma deposition,[13,14,16] and sputtering with postdeposition oxidation.[21] The coordination of the oxygen atoms in these RCP structures is typically four.

Table I includes (i) a listing of representative elemental

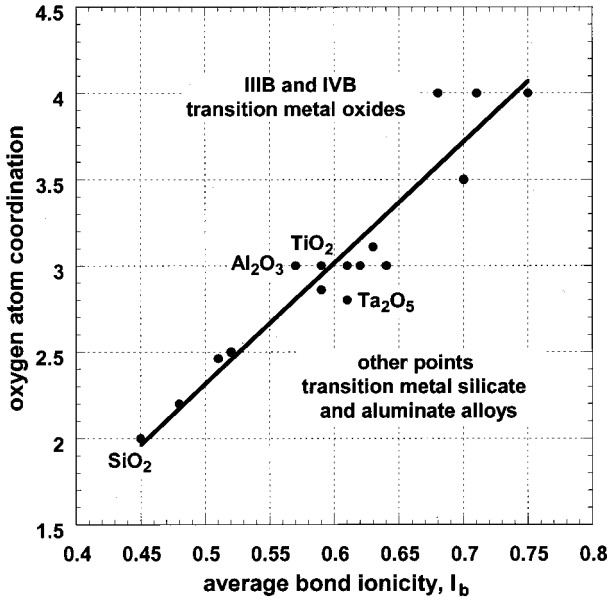

FIG. 2. Plot of oxygen atom coordination as function of average bond ionicity, I_b, is shown.

and binary oxide alloys, (ii) values of ΔX, (iii) values of a Pauling bond ionicity, I_b, and (iv) an average O-atom and cation bonding coordinations, $N_{\mathrm{av}}(\mathrm{O})$. Values of ΔX of ~ 1.6 and ~ 2.0 separate these elemental and binary oxides into three different groups: (i) the CRNs for ΔX up to 1.6, (ii) the MCRNs for ΔX between about 1.6 and 2.0, and (iii) the RCP ionic structures with $\Delta X > 2$. The effective coordination of O atoms equals 2 for the CRNs, is between 2 and 3 for the MCRNs and increases to 4 for the RCP oxides. The values of I_b, calculated from Eq. (2), that separate these three groups are approximately 47% and 67%. The use of other electronegativity scales, e.g., Sanderson,[20] and different definitions of bond ionicity would change the values of bond ionicity that establish the boundaries, but do not change the separation of oxides and oxide alloys into the same three classes, and therefore do not modify any of the qualitative aspects of the proposed classification scheme.

Figure 2 contains a plot of the average oxygen atom coordination as a function of the average bond ionicity, I_b. The values of oxygen atom coordination are either inferred from measurements or measured directly,[13,16] or determined from crystalline compounds.[17] Since I_b is a linear function of ΔX in the regime of interest, the average oxygen atom coordination also scales linearly with ΔX. The monotonic and nearly linear variation suggests a fundamental relationship between charge localization of the oxygen atom and bonding coordination. This is being investigated by our group through molecular orbital theory studies of local bonding groups that are representative of the molecular structure of transition metal–oxygen bonds in the noncrystalline state.[22]

IV. TRENDS WITH BOND IONICITY FOR THE HIGH-*k* DIELECTRICS

There are several trends with increasing bond ionicity for properties of the high-*k* gate dielectrics of the second and third groups that are important for gate dielectric applications. The plot in Fig. 2 suggests that oxygen atom coordination is an important microscopic metric that drives some of the trends which will be discussed. First, the dielectric constants of metal and transition metal oxides, silicates, and aluminates increase as bond ionicity increases; however the thermal stability against crystallization decreases. The trends in the dielectric constant behavior are a reflection of the electronic structure and are beginning to be addressed by band theory calculations.[23,24] For example, $k = 3.8$ for SiO_2, increases to 9 in Al_2O_3, varies between 8 and 11 in Zr and Hf silicate alloys with x in the range of 0.1 to 0.2, and then increases to more than 20 in transition metal oxides ZrO_2, HfO_2, La_2O_3, and Yr_2O_3. The dielectric constants of TiO_2 and Ta_2O_5 are significantly higher than those of the silicates and aluminates of the MCRN group, and are correlated with their relative small band gaps, less than 4.2 eV as compared to greater than 5–6 eV in Al_2O_3, and the other transition metal oxides and silicates.[21,25,26]

Effective crystallization temperatures show the opposite trend, decreasing as the bond ionicity increases.[13] The results obtained to date are for relatively thick films, >10–50 nm, and may not apply to significantly thinner films <5 nm, that are targeted for device applications. SiO_2 cannot be crystallized below its melting temperature of more than 1600 °C; Al_2O_3 crystallizes at 900 °C after a 30 s anneal in Ar, and Ta_2O_5 crystallizes at 800 °C under the same annealing conditions.[13,14] Zr silicate alloys with $x < 0.5$ phase separate after a 30 s 900 °C anneal in Ar into SiO_2 and ZrO_2 with no evidence for crystallization of either phase, whereas alloys with $x > 0.5$ phase separate into noncrystalline SiO_2 and crystalline ZrO_2 under the same annealing conditions.[10,11] Films of ZrO_2 and HfO_2 crystallize at temperatures below 500 °C. The group III B silicates show a qualitatively different behavior that is correlated with a difference in the equilibrium phase diagram between the group III B and IV B silicates. The group III B silicates have congruent melting points for at least two distinct silicate phases, whereas the group IV B Hf and Zr silicates do not a congruent melting point for the single silicate phase.[27] For example, chemical phase separation into SiO_2 and a SiO_2-rich silicate phase occurs for low Y and La concentrations corresponding to a 2:1 mix of SiO_2 and $Y(La)_2O_3$. For lower SiO_2 concentrations, there are two regimes of crystallization and/or chemical phase separation, neither of which involves SiO_2. As noted above, these trends between bond ionicity and crystallization have been identified for relatively thick samples. Crystallization is expected to be dependent on film thickness where quantitatively different behavior can occur when the film thickness and stable crystallite size become comparable.

Band gaps and conduction band offset energies generally both decrease with respect to SiO_2 as k increases.[25,26] The determinant factors are different for the oxides that do not contain transition metals and those which do. The lowest lying conduction bands in nontransition metal oxides, silicates and aluminates are associated with *s* and *p* states of the metals and/or silicon or aluminum, whereas in transition

metals oxides and compounds they are associated with *d* states of the transition metals, and are significantly lower energies relative to the oxygen atom states that determine the approximate position of the top of the valence band.[22]

V. TECHNOLOGY GENERATIONS OF ALTERNATIVE GATE DIELECTRICS

The transition from thermally grown SiO_2 to higher-*k* dielectric materials is presently proceeding in two steps; (i) first, the replacement of thermally grown SiO_2, initially with plasma and thermally-nitrided oxides and/or deposition of nitrides and oxynitride alloys, and then (ii) the replacement of nitrides and oxynitride alloys with alternative high-*k* dielectrics.

There are several basic materials and processing issues that are common to second and third generation gate dielectric structures. Most of these are related to the nature of the interface, or interfacial layer between the Si substrate and the alternative dielectric. Based on studies reported to date, and using a direct tunneling current of 1 $A\,cm^{-2}$ to define the EOT of the replacement dielectric. Si oxynitride alloys with optimized interfaces comprised of nitrided SiO_2 are viable to EOT $\sim$1.1–1.2 nm,[28] and alternative high-*k* transition metal oxides, and silicate and aluminate alloys will be required to extend EOT significantly below 1.0 nm.[1] There are also application specific differences in the level of tunneling current that can be tolerated. For example, mobile devices in general require off state tunneling leakage to be less than about $10^{-3}\,A\,cm^{-2}$, whereas many order of magnitude higher tunneling currents will not degrade performance of devices for desk top applications.[1]

Previously reported studies have demonstrated the need for separate and independent process steps for interface formation and film deposition for deposited SiO_2, and Si nitride and oxynitride alloys.[5,29] Direct deposition in an oxidizing, or other chemically active environment, generally results in subcutaneous chemical reactions leading the formation of defective interfacial layers that have been shown to degrade device performance.[30] The results reported in Ref. 29 demonstrated separate and independent processing steps for interface formation, a 300 °C remote plasma assisted oxidation, followed by a 300 °C remote plasma enhanced chemical vapor deposition (RPECVD) of SiO_2. Remote plasma processing at 300 °C for interface formation and film deposition, combined with a 30–60 s, 900 °C rapid thermal anneal in an inert ambient, Ar, has yielded state-of-the-art MOS devices.[5,31]

These studies have identified two issues relative to deposited dielectrics: (i) high densities of interfacial defects that result from subcutaneous interfacial reactions occurring concurrently with film deposition as for SiO_2 on Si,[30] and (ii) interfacial defects that result from a overconstrained interfacial bonding as occurs for the direct deposition of Si_3N_4 on Si.[11,12] These issues have been studied in detail for the silicon nitrides and oxynitride alloys, and apply to the high-*k* alternative oxide dielectrics as well.

The deposition of high-*k* oxides directly onto crystalline Si can result in interfacial reactions that give rise to defects,[31] and more generally results in high levels of fixed charge which are a direct result of increased bond ionicity.[32] The next two sections will deal respectively with first generation replacement dielectrics, the Si nitrides and oxynitride alloys, and second generation replacement dielectrics, metal oxides and alloys, with an emphasis on transition metal oxides, and their silicate and aluminate alloys.

VI. DEPOSITED Si NITRIDE AND OXYNITRIDE ALLOYS

Remote plasma-deposited and processed Si nitride and Si oxynitride alloy replacement dielectrics have recently been reviewed by the author,[5] and the reader is referred to that article and the references therein. The main points made in that review article are summarized below:

(a) Direct deposition of silicon nitride onto Si results in interfaces with high densities of defects. The defects are predominantly donor-like and are in the lower half of the Si band gap and therefore effect *p*-channel FETs more adversely than they do *n*-channel FETs.[11,12] High defect concentrations have been attributed to overconstrained interface bonding with Si_3N_4 and Si by an extension of the bond constraint considerations of Phillips from glasses to CRN thin films[10] and interfaces and thin CRN thin films and their interface with Si.[11]

(b) Insertion of thin SiO_2 interface layers between the Si substrate and the Si nitride or oxynitride layers as thin as $\sim$0.5 to 0.6 nm reduce defect levels at the Si–dielectric interface to values characteristic of thermally grown Si–SiO_2 interfaces, provided that the interfaces have been subjected to a 900 °C anneal in a nonreactive ambient such as Ar.[5] However, due to a mismatch in N_{av} and C_{av} at the internal dielectric interface, defects are induced at that interface. Typical defect density levels at SiO_2–Si_3N_4 interfaces are $\sim > 10^{12}\,cm^{-2}$, and are reduced to $< 10^{11}\,cm^{-2}$ at SiO_2–Si oxynitride interfaces $(N{:}O = 2{:}1)$.[12] These defect densities are sufficiently high to produce remote scattering of holes in the channels of *p* MOSFETs, and shift threshold voltages in 0.1 V range for devices in which the EOT is $\sim$2.5 nm.[5]

(c) Monolayer scale interface nitridation of 0.5–0.6 nm Si oxide layers has been achieved by exposure to active species generated in a remote N_2/He plasma.[33] The incorporation of approximately one monolayer of nitrogen, as determined by secondary ion mass spectrometry, reduced direct tunneling by approximately one order of magnitude with respect to nonnitrided interfaces.[33] A model based on x-ray photoelectron spectroscopy (XPS) studies of Si suboxide bonding has demonstrated that the decrease in direct tunneling can result from an increased interfacial barrier height that is derived from a direct bonding of N atoms to the Si substrate.[34] The XPS studies indicate that the suboxide interfacial bonding is still present after interfacial nitri-

dation, and that the interfacial bonding sequence has changed from substrate Si to O to SiO_2, to substrate Si to N to Si to O to SiO_2.

(d) In order to optimize both performance and reliability in *p*-channel devices, it is necessary to stop the transport of boron atoms out of the polycrystalline Si gate electrode during the dopant activation anneal. This can be accomplished in oxide–nitride, and oxynitride stacked dielectrics. Reference 35 identifies the areal density of nitrogen atoms necessary to effectively stop boron penetration, $\sim 4-5\times 10^{15}\,\mathrm{cm}^{-2}$, and this can be realized in 0.8 nm of Si_3N_4, or in an optimized oxynitride alloy with a 2:1 ratio of N: O, and a physical thickness of approximately 1.2 nm.

(e) Gate stacks have been prepared by controlling the nitrogen profile in the Si–dielectric structures. Device optimization based on the minimization of direct tunneling and reductions in remote carrier scattering and threshold voltage shifts due to interfacial charged defects have been achieved in oxide–oxynitride alloy stacks with approximately 0.6 nm of interfacial SiO_2, monolayer interface nitridation, and a physical thickness of an optimized oynitride alloy that is adjusted for achieving a particular value of EOT. This approach, and an approach based on oxide–nitride stacks, can yield direct tunneling currents of less than $1\,\mathrm{A\,cm}^{-2}$ for EOT values between 1.1 and 1.2 nm.[5,28] These studies have shown that reliability of the stacked devices improves with decreasing tunneling current.

(f) Other processing approaches based on jet vapor deposition and thermal CVD, with either predeposition or postdeposition interface nitridation have yielded similar results with respect to minimum values of EOT.[5]

VII. DEPOSITED HIGH-*k* DIELECTRICS

There are several significant issues that impact on the performance of devices with deposited transition gate dielectrics. The first group relates to process integration issues. The second group derives from inherent relationships between (a) chemical bonding and physical properties, and (b) device operation. These include (i) reduced interfacial band offset energies, (ii) interfacial fixed charge, (iii) interfacial trapping of electrons, and (iv) coordination dependent dielectric constants. These are illustrated with respect to studies in which the gate dielectrics stacks were prepared by RPECVD. Similar results have been obtained by other deposition approaches, including physical vapor deposition followed by postdeposition oxidation (See Ref. 2 and references therein).

One of the paramount issues in the integration of deposited metal–oxide and metal–oxide alloy dielectrics in Si CMOS devices is the formation of the Si–dielectric interface. As in the case of Si nitride and the Si oxynitride alloys, the interface can be formed (i) during deposition of the alternative dielectric (ii) in a separate and distinct predeposition step, or (iii) during a postdeposition anneal in an oxygen containing ambient. Since there are significant differences in the reaction kinetics for different approaches to film deposition, e.g., thermal CVD, plasma-assisted CVD, atomic layer deposition, and physical vapor deposition, followed by postdeposition oxidation, it is difficult to make any generalizations. At this point in time, the *ideal* or best choice alternative metal oxide containing dielectric has not as yet been identified, and until the *winner* or *winners* emerge, it is not productive to dwell on this point. It is however likely, that the optimum choice of a high-*k* dielectric must have an optimized interface, and will be based on an approach to processing that achieves a compositional profile that yields performance and reliability that are in line with roadmap expectations.

A second process integration issue relates to the replacement of polycrystalline Si and Si, Ge with a gate metal. This change is driven by issues: (i) the high temperature processing required for dopant activation in polycrystalline Si and Si, Ge, and (ii) the additional capacitance that is associated with a depletion region in the polycrystalline films. Metal gate materials fall into several categories, elemental metals, metal alloys, and metal oxides, and nitrides. In order to have low threshold voltages, two different metals with *effective work functions* close to the valence and conduction bands of Si, respectively are required. Several different processing options have been proposed, and the final choice will involve the identification of a gate stack materials package that can be manufactured, and meets device performance and reliability targets.

Due to the inherent differences between the high-*k* and Si-based oxide, nitride, and oxynitride dielectrics are easier to deal with. The major differences in bulk and interface properties derive from the increased bond ionicity of the high-*k* oxides, and in particular on the transition metal oxides and silicate and aluminate alloys which have emerged as candidate materials (see Table I).

The most significant limitation in scaling down the physical thickness of a gate dielectric to increase gate capacitance is the rapid increase in direct tunneling. For SiO_2, this onset occurs for a physical thickness <0.2 to 0.25 nm. The tunneling current also depends on the barrier height for tunneling. Equation (3) illustrates this by giving the tunneling transmission, T, through a square barrier of thickness, t, with a barrier height of E_b, and a tunneling electron mass, m^*,

$$T\sim \exp(-2at[E_b\times m^*]^{0.5}), \tag{3}$$

where a is a constant. Equation (3) identifies the scaling trade-off between t and E_b. It is convenient to normalize the effective capacitive thickness to SiO_2 so that $t=(k/k(SiO_2))t_{ox}$, where t_{ox} is the EOT scaled to thermally grown SiO_2. In general, as k increases, E_b decreases. For example, $k=3.8$ for SiO_2 and $E_b\sim 3.1$ eV, whereas $k=7.6$ for Si_3N_4 and E_b is reduced to ~ 2.1. However, there are *additional inherent* decreases in E_b associated with the transition metal oxides, silicates, and aluminates. These decreases are derived from the *d*-band character of the conduction bands of the transition metal dielectrics.[25] Typical band offset energies are reduced from 3.1 eV for $Si–SiO_2$, to $\sim 2.1–2.5$ eV for $Si–Si_3N_4$ and Al_2O_3, and to less than 1.5

eV for many of the transition metal oxides and silicates, e.g., Ti and Ta oxide, and the group IV B oxides and their silicate alloys.[21] These negate significant decreases in T from increased physical thickness. The band offset energies in the group III B oxides are larger, ~2 eV, or more, and these oxides and their silicate alloys are beginning to receive increased attention.

In transition metal silicate and aluminate alloys, the lowest lying conduction bands or antibonding states of transition metals are derived from atomic *d* states. The conduction band, or antibonding states associated with Si and Al are derived from atomic *s* and *p* states, and are at significantly higher energies.[22] However, the differences in symmetry character, spatial extent, and absolute energies between the *d*-state derived bands of the transition metals and the *s*- and *p*-state derived bands of the Al and Si atoms states makes it unlikely that they will mix in any significant way. This means that transition metal silicate and aluminate alloys are expected to have effective band offset energies similar to the transition metal constituent of the alloy as demonstrated in the results presented in Refs. 21, 25, and 26.

The dielectric properties of the group IV B silicate alloys, $(Zr(Hf)O_2)_x(SiO_2)_{1-x}$, have been shown to display a nonlinear behavior with alloy composition. MOS capacitors with SiO_2-rich Zr and Hf silicates with 3–6 at % Zr(Hf), or $x \sim 0.2$ to 0.3 have been reported to have increased dielectric constants and reduced tunneling currents.[36–39] Reported values of *k* from capacitance–voltage ($C-V$) curves are ~8 to 11, and more than 50% larger than values estimated from a linear extrapolation of *k* between SiO_2, ~3.9, and the compound silicates, ~12. These *enhanced* values of *k* cannot be reconciled with macroscopic dielectric theory that is based on an effective medium approach.[40] This theory applies to mixtures in which chemical bonding of constituents does not change with composition.

The increases in *k* at low Zr and Hf concentrations have been explained by changes in coordination of the Zr and Hf atoms as the alloy fraction of ZrO_2 or HfO_2 is increased.[16] Changes in coordination have been inferred from an interpretation of infrared spectra in the Zr silicate alloys, and confirmed by direct measurement of coordination by EXAFS. At low concentrations, the Zr and Hf atoms are four-fold coordinated, and the coordination increases to eight-fold as the alloy fraction increases to 0.5, the stoichiometric silicate composition. The bonding of Zr and Hf is more covalent at the lower concentrations and the there is a dynamic enhancement in the infrared effective charge. As the coordination increases, and bonding becomes more ionic this enhancement decreases. These changes have been incorporated into a model equation that uses bond order as a scaling parameter. This approach fits the data remarkably well, and is underpinned by microscopic model based on changes in the way Zr and Hf atomic *d*-states contribute to the bonding orbitals as the coordination changes. This effect may not occur in group III B silicate alloys due to the inherent properties of the group three ions and their bonding with oxygen atoms. Coordinations greater or equal to six are the norm, and it is unlikely that a covalent coordination of three is possible.

Values of interfacial fixed charge have been determined in the conventional manner from shifts in flat band voltage relative to devices with the same equivalent oxide thickness with SiO_2 dielectrics, and by studying the flat band voltage as a function of decreasing gate stack capacitance.[32,41] The sign of the charge is reflected in the negative or positive voltage flat band shifts, and the magnitude of the charge in the magnitude of these shifts. The sign of charge has been shown to show a correlation with the local atomic structure of the particular high-*k* dielectric, and the bonding that is anticipated at a Si surface. Based on these results and a propensity for covalent bonding of O to Si at the Si–dielectric interface, the respective bonding arrangements at the $Si-Al_2O_3$, $Si-Ta_2O_5$, and Si–Zr silicate (e.g., $x \sim 0.3$) interfaces are respectively proposed to be $Si-O-Al^{-\delta}$, $Si-O-Ta^{+\delta}$, and $Si-O-Zr^{+\delta}$, where the values of δ are the partial charges on the interfacial atoms due to the ionic character of the high-*k* dielectrics. This in turn suggests a correlation between the sign of the fixed charge and the metal ion of the network component of the oxide in Al_2O_3 and Ta_2O_5, and of the metal ion in the silicates. Studies of group III B silicates in which the bond ionicity is higher than in Al_2O_3, e.g., Y and La silicates, indicate fixed positive charge consistent with the incorporation of Y and La as 3+ ions rather than as network constituents with a negative charge. The magnitude of the fixed charge is more difficult to account for. The convention has been to calculate an areal density of fixed charge defects of unit charge based on the macroscopic charge density obtained from the analysis of $C-V$ data. Alternatively, it can be described in terms of partial charges on atoms, or differences between charge distributions of both signs as in interfacial dipoles. Since the way the fixed charge is distributed influences the channel mobilities of charged carriers, analysis of mobility data from MOSFETs using different microscopic models for the fixed charge should help to resolve this issue.

Another issue relates to electron trapping at S–dielectric interfaces. This has been found in our research group for Ta_2O_5, and $Ta_2O_5-Al_2O_3$ alloys.[42] A simple and effective way to distinguish between fixed charge and injected trapped charge is to perform temperature dependent $C-V$ and current–voltage, $I-V$, studies on, MOS capacitors MOSCAPs with *p*-type and *n*-type substrates. This is the approach used to identify electron trapping in the oxides and oxide alloys of Ref. 42. Analysis of the $C-V$ and $I-V$ data indicates an electron trap about 0.3 eV above the Si conduction band edge for the Ta_2O_5 and the $Ta_2O_5-Al_2O_3$ alloys. This is consistent with the state being an antibonding state of Ta bonded to oxygen.

VIII. SUMMARY

This article has demonstrated that noncrystalline high-*k* dielectrics can be separated into three groups with different amorphous morphologies, and a convenient metric for this separation is bond ionicity. Several important properties of the alternative high-*k* dielectrics, such as dielectric constant

and thermal stability against chemical phase separation and/or crystallization scale with bond ionicity.

At the present time, there is no one single elemental or binary oxide that has emerged as an ideal replacement for thermally grown SiO_2. On the other hand, it is clear that the replacement of SiO_2 by alternative dielectrics will proceed in two steps (i) the introduction of heavily nitrided oxides prepared by thermal, plasma, and eventually deposition techniques, and (ii) the introduction of deposited high-k elemental or binary oxide silicate or aluminate alloys. At the present time, the most promising materials for second generation replacement dielectrics include Zr and Hf group IV B silicate alloys, and the group IV B Zr and Hf oxides. The group III B oxides of Y and La, and their silicate alloys and compounds are beginning to receive increased attention as are aluminate alloys involving group III B and IV B oxides.

Finally, recent studies by x-ray absorption spectroscopy of the conduction band states derived from $4d$ atomic states of Zr in Zr silicate, $(ZrO_2)_x(SiO_2)_{1-x}$ alloys have demonstrated that the energies of these states are independent of alloy composition, film morphology, i.e., the state of crystallinity.[43] The experimental results are supported by a molecular orbital model in which the orbital energies of the top of the valence band in transition metal oxides and their silicate and aluminate alloys are correlated with the atomic $2p$ state energies of oxygen, and the orbital energies of the lowest conduction bands are correlated with the atomic d-state energies of the transition metal atoms.[22]

ACKNOWLEDGMENTS

The research reported in this article reflects the contributions of many of the authors' research students and postdoctoral fellows. In chronological order, the author would like to acknowledge: Yi Ma, T. Yasuda, S. Hattangady, H. Niimi, Y. Wu, H. Yang, R. Therrien, G. Rayner, R. Johnson, J. G. Hong, and H. Kang. The author would also like to acknowledge important discussions with faculty at NC State University, Professors V. Misra, G. Parsons, and J. Hauser. Finally, research support for the work presented in this invited article comes from the National Science Foundation Engineering Research Center for Advanced Electronic Materials Processing, the Office of Naval Research, the Air Force Office of Scientific Research, the Semiconductor Research Corporation, SRC, and the SRC/Sematech Center for Front End Processes.

[1]*The International Technology Roadmap for Semiconductors* (SIA, Semiconductor Industry Association, San Jose, CA, 1999).
[2]G. D. Wilk, R. W. Wallace, and J. M. Anthony, J. Appl. Phys. (in press).
[3]R. A. Mckee, F. J. Walker, and M. F. Chisholm, Phys. Rev. Lett. **81**, 3104 (1999).
[4]Z. Yu, R. Droopad, J. Ramdani, J. A. Curless, C. D. Overgaard, J. M. Finder, K. W. Eisenbeiser, J. Wang, J. A. Hallmark, and W. J. Ooms, Mater. Res. Soc. Symp. Proc. **567**, 427 (1999).
[5]G. Lucovsky, IBM J. Res. Dev. **43**, 301 (1999).
[6]L. Pauling, *The Nature of the Chemical Bond*, 3rd ed. (Cornell University, Ithaca, NY, 1948).
[7]R. Zallen, *The Physics of Amorphous Solids* (Wiley, New York, 1983).
[8]J. C. Phillips, J. Non-Cryst. Solids **34**, 153 (1979).
[9]J. C. Phillips, J. Non-Cryst. Solids **43**, 37 (1981).
[10]G. Lucovsky and J. C. Phillips, J. Non-Cryst. Solids **227**, 1221 (1998).
[11]G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, Appl. Phys. Lett. **74**, 2005 (1999).
[12]G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, J. Vac. Sci. Technol. B **17**, 1806 (1999).
[13]B. Rayner, H. Niimi, R. Johnson, R. Therrien, G. Lucovsky, and F. L. Galeener, AIP Conf. Proc. **550**, 149 (2001).
[14]B. Rayner *et al.*, Mater. Res. Soc. Symp. Proc. **661**, c1 3.1 (2001).
[15]J. C. Phillips and X. Kerner, Solid State Commun. **117**, 47 (2001).
[16]G. Lucovsky and G. Rayner, Appl. Phys. Lett. **77**, 2912 (2000).
[17]J. E. Huheey, *Inorganic Chemistry, Principles of Structure and Reactivity* 2nd ed. (Harper and Row, New York, 1978).
[18]F. A. Cotton and G. Wilkinson, *Inorganic Chemistry*, 3rd ed. (Interscience, New York, 1972).
[19]L. Pauling, Am. Mineral. **65**, 321 (1960).
[20]R. T. Sanderson, *Chemical Bonds and Bond Energy* (Academic, New York, 1970).
[21]S. Miyazaki and M. Hirose, AIP Conf. Proc. **550**, 89 (2000).
[22]Y. Zhang, J. L. Whitten, and G. Lucovsky, to be presented at the International Conference on Formation of Semiconductor Interfaces in Sapporo, Japan, June 2001; Appl. Surf. Sci. (to be published).
[23]K. G. Cho, presented at Physics and Chemistry of Semiconductor Interfaces 2000, Orlando, FL; J. Vac. Sci. Technol. B (to be published).
[24]G.-M. Rignanese, X. Gonze, and A. Pasquerello, Phys. Rev. B **63**, 104 (2001).
[25]J. Robertson and C. W. Chen, Appl. Phys. Lett. **76**, 1168 (1999).
[26]J. Robertson, J. Vac. Sci. Technol. B **18**, 1785 (2000).
[27]N. A. Toropov and I. A. Bondar, Izv. Akad. Nauk SSSR, Otd. Tekh. Nauk, Metall. Topl. **4**, 547 (1961).
[28]H. Yang and G. Lucovsky, Tech. Dig. Int. Electron Devices Meet. 245 (1999).
[29]T. Yasuda, Y. Ma, S. Habermehl, and G. Lucovsky, Appl. Phys. Lett. **60**, 434 (1992).
[30]S. S. Kim and G. Lucovsky, J. Vac. Sci. Technol. A **8**, 2039 (1990).
[31]H. Luan *et al.*, Mater. Res. Soc. Symp. Proc. **567**, 385 (1999).
[32]G. Lucovsky, J. C. Phillips and M. F. Thorpe, AIP Conf. Proc. **550**, 149 (2001); AIP Conf. Proc. **550**, 154 (2000).
[33]H. Niimi and G. Lucovsky, J. Vac. Sci. Technol. A **17**, 3285 (1999); J. Vac. Sci. Technol. B **17**, 2610 (1999).
[34]H. Yang, H. Niimi, J. W. Keister, G. Lucovsky, and J. E. Rowe, IEEE Electron Device Lett. **21**, 76 (2000).
[35]Y. Wu, H. Niimi, H. Yang, G. Lucovsky, and R. B. Fair, J. Vac. Sci. Technol. B **17**, 1813 (1999).
[36]G. D. Wilk and R. M. Wallace, Appl. Phys. Lett. **74**, 2854 (1999).
[37]G. D. Wilk, R. M. Wallace, and J. M. Anthony, J. Appl. Phys. **87**, 484 (2000).
[38]G. D. Wilk and R. M. Wallace, Appl. Phys. Lett. **76**, 112 (2000).
[39]V. Misra (unpublished).
[40]N. Jayasundere and B. V. Smith, J. Appl. Phys. **73**, 2462 (1993).
[41]R. S. Johnson, G. Lucovsky, and I. Baumvol, J. Vac. Sci. Technol. A, these proceedings.
[42]R. S. Johnson and G. Lucovsky, Appl. Surf. Sci. (to be published).
[43]G. Lucovsky *et al.*, Appl. Phys. Lett. (submitted).