Ultrathin nitrided gate dielectrics: Plasma processing, chemical characterization, performance, and reliability

by G. Lucovsky

The incorporation of nitrogen (N) atoms into ultrathin gate dielectrics 1) at monolayer levels at Si–SiO_2 interfaces reduces tunneling current and defect generation; 2) in bulk nitrides, as in oxide–nitride–oxide (ONO) or oxide–nitride (ON) composite structures, allows the use of physically thicker films without reduced capacitance compared to single-layer oxides; and 3) in nitrided layers at the polycrystalline Si–dielectric interface or in ON dielectrics reduces boron (B) atom out-diffusion from heavily doped p^+ polycrystalline silicon gate electrodes into oxide gate dielectrics. The results presented in this review demonstrate that N atoms can be selectively and independently incorporated into different parts of the gate dielectric by low-temperature remote-plasma-assisted processing. When combined with low-thermal-budget rapid thermal annealing, this yields ultrathin gate dielectrics with performance and reliability which generally exceeds that of single-layer thermally grown oxides. The devices addressed in this paper include n-MOS and p-MOS field-effect transistors (FETs) with oxide-equivalent thicknesses of less than 2 nm.

Introduction

As lateral device dimensions are scaled into the deep-submicron regime to achieve high levels of speed and integration, there must be corresponding decreases in the gate oxide-equivalent thickness, $t_{ox\text{-}eq}$, to maintain current

0018-8646/99/$5.00 © 1999 IBM

levels required for circuit operation. Values of $t_{ox\text{-}eq}$ are calculated from experimentally determined gate capacitance in the accumulation region, C_{ac}, by assuming that the actual dielectric film is equivalent to an oxide film with a static dielectric constant, k_0, of 3.8, so that

$$t_{ox\text{-}eq} = k_0 \varepsilon_0 A / C_{ac}, \quad (1)$$

where ε_0 is the permittivity of free space and A is the area of the capacitor. The value of C_{ac} can be obtained directly from experimental C–V data for thicker films ($t_{ox\text{-}eq}$ >2.5 nm at accumulation voltages >2–3 V) or from a fit to capacitance–voltage data for ultrathin dielectrics when $t_{ox\text{-}eq} < 2.5$ nm.

Selective incorporation of nitrogen into ultrathin silicon dioxide (SiO_2) dielectrics provides improvements in device performance and reliability. A practical definition of the term *ultrathin* consistent with emerging technology requirements is $t_{ox\text{-}eq} < 3$ nm. This is the thickness of an SiO_2 film at which direct tunneling becomes the dominant mechanism for current transport through that film [1, 2]. One objective of gate dielectric engineering is to reduce direct tunneling by increasing the physical thickness of gate dielectrics while maintaining an oxide-equivalent thickness that corresponds to a significantly thinner SiO_2 film. This can be realized by using alternative dielectric materials with dielectric constants higher than SiO_2, as for example silicon nitride or transition-metal oxides such as TiO_2 or Ta_2O_5 [1, 2]. These transition-metal oxides can be incorporated into composite structures by substituting these oxides for the Si_3N_4 layers of the ON or ONO dielectrics of this review. If the transport mechanism at a particular value of $t_{ox\text{-}eq}$ is by direct tunneling, the use of physically thicker films is anticipated to significantly reduce gate leakage current [1, 2].

This paper focuses primarily on stacked ON structures in which the oxide and nitride layers have been deposited by low-temperature remote-plasma-assisted processing [3] and then subjected to rapid thermal annealing (RTA) at ~900°C in nonoxidizing ambients [4]. Defect minimization in the plasma-deposited nitrides is accomplished during RTA by reducing the bonded hydrogen content by a factor of about 2, from levels of 20–30 at.% in the as-deposited films to about 10–15 at.% after a 900°C anneal[1] [5, 6]. This review does not discuss alternative processing approaches such as rapid thermal chemical vapor deposition (RTCVD) [7] or jet vapor deposition (JVD) [8], which have also been used to deposit nitride and/or oxynitride films for stacked gate dielectric structures. Defect reduction in these films is generally achieved by postdeposition annealing in an oxidizing environment, converting the as-deposited nitride films to oxynitride alloys with a relatively high nitrogen content [9]. It is still an open question as to which type of nitride film, the hydrogenated plasma-deposited nitrides or the reoxidized RTCVD or JVD nitrides, will eventually offer the best combination of performance and reliability for aggressively scaled devices with $t_{ox\text{-}eq} < 2$ nm.

Spatially selective incorporation of nitrogen into advanced gate dielectrics 1) reduces defect generation at the Si–SiO_2 interface when N atoms are incorporated at submonolayer-to-monolayer concentrations [10–13]; 2) allows the use of physically thicker stacked dielectric structures when silicon nitride layers are incorporated into the body of the dielectric, as in ONO or ON composites [6, 13, 14]; and 3) reduces B-atom transport out of heavily doped p^+ polycrystalline silicon gate electrodes when nitrided layers are formed at the polycrystalline Si–dielectric interface, or when nitride layers are included in the body of the dielectric, as in ON stacks [15, 16]. This paper presents results from an ongoing research program at North Carolina State University sponsored by the National Science Foundation (NSF), Semiconductor Research Corporation (SRC), and SEMATECH, which has addressed all of these issues. This research has demonstrated separate and independent control of nitrogen incorporation in the different parts of the gate dielectric structure through the use of low-temperature plasma-assisted processing at 300°C combined with low-thermal-budget rapid thermal annealing (RTA), e.g., 30 s at 900°C [6, 10, 13–18].

The next section presents a brief introduction to remote-plasma-assisted processing as it has been applied to the formation of gate dielectrics. The following three sections deal with plasma processes for 1) interface nitridation, 2) top-oxide surface nitridation of oxides, and 3) deposition of silicon nitride thin films. Each section includes a discussion of processing, chemical and structural characterizations of the plasma-processed interfaces and films, and experimental results on test devices that demonstrate the potential advantages of nitrogen incorporation. The final sections of the review focus on stacked ON gate dielectrics with nitrided Si–SiO_2 interfaces, i.e., NON structures, which are emerging as possible alternative composite dielectrics for inclusion in future generations of devices.

Remote-plasma-assisted processing

As stated above, this review is restricted to nitrided gate dielectrics formed by low-temperature remote-plasma processing combined with postdeposition rapid thermal annealing in nonoxidizing environments. It is important to note that the annealing step is crucial. It reduces the hydrogen content of the plasma-deposited nitrides, thereby reducing densities of defects and defect precursors and yielding nitride films with device-quality performance and

[1] V. Misra, Z. Wang, Y. Wu, H. Niimi, G. Lucovsky, J. Wortman, and J. Hauser, submitted to *IEEE Trans. Electron Devices* (1999).

reliability for aggressively scaled complementary metal–oxide–semiconductor (CMOS) devices [5, 6, 10, 14].

Figure 1 is a schematic diagram of a remote-plasma-processing chamber that has been used for Si–SiO_2 interface formation, SiO_2 and silicon nitride (Si_3N_4) film deposition, and SiO_2 top-surface nitridation [3]. Plasma-assisted activation of reactive species has been achieved by radio-frequency (rf) excitation at 13.56 MHz typically at power levels from 15 to 60 W. The coupling of rf power to the plasma is capacitive, with the coil being the hot electrode and the metal plate at the top of the chamber being grounded. Similar chambers have been incorporated into multichamber systems that include additional chambers for surface chemical analysis by Auger electron spectroscopy (AES); rapid thermal annealing (RTA); metal deposition by reactive magnetron sputtering (RMS); and plasma and rapid thermal deposition of polycrystalline silicon for gate electrodes. Gas injection in the remote-plasma chambers is through a plasma excitation inlet tube, and injection within the chamber is through showerhead dispersal rings (see Figure 1). Remote-plasma processing is differentiated from conventional or direct-plasma processing in three ways: 1) It provides selective excitation of source and carrier gases as determined by their point of injection into the chamber, either through the remote-plasma tube or through a showerhead injection ring; 2) the deposition substrate is outside the plasma glow region; and 3) source gases injected from the plasma tube are prevented from backstreaming into the plasma-generation region by gas flow and process pressure [3]. Plasma chambers have been configured for *in situ* process diagnostics, including mass spectrometry (MS) and optical emission spectroscopy (OES) [19]. The implementation of remote-plasma-processing chambers into multichamber cluster systems has made it possible to interrupt plasma-assisted oxidation, nitridation, and/or film deposition and then, without removing the sample from an ultrahigh-vacuum (UHV)-compatible environment, perform on-line chemical analysis by AES. It is also possible to integrate on-line sequences of plasma processing with RTA and metal gate electrode deposition [4]. Even though electron cyclotron resonance (ECR) plasma chambers have structural features in common with the remote-plasma chamber in Figure 1, they are operated at significantly lower pressures [~1 to 20 mTorr (compared to 300 mTorr)] at which all of the injected gases can eventually be subjected to plasma activation. **Table 1** outlines a typical remote-plasma-processing sequence for the formation of device-quality gate oxides. For applications in CMOS devices, it is important to focus on two different parts of the gate dielectric structure: the Si–SiO_2 interface and the bulk dielectric film.

The Si–SiO_2 interfaces for CMOS devices may be formed in different ways:

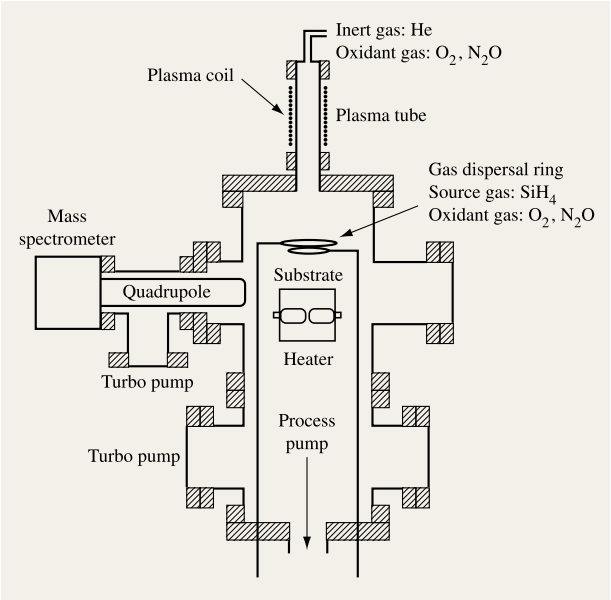

Figure 1

Schematic diagram of remote-plasma-processing chamber.

1. By thermal oxidation of the Si substrate, so that the Si–SiO_2 interface is continuously regenerated and buried below the original surface of the Si wafer.
2. By an ideal deposition process, in which the metallurgical boundary between the Si and the deposited SiO_2 film is maintained at the original surface of the Si.
3. By a real deposition process, in which plasma or thermally generated, chemically active oxygen species (such as O_2^* metastables or O atoms) interact with the Si substrate during deposition so that oxidation takes place at the Si substrate during the initial stages of the deposition and displaces the Si–SiO_2 interface into the bulk of the Si substrate layer [20, 21].
4. By a two-step plasma oxidation/deposition sequence, in which a thin passivating layer of SiO_2 is created during the remote-plasma-assisted oxidation (RPAO) step that forms the interface and prevents further oxidation of the Si substrate during the deposition of an oxide by remote-plasma-enhanced chemical vapor deposition (RPECVD) [4, 22].

The two-step process described last is a focal point of this review; it has been extended to multistep processes which include additional plasma-deposited thin-film layers in the gate dielectric, as in ON and ONO stacks.

The two-step (or multistep) approach provides separate and independent control over interface formation and film

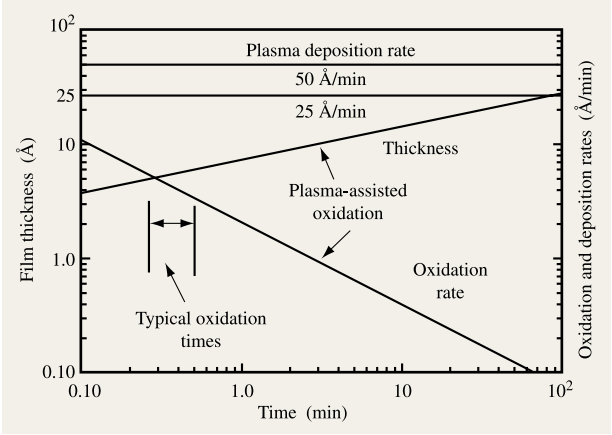

Figure 2

Log–log plots of RPAO oxide thickness and oxidation rates, and RPECVD deposition rates as a function of time.

deposition. The basis for this independence is illustrated in **Figure 2**, which displays rates for interface formation by the RPAO process and for film deposition by RPECVD.

The kinetics for the RPAO process have been determined from experimental studies of oxide layer thickness as a function of oxidation time using on-line AES (see **Figures 3** and **4**) [4]. The oxide thickness, t_{ox}, was determined by analysis of AES as described in Figure 5(a), which is shown later. Using these AES results, it has been demonstrated that the initial stages of the RPAO process follow a power-law time dependence, $t_{ox} = \tau_0 t^{\beta}$, for film thicknesses up to ~2 nm, where t is the oxidation time in minutes, t_{ox} is the oxide thickness in nm, and τ_0 and β are the experimentally determined fitting parameters (Figure 3). For the oxidation of Si(100) using the oxidation process conditions with O_2 as the O-atom source gas in [4, 13] and Table 1, the fitting parameters are given by $\tau_0 \approx 0.7 \pm 0.1$ nm and $\beta \approx 0.28 \pm 0.01$. It is important to note that the power-law form is empirical and that it has also been used to characterize low-temperature thermal oxidation with oxygen-atom partial pressures well below atmospheric levels [23]. The values of β obtained in the low-pressure oxidation studies are essentially the same as those obtained in the RPAO process.

Returning to Figure 2, for typical oxidation times of 15 to 40 s, the oxidation rate after 0.5–0.6 nm of oxide has been formed is ~0.3 to 0.4 nm/min, whereas deposition rates for bulk oxide films are typically higher, ~2.5 to 5 nm/min. In the temperature range between 200°C and 300°C, τ_0 shows a weak temperature dependence (see Figure 3), but β is temperature-independent, displaying different values for different silicon surfaces [0.28 ± 0.01 for Si(100), increasing to 0.30 ± 0.01 for Si(111)]. This suggests that β may scale with the Si surface-atom density. A physical basis for the empirical power-law relationship has been identified through theoretical studies of the initial stages of silicon oxidation [24]. This study has shown that chemical attack of Si surface and subsurface atoms by active oxygen species, such as atoms or O_2^* metastables, initially proceeds very rapidly owing to the surface-bonding geometry of crystalline Si. The Si surface dangling bonds and the first layer of back-bonds to these surface atoms are the most susceptible to chemical attack, accounting for the relatively short times required to form

Table 1 Process sequence for forming device-quality stacked gate oxide dielectrics [3].

Process step	*Processing conditions*	*Processing results*
RPAO	Substrate temperature: 300°C Process pressure: 300 mTorr Plasma-excited mixture: He/O_2 (200 sccm He, 20 sccm O_2) Time: ~15–30 s	*In situ* substrate cleaning (reduces C and F levels) Forms Si–SiO_2 interface Grows passivating oxide: ~0.5 nm
RPECVD	Substrate temperature: 300°C Process pressure: 300 mTorr Plasma-excited mixture: He/O_2 (200 sccm He/20 sccm O_2) Downstream mixture: He/SiH_4 (20 sccm He/0.4 sccm SiH_4)	Forms body of dielectric film Deposition rate: 2.5–5.0 nm/min Stoichiometric SiO_2 No IR-detectable Si–H or Si–OH Low Si–OH ($\ll$5 at.% H)
RTA	Temperature: 900°C Time: ~30 s Low-pressure or atmospheric inert gas ambient (Ar)	Reduces oxidation-induced suboxide bonding at Si–SiO_2 interface Promotes densification of oxide films Reduces bonded H (mostly in nitrides)

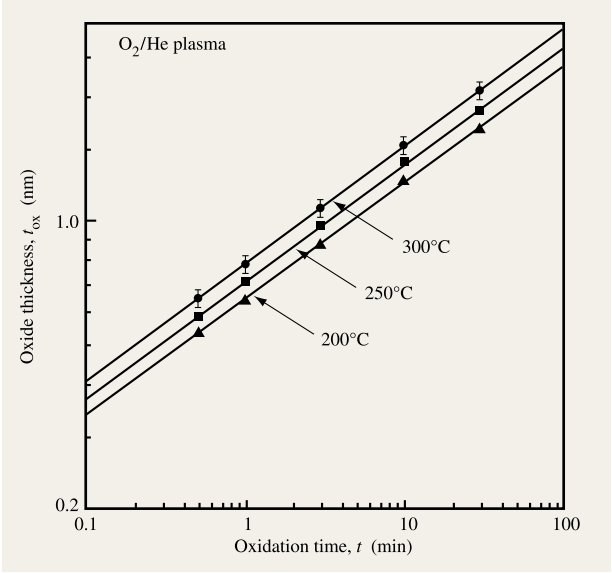

Figure 3

Log–log plots of oxide thickness (t_{ox}) as a function of time (t) for RPAO of Si using O_2 as the O-atom source gas. The straight lines connecting the data points represent a power-law dependence, i.e., $t_{ox} = \tau_0 t^{\beta}$, where τ_0 and β are fitting parameters.

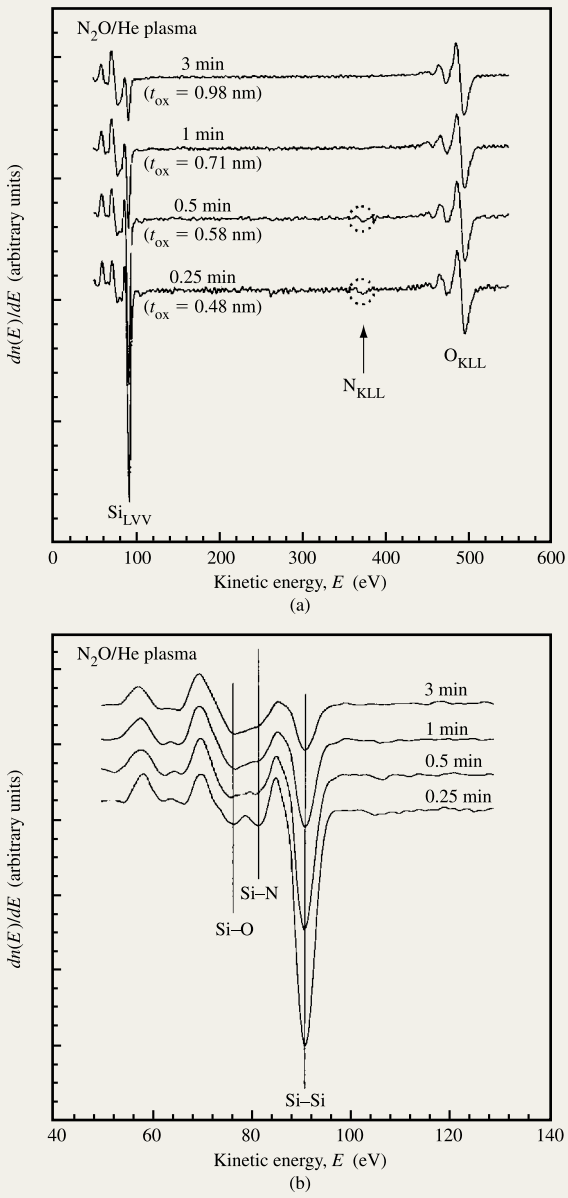

Figure 4

Time evolution of differential AES spectra for the N_2O RPAO process for oxidation times from 0.25 min (15 s) to 3 min: (a) from 0 eV to 600 eV; (b) in the Si_{LVV} regime from 40 eV to 140 eV.

approximately two molecular layers with Si–O bonds, or equivalently, 0.5 to 0.6 nm of oxide.

The slowing down of the oxidation rate in Figure 2 to ~0.3–0.4 nm/min after formation of ~0.5–0.6 nm of oxide means that during plasma-enhanced deposition at rates of ~2.5–5 nm/min, plasma-activated O species are consumed faster by deposition reactions with SiH_4 than by continued oxidation at the buried Si–SiO_2 interface. Experimental results have demonstrated that the two-step remote-plasma approach achieves independent control over interface formation and film deposition [22]. For example, substrate temperatures of 400°C for deposition of SiO_2 directly onto H-terminated Si(100) surfaces by RPECVD result in high densities of interfacial defects (D_{it}) [22]. However, formation of 0.5–0.6 nm of a plasma-grown oxide by RPAO prior to oxide deposition has demonstrated no significant increases in D_{it}, even when deposition temperatures are increased to as high as 450°C.

The predeposition oxidation step also provides *in situ* surface cleaning, e.g., reducing residual levels of carbon and fluorine interface contamination to the low 10^{12}-cm^{-2} range. This is presumed to occur through direct oxidation of the surface contaminants to volatile species such as CO [10].

This two-step process sequence, RPAO followed by RPECVD, forms the conceptual basis for all of the other processing approaches presented in this review. For example, interface plasma nitridation has been accomplished in two different ways: 1) by performing the RPAO process using N_2O or N_2O/O_2 mixtures in place of O_2 [10]; and 2) by performing the RPAO process using O_2 and then subjecting the interface and ultrathin passivating SiO_2 film to a plasma-assisted nitridation process using

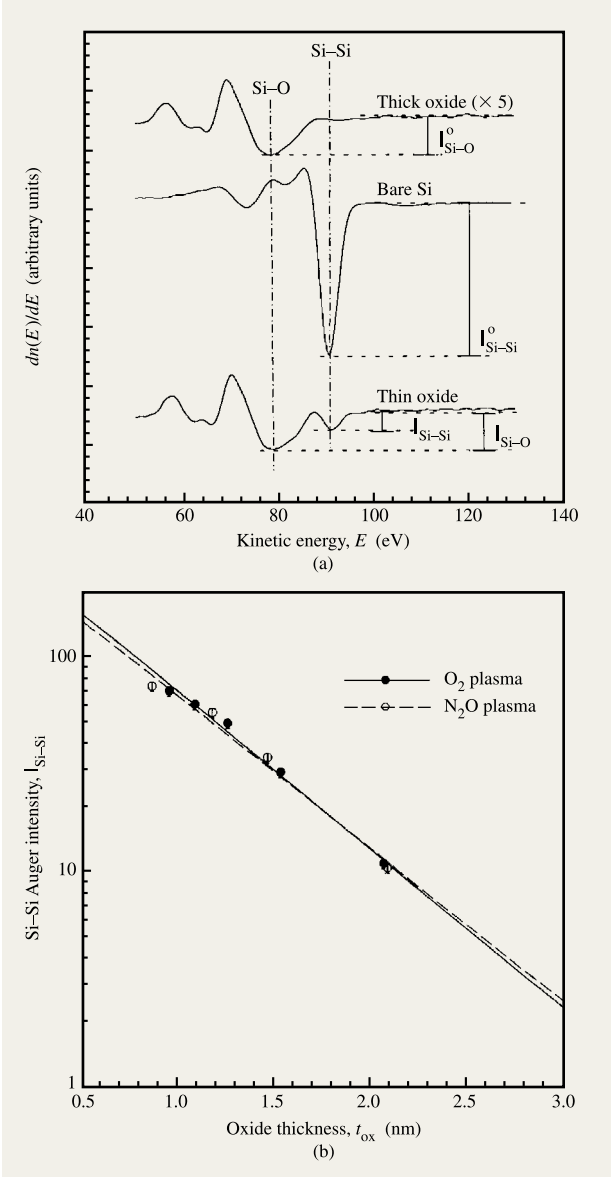

Figure 5

Calibration of oxide thickness from Si_{LVV} spectral features: (a) AES traces for a thick oxide, a bare Si wafer, and a thin oxide which displays both Si–O oxide and Si–Si substrate features; (b) self-consistency check in which the Si–Si substrate signal is plotted as a function of the calculated oxide thickness. I^o_{Si-Si} and I^o_{Si-O} are respectively the Auger amplitudes for the reference Si and Si–O signals from the Si substrate and thick oxide film; I_{Si-Si} and I_{Si-O} are the Auger amplitudes for the Si and Si–O signals from the thin film being characterized. If λ is the escape length of Auger electrons at the Si_{LVV} energy of ~80–90 eV, the film thickness t_{ox} is given by $t_{ox} = \lambda \ln[(I^o_{Si-Si} \times I_{Si-O})/(I_{Si-Si} \times I^o_{Si-O})]$.

species extracted from a remote N_2/He plasma [4]. Following either of these steps, the deposited gate dielectric can be an SiO_2 film [10, 25], an ONO stack [14], or an ON stack [16–18]. If the gate dielectric film is an oxide, top-surface nitridation can be accomplished by a downstream plasma-assisted process [15]. Finally, to fabricate device-quality gate dielectrics, it is necessary to include a postdeposition rapid thermal anneal at 900°C for at least 30 s, or an equivalent thermal exposure at temperatures of at least 900°C [4, 26]. This provides chemical and structural relaxation of the oxide and nitride films, as well as chemical and structural relaxation of the Si–SiO_2 and nitrided Si–SiO_2 interfaces.

Prior to remote plasma processing, wafers are either subjected to conventional RCA cleanings or simply etched in dilute HF. Sacrificial oxides ~10 nm thick are grown by conventional thermal oxidation at 800°C–900°C, and then annealed at 900°C in a nonoxidizing environment to minimize interfacial suboxide bonding [26]. Immediately before introduction into the plasma-processing chamber, the sacrificial oxide layer is removed by rinsing in dilute (~1%) HF. No additional *in situ* cleanings are performed prior to the initiation of the RPAO process. Both the RPAO and RPECVD processes are relatively insensitive to processing temperatures between 200°C and 300°C [22]; however, a processing temperature of 300°C provides a convenient way of minimizing OH incorporation into the RPECVD SiO_2 films [3]. The use of a 300-mTorr process pressure, combined with the high flow rates of reactant and He diluent gases through the plasma tube (~200 sccm), prevents backstreaming of downstream-injected source gases such as SiH_4 into the plasma-generation region [3]; it also provides effective gas-phase mixing of process gases, which is important in obtaining uniformly thick depositions over the entire substrate wafer surface. A top-surface nitridation process described below has been performed at a reduced process pressure of 100 mTorr in order to allow the plasma afterglow to penetrate into the processing chamber and speed up the process [15]. Because of this reduced pressure, the top-surface nitridation cannot be characterized as a strictly remote-plasma process, even though the substrate is downstream from the region of direct-plasma excitation. On the other hand, depositions of nitride films at 200 mTorr process pressures are effectively remote in the context of the distinctions made above.

Interfacial nitridation

- *RPAO in N_2O and N_2O/O_2 mixtures*

Interface nitridation was first demonstrated during an RPAO process at 300°C that uses excited species from a remote He/N_2O plasma [10]. Reaction pathways for this process have been discussed [4, 19]. Figures 4(a) and 4(b) indicate the time evolution of differential AES spectra for this process, displaying Si_{LVV}, N_{KLL}, and O_{KLL} features.

Changes in Si_{LVV} spectra coupled with changes in N_{KLL} and O_{KLL} features have established that the N atoms are incorporated in the immediate vicinity of the Si–SiO_2 interface rather than uniformly in the RPAO passivating oxide film. Three features in the Si_{LVV} region [Figure 4(b)] have been used to establish interfacial nitridation: the Si–O feature at ~76 eV, the Si–N feature at ~83 eV, and the Si–Si substrate feature at ~91 eV. The 83-eV feature also includes a contribution from Si–suboxide bonding [19, 26]. The 76-eV feature has been assigned to Si atoms with four O-neighbors; the 83-eV feature has contributions from Si atoms bonded to fewer than four O atoms and more than one Si atom, as well as Si bonded to at least one N atom; and the 91-eV feature is assigned to Si atoms with four Si-atom neighbors. The oxide layer thickness has been obtained to ±0.1 nm from the relative signal strengths of the 76-eV and 91-eV features using a nominal electron escape depth of 0.6 nm [**Figure 5(a)**] [22]. This procedure is internally consistent in the sense that a plot of the Si–Si substrate signal as a function of the calculated oxide-layer thickness is fitted by an exponential function with a characteristic decay length equal to the electron escape depth used to determine the film thickness [**Figure 5(b)**]. The Si–N Si_{LVV} and the N_{KLL} features are present at the initial stages of the oxidation process, but are not observed once t_{ox} is increased to about two to three times the electron escape depth of ~0.6 nm. This means that the N atoms are either buried at the interface or removed from the film during the continuation of the RPAO process. Secondary ion mass spectrometry (SIMS) [10] and X-ray photoelectron spectroscopy (XPS) [27] results demonstrate that N atoms are not eliminated during the continuation of the RPAO process. By combining the AES results with SIMS data [10], XPS data [27], and an analysis of optical second-harmonic generation (SHG) data [4, 28], it has been established that the N_2O RPAO process produces essentially one monolayer of interfacial nitridation. The areal concentration of N atoms at the Si(100) interface is $7 \pm 1 \times 10^{14}$ cm^{-2}, which is approximately equal to the Si atom density at that surface. In addition, it has been demonstrated by SIMS that substituting plasma-excited N_2O/O_2/He mixtures for a N_2O/He mixture results in submonolayer interface nitridation, with the N concentration decreasing as the N_2O fraction of the N_2O/O_2/He source-gas mixture is reduced [10].

- *Post-RPAO nitridation*

The post-RPAO nitridation process uses two separate interface formation steps: first, 300°C remote-plasma-assisted oxidation of a Si surface to form the interface and a superficial oxide layer ~0.5–0.6 nm thick, followed by a remotely activated N_2/He plasma nitridation step to incorporate nitrogen at the Si–SiO_2 interface [4, 13]. The active species for the nitridation process have been identified as neutral nitrogen metastable molecules (N_2^*) [19]. The amount of nitrogen incorporated at the interface is controlled by varying the N_2 plasma exposure time. On-line AES was performed to quantify the initial stages of oxidation of the Si surface and nitridation of the superficially thin oxide layer. SIMS analyses using CsN^+ and SiN^- ions for depth profiling of the interfacial nitrogen were performed on structures which included a deposited oxide layer about 5 nm thick on top of the plasma-processed interface. In addition, nuclear reaction analysis (NRA) was also performed to determine the N concentration.[2]

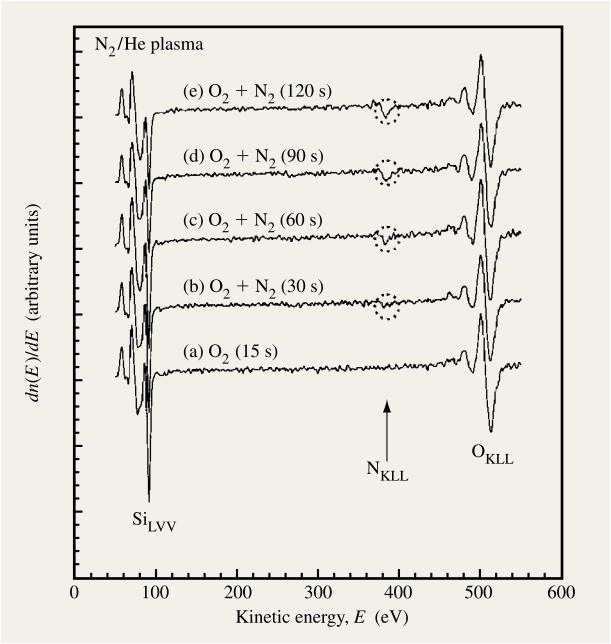

Figure 6

Time evolution of differential AES spectra that combines the O_2 RPAO process (15-s exposure) with postoxidation nitridation from an N_2/He plasma for exposure times ranging from 30 s to 120 s.

Figure 6 shows on-line AES data after (a) a 15-s O_2 plasma exposure. The initial RPAO provides an ~0.5–0.6-nm-thick oxide on the silicon surface. When followed by N_2/He (respective flow rates are 60 and 160 sccm) plasma exposures of (b) 30 s, (c) 60 s, (d) 90 s, and (e) 120 s, this step results in nitridation of the Si–SiO_2 interface. The intensity of the nitrogen KLL peak (N_{KLL}) at ~375 eV increases with increasing exposure time to the nitrogen plasma, demonstrating that longer exposures result in an

[2] D. A. Buchanan, 1998.

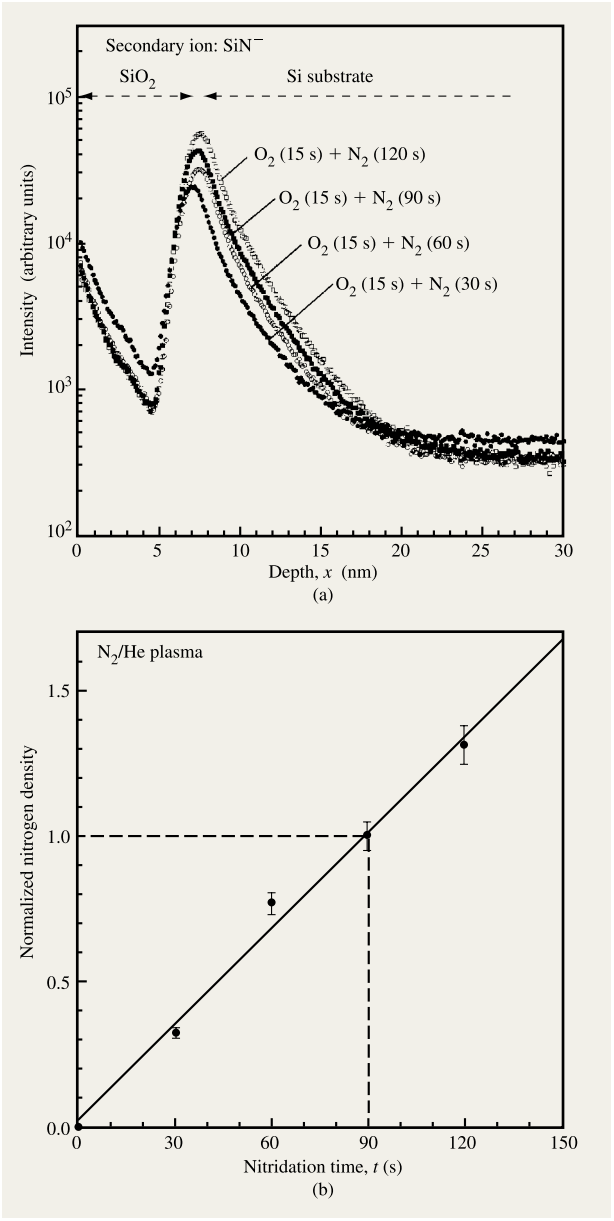

Figure 7

(a) SIMS depth profiles for the post-RPAO nitridation process. (b) Nitrogen concentration vs. nitridation time. The value of one (1) corresponds to approximately one monolayer of interfacial nitrogen atoms ($7 \pm 1 \times 10^{14}$ cm^{-2}).

increasing nitrogen incorporation at either the Si–SiO_2 interface or in the bulk of the 0.5–0.6-nm oxide film.

The following set of experiments has been used to demonstrate that the nitrogen is incorporated at the Si–SiO_2 interface, and not in the bulk of the thin oxide film, so that this process and the RPAO process using N_2O as the N-atom source gas achieve essentially the same result. First, the nitrogen content at the interface was determined by secondary ion mass spectroscopy (SIMS), using both CsN^+ and SiN^- ions to monitor the nitrogen content. Since the data are essentially the same, only the SiN^- ion data are shown. **Figure 7(a)** indicates a depth profile obtained from the SIMS data in the usual way by combining the raw data with a determination of the average sputtering rate. **Figure 7(b)** presents normalized areal N-atom densities as a function of time as obtained by integration of the nitrogen depth profile of Figure 7(a). Figure 7(b) indicates that the interfacial nitrogen concentration is essentially a linear function of the nitrogen plasma exposure time. In addition, a calibration of the SIMS data of Figure 7(a) using an ion-implanted calibration standard, as well as comparisons of the SIMS data of this experiment with those in [10], indicates that approximately one monolayer of nitrogen is incorporated at a plasma exposure time of 90 s. This has been further confirmed by using an NRA technique; analysis of the SIMS data gives a N content of $7 \pm 1 \times 10^{14}$ cm^{-2}, and analysis of the NRA data[3] gives a concentration of $\sim 8 \pm 1 \times 10^{14}$ cm^{-2}.

Figure 8 contains a plot of the ratio of the amplitudes of the N_{KLL} feature and the substrate Si_{LVV} feature at 91 eV vs. nitridation time. A plot of the N_{KLL} strength vs. time in Figure 8 shows a sublinear dependence (slope < 1) due to AES signal losses associated with increasing film thickness during the nitridation process. Normalizing the N_{KLL} signal by the Si_{KLL} substrate signal at 91 eV yields an approximately linear dependence on time (slope ≈ 1). Normalizing the N_{KLL} signal amplitude through division by the Si_{LVV} substrate signal takes into account reductions in signal amplitude due to the passage of the N_{KLL} electrons through the oxide film. The approximate linearity of this plot suggests that the N atoms are not incorporated uniformly throughout the SiO_2 film, but instead are localized closer to the Si–SiO_2 interface. The plasma-nitrided interfaces have also been studied by angle-resolved X-ray photoelectron spectroscopy (ARXPS),[4] and these experiments have confirmed that the N atoms are not uniformly distributed in the bulk of the oxide, but rather are incorporated closer to the Si–SiO_2 interface. Neither the AES data of Figure 7, the reduced AES data in Figure 8, nor the ARXPS data establish that N atoms are located at the metallurgical Si substrate dielectric interface. However, they are consistent with a nonuniform incorporation of nitrogen atoms, with their density being higher at the interface than in the bulk of the ultrathin oxide film.

[3] D. A. Buchanan, 1998.
[4] J. M. White, 1998.

• *Electrical performance*

Two important aspects of device reliability and performance have been directly correlated with interfacial nitridation: improvement of hot-carrier and current-stress reliability, and reduction of direct and Fowler–Nordheim (F–N) tunneling currents. Improvements in device reliability have been discussed in [10] and are illustrated by the data included in **Table 2** for FETs with a channel length of ~500 nm and a gate oxide thickness of ~5.5 nm. Nitrided interfaces for these devices were formed directly by RPAO using the N_2O source-gas process. For these devices the dielectric was a stacked structure comprising ~0.5–0.6 nm oxide with a nitrided interface formed by the RPAO process, and the remaining ~5 nm of oxide was prepared by RPECVD using N_2O as the oxygen-atom source gas and silane (SiH_4) as the silicon-atom source gas. The formation of RPECVD oxides using N_2O does not incorporate N atoms in the bulk oxide layer, as determined by SIMS analysis with sensitivity limits of $\sim 10^{17}$ cm^{-3}. There was no separate RTA step in the processing, so that interface and bulk film structural and chemical relaxations occurred during the $POCl_3$ doping process/dopant activation anneal for the polycrystalline Si gate electrode, and source and drain contacts at ~950°C for ~30 min in a nonoxidizing ambient [10]. As shown in Table 2, the test data demonstrate that device reliability, as monitored by hot-electron electrical stress-induced changes in the threshold voltage ($\Delta V_{th}/V_{th}$, where V_{th} is the threshold voltage before stressing) and the peak transconductance ($\Delta g_m/g_m$, where g_m is the peak transconductance before stressing) in devices with interface nitridation are smaller when referenced to a control device with a thermally grown oxide, and larger with respect to plasma-processed devices without interface nitridation. Additionally, as shown in **Figure 9**, for partially nitrided interfaces, the stress-induced changes decrease as the degree of interfacial nitridation increases. The partially nitrided interfaces were prepared by an RPAO process that used O_2/N_2O/He as the plasma-excited source-gas mixture. Partial nitridation was verified directly by analysis of SIMS measurements [10].

Correlations between reductions in direct and Fowler–Nordheim tunneling currents and increasing

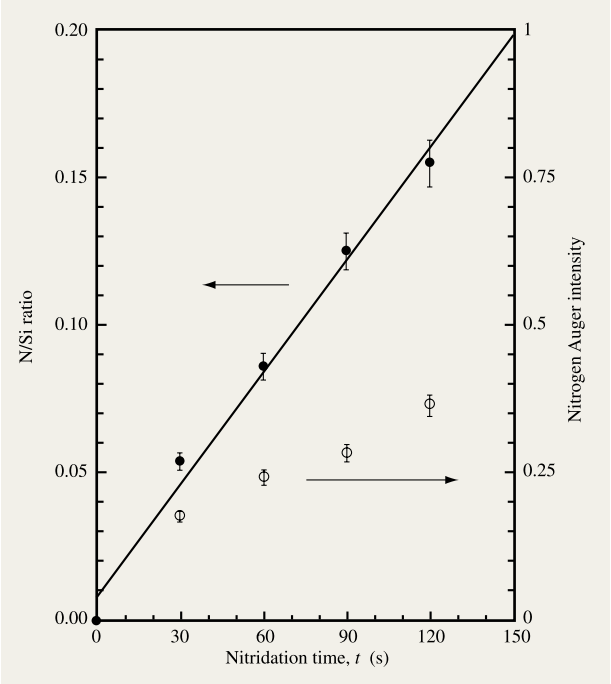

Figure 8

Amplitude of the N_{KLL} peak, and the ratio of the amplitude of the N_{KLL} peak divided by the amplitude of the substrate Si_{LVV} peak, each as a function of nitridation time.

Table 2 Electrical properties of FETs: Interfaces by N_2O RPAO.

Initial performance of FETs—threshold voltage (V_{th}) and mobility [10]		
Interface preparation	*Threshold voltage* (V)	*Peak mobility* (cm^2 V^{-1} s^{-1})
Thermal oxidation 900°C	0.33 ± 0.03	450 ± 5
Plasma oxidation 300°C + RTA	0.38 ± 0.03	460 ± 5
Plasma nitridation 300°C + RTA	0.41 ± 0.03	460 ± 5
Reliability of short-channel FETs: 5.5-nm gate oxide—0.5-μm channel length [10]		
Interface preparation	$\Delta g_m/g_m$ (%)	$\Delta V_{th}/V_{th}$ (%)
Thermal oxidation 900°C	−16.2	+9.9
Plasma oxidation 300°C + RTA	−20.5	+14
Plasma nitridation 300°C + RTA	−12.5	+6

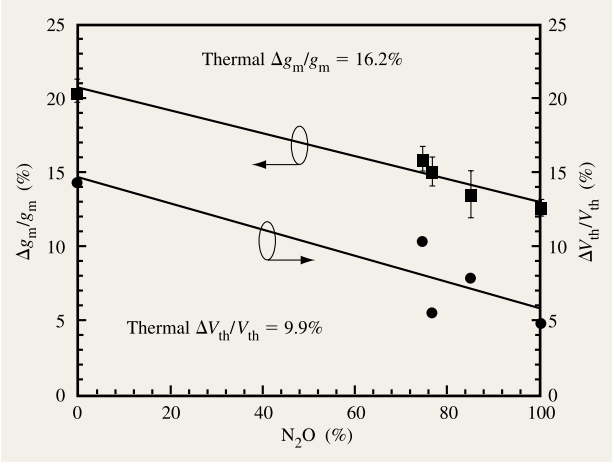

Figure 9

Stress-induced changes in the threshold voltage ($\Delta V_{th}/V_{th}$) and the peak transconductance ($\Delta g_m/g_m$) as a function of the relative percentage of N_2O in the N_2O/O_2 mixtures used in the RPAO process. A value of 100% corresponds to monolayer interface nitridation.

interfacial nitridation are shown in **Figure 10** for nitrided interfaces prepared by the alternative two-step interface nitridation process discussed above, i.e., RPAO using O_2/He, followed by plasma-assisted nitridation using N_2/He [13]. The post-RPAO nitridation process was used in the fabrication of MOS capacitors with oxide thicknesses ranging from 5 nm to 2 nm on n-type and p-type Si(100). After removal of sacrificial thermal oxide layers, the process sequence included three 300°C plasma steps: 1) RPAO using the O_2/He plasma excitation; 2) interface nitridation using the N_2/He plasma process; and 3) RPECVD oxide deposition using plasma-excited O_2/He and downstream-injected SiH_4 as the process gases. The devices fabricated on the n-type Si had either Al or n^+ polycrystalline-Si gate electrodes, and those fabricated on p-type Si had Al gate electrodes. The oxide layer thickness was estimated from capacitance–voltage (C–V) data in the accumulation capacitance regime at applied voltages ≈2–3 V [Equation (1)]. Figure 10(a) includes a series of current-density–gate-voltage (J–V) traces for substrate injection [n-Si(100)] for a 4.5-nm-thick oxide over a wide range of interface nitridation. Figure 10(b) includes J–V traces for substrate injection for thinner oxide films (2 nm and 3 nm), but only for monolayer-nitrided ($\sim 7 \pm 1 \times 10^{14}$ N atoms/cm^2) and non-nitrided interfaces, i.e., the 90-s exposure. These traces demonstrate the effectiveness of interfacial nitrogen in reducing tunneling current in both the Fowler–Nordheim and direct-tunneling regimes. As the amount of nitrogen at the interface is increased, the flat-band voltage, as determined from C–V data, remains essentially constant with increasing interface nitridation (see **Table 3**). This establishes that the reduction in tunneling current is not due to a flat-band voltage shift owing to fixed charge at the Si–SiO_2 interface derived from incorporation of interfacial nitrogen. However, the decrease in tunneling current with increasing

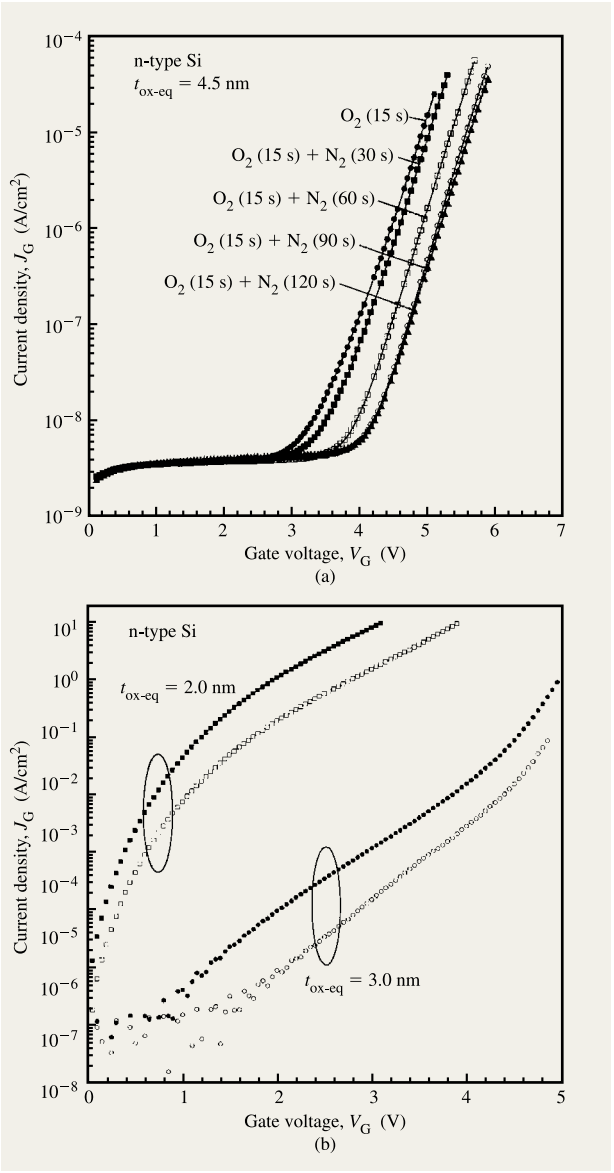

Figure 10

J–V traces for substrate injection from lightly doped n-type Si(100) for capacitors (Al gate electrodes) with interfaces nitrided by the post-RPAO nitridation process: (a) $t_{ox} \approx 4.5$ nm as a function of nitridation time; (b) $t_{ox} \approx 3$ and 2 nm for a 90-s nitridation time. For (b), the solid symbols are for non-nitrided interfaces and the open symbols are for interfaces with approximately one monolayer of nitrogen.

interfacial nitridation is systematic and significant in magnitude. For example, similar decreases in tunneling current have been observed as a function of increasing interface nitridation using p-type substrates and gate injection (**Figure 11**). The oxide thicknesses range from direct tunneling for the 2.0-nm oxides to Fowler–Nordheim tunneling for the 5.0-nm oxides, with comparable reductions in current with increasing nitridation.

Top-surface nitridation

- *Processing and characterization*

Top-surface plasma-assisted nitridations were first performed with substrates at ambient temperature (~23°C) and at 300°C on thermally grown SiO_2 films in a thickness range from 5 to 10 nm [15]. A remote He/N_2 discharge was initiated at a process pressure of 100 mTorr (0.1 Torr) which allows the plasma afterglow to penetrate into the processing chamber so that this process is not remote in the context of definitions presented above. It is important to differentiate this process from the interface nitridation process discussed above. Interfacial nitridation is performed at process pressure of 300 mTorr, and the active species that promote interfacial nitridation are N_2^* metastables [19]. In contrast, when the process pressure is reduced to 100 mTorr, and the plasma afterglow extends to the substrate, atomic nitrogen as well as charged species, e.g., N_2^+, are present in higher densities [19], and these species are responsible for promoting the top-surface nitridation discussed below. Following nitridation, rapid thermal annealing (RTA) at 900°C in N_2 or N_2O was performed in the remote-plasma-processing chamber of the multichamber system at a pressure of 0.5 Torr in an Ar ambient for a time of approximately 40 s. On-line AES was performed in the analysis chamber at intermediate stages of processing. Angle-resolved XPS (ARXPS) was performed *ex situ* using an Al–Kα X-ray source (1.486 keV) in both normal (detector at ~80° with respect to sample surface) and glancing-angle modes (detector at approximately 80° with respect to the normal to the sample surface); see **Figure 12**. SIMS analyses were performed *ex situ* using Cs^+ ions and detecting $CsSi^+$, CsN^+, and CsO^+ species for analysis [15].

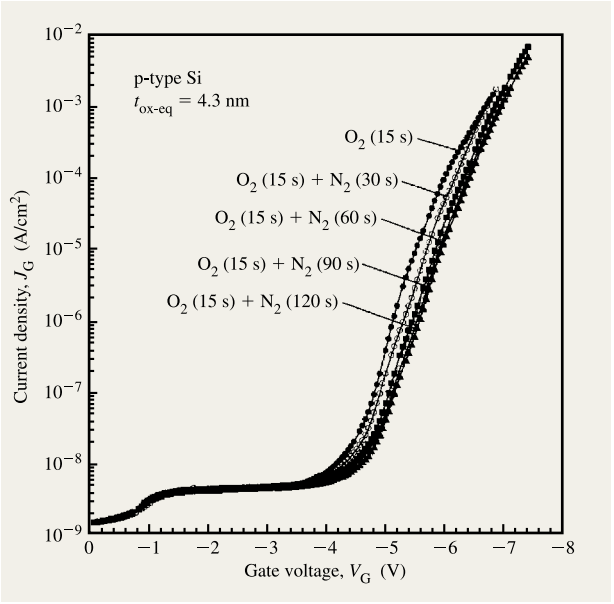

Figure 11

J–V traces for substrate injection from lightly doped p-type Si(100) for capacitors (Al gate electrodes) with interfaces nitrided by the post-RPAO nitridation process as a function of interface nitridation time.

Figure 13 tracks top-surface nitrogen concentrations by displaying AES spectra for a clean thermal oxide surface (~5 nm thick) that indicates only Si_{LVV} and O_{KLL} features (Curve A); following a 23°C, 50-min He/N_2 plasma treatment (Curve B); and following a 300°C, 50-min He/N_2 plasma treatment (Curve C). The surfaces exposed to the N_2 plasma treatments display an additional N_{KLL} feature with respect to Curve A, demonstrating nitridation of the oxide surface. AES spectra in the exploded view of the Si_{LVV} region show a shift of a feature in that manifold from a position indicative of an oxide surface (~76 eV) to a position indicative of a nitrided surface (~81 eV). Reduction of the AES data of **Figure 14** is used to determine top-surface nitrogen incorporation vs. time in

Table 3 Thickness and flat-band voltages determined from *C–V* analysis for ~4.5-nm devices of Figures 10(b) and 11.

Substrate conductivity: p-type (Figure 10(b))					
Nitridation time (s)	0	30	60	90	120
Thickness (±0.05 nm)	4.48	4.47	4.49	4.48	4.47
Flat-band voltage (±0.03 V)	0.29	0.28	0.28	0.27	0.27
Substrate conductivity: n-type [Figure 11]					
Nitridation time (s)	0	30	60	90	120
Thickness (±0.05 nm)	4.29	4.28	4.30	4.29	4.28
Flat-band voltage (±0.03 V)	−0.67	−0.68	−0.70	−0.71	−0.72

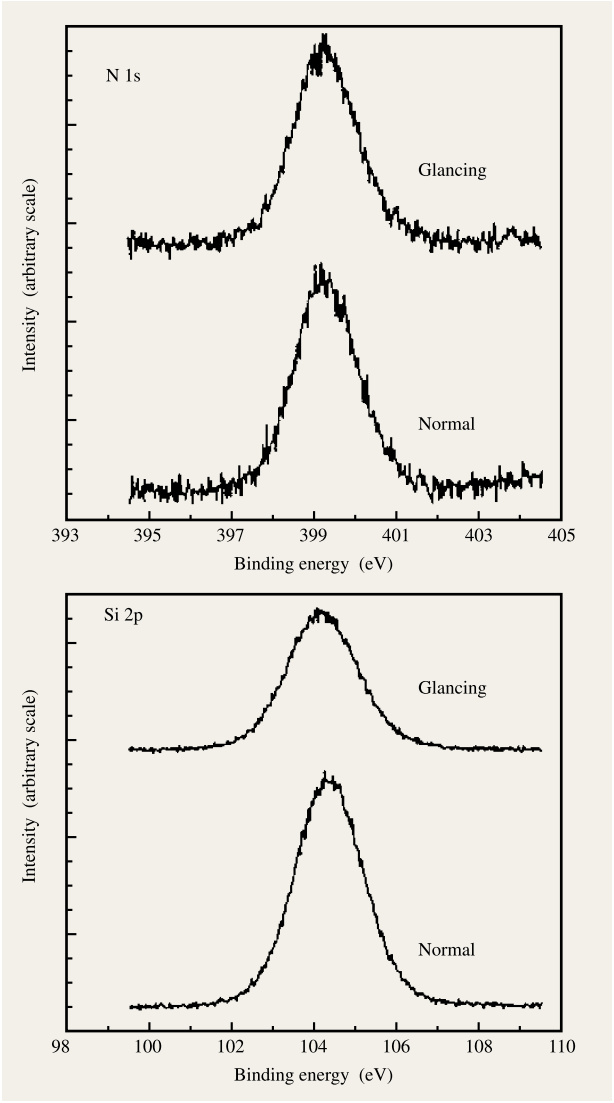

Figure 12

Normal and glancing-angle XPS spectra of a Si surface exposed to plasma-assisted nitridation for 50 min at 300°C. Since the N 1s feature is unchanged, the N atoms are localized in the vicinity of the top surface. The grazing-angle spectra are obtained with an analyzer at an angle of 80° with respect to the normal to the film surface.

Figure 15 for plasma processing at 23°C and 300°C. The reduced data demonstrate that the nitridation efficiency is greater at 300°C than at 23°C. In addition, the nitrogen content tends to saturate with increasing exposure time more quickly at 23°C than at 300°C. High-resolution ARXPS spectra in normal- and glancing-angle modes of the Si 2p and N 1s features have been obtained (Figure 13). The observation of a higher N/Si ratio for the glancing-angle mode establishes that the N atoms are localized at the top surface of the film, rather than being uniformly distributed in the bulk of the oxide layer [15]. The AES analysis gives a surface nitrogen concentration of the order of 10–20 at.%. SIMS analysis shows that the nitrogen is near the surface, with areal densities of 3×10^{14} cm^{-2} and 5×10^{14} cm^{-2}, respectively, for 10-min and 50-min exposures at 300°C. The SIMS and AES results

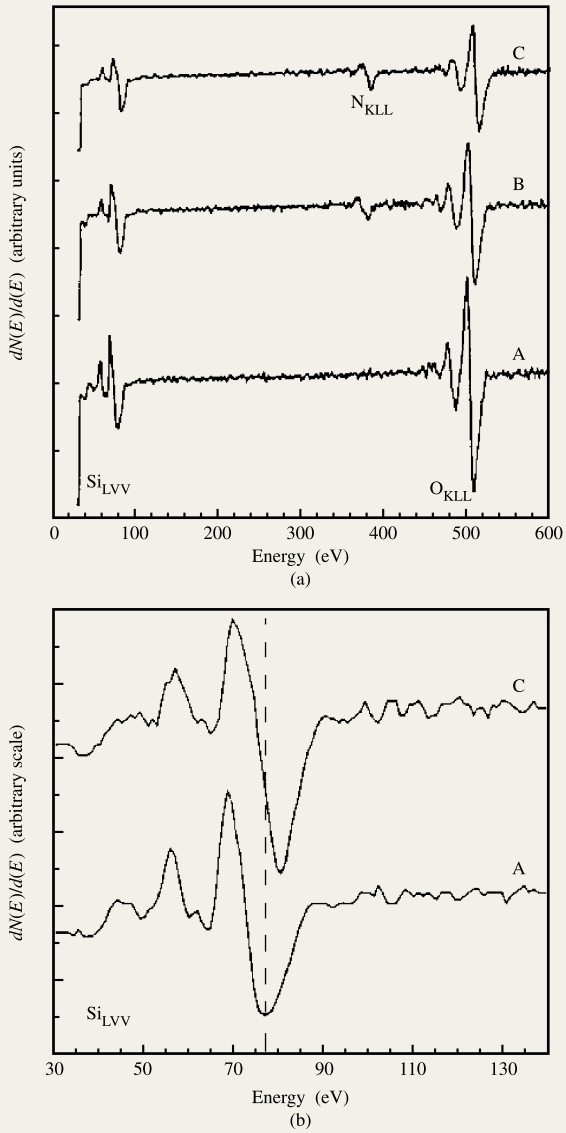

Figure 13

(a) Differential AES spectra: Curve A, for a clean thermal oxide surface; Curve B, following a 23°C, 50-min He/N_2 plasma treatment; Curve C, following a 300°C, 50-min He/N_2 plasma treatment. (b) Exploded view of the Si_{LVV} feature indicating a shift from Si–O bonding to Si–N bonding after the nitridation process.

give essentially the same atomic fractions, establishing that the nitrogen is confined with ~1 nm of the surface region, consistent with the ARXPS results. More recent studies [29], based on a reassessment of SIMS relative sensitivity factors,[5] have shown that the nitrogen concentrations are higher by factors of at least 2 to 3, providing an excellent barrier to boron transport out of heavily doped p^+ polycrystalline gate electrodes in p-MOS devices during dopant activation anneals.

- *Electrical performance in p-MOS devices*

A more efficient variation of the top-surface nitridation process described above was implemented by Hattangady and co-workers at Texas Instruments [15]. Their electrical test data on p-MOS devices with B-doped p^+ polycrystalline silicon gate electrodes showed that plasma-initiated top-surface nitridation of oxides was effective in reducing or eliminating B-atom out-diffusion from the p^+ gate electrodes.

Electrical characterizations have also been discussed in [29], using yet another improved top-surface nitridation process, which has yielded results complementing those of [15]. These studies used tunneling *J–V* data to demonstrate the effectiveness of top-surface nitridation in suppressing B-atom transport out of p^+ polycrystalline silicon gate electrodes. **Figure 16** shows *J–V* data for two p-MOS capacitors: one with top-surface nitridation and a non-nitrided interface, and the second with top-surface nitridation and a monolayer nitrided interface. The flat-band voltages, as determined by *C–V* analysis, are the same to ±0.05 V, indicating that B atoms have not been transported to the Si–SiO_2 interface in either of these devices. The reduction in tunneling current in the device with the nitrided interface is similar to that shown in Figures 10(a), 10(b), and 11, and is characteristic of the previously described interface plasma-assisted nitridation process. The reduction in tunneling in the devices with nitrided interfaces is independent of $t_{ox\text{-}eq}$, including both the direct and Fowler–Nordheim tunneling regimes, and of substrate or gate injection. In the spirit of the WKB approximation [1, 2], they are attributed to different interfacial barriers.[6]

Bulk nitride layers

- *Nitride deposition and chemical and structural characterizations*

Two different RPECVD processes have been used for deposition of bulk nitride films: one process used NH_3 as the N-atom source gas [14], and the other used N_2 [6]. Both processes resulted in heavily hydrogenated films as deposited at 300°C with a bonded hydrogen concentration,

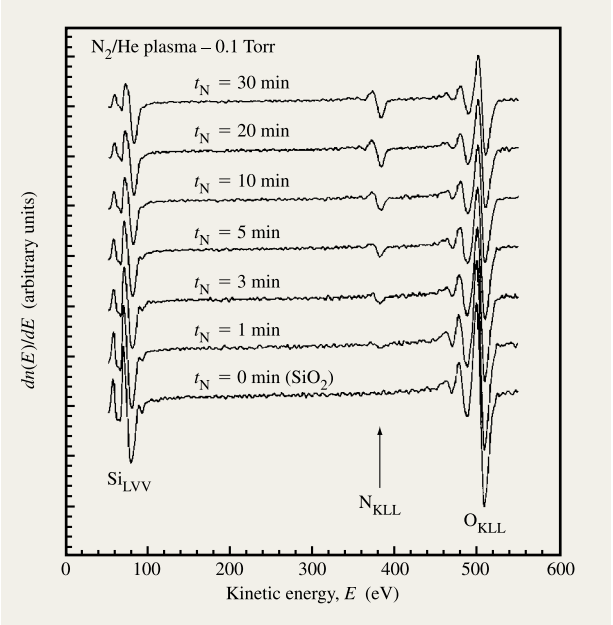

Figure 14

Differential AES spectra for top-surface nitridation at 0.1 Torr of an ~5-nm oxide for nitrogen process times from 1 min to 30 min at a substrate temperature of 300°C.

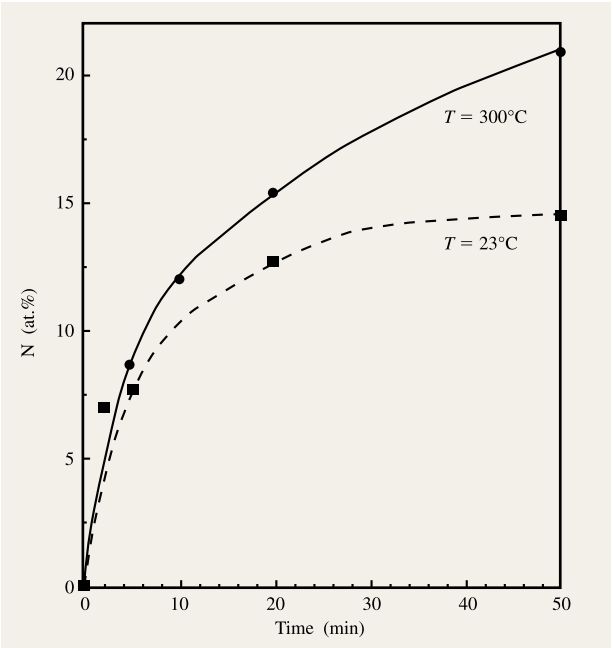

Figure 15

N-atom concentration vs. time for the nitridation of Si at 23°C and 300°C.

[5] Evans East, 1998.
[6] G. Lucovsky and H.-Y. Yang, 1999.

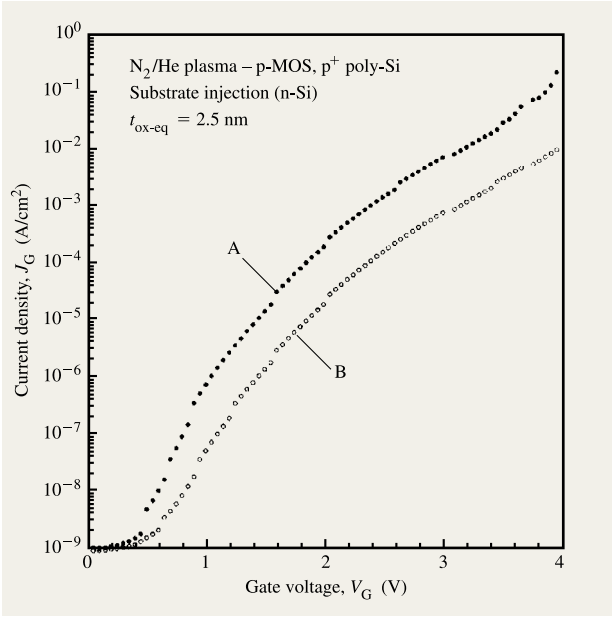

Figure 16

J–V data for substrate injection for two p-MOS capacitors with $t_{ox\text{-}eq}$ = 2.5 nm: Curve A, ON structure with top-surface nitridation and non-nitrided interface (solid symbols); Curve B, NON structure with top-surface nitridation and a monolayer nitrided interface (open symbols). The flat-band voltages determined by *C–V* analysis are the same to ±0.05 V, indicating that B atoms from the p^+ polycrystalline gate electrodes have not been transported to the Si–SiO_2 interface in either of these devices.

[H], of ~20–30 at.%, and both processes were used to produce nitride films for p-MOS devices. The discussion below of chemical bonding changes between initial deposition and after a 900°C RTA emphasizes films produced from the NH_3 source gas; the films produced from the N_2 process show a similar behavior with respect to postdeposition annealing and subsequent device performance [17].

The hydrogenated plasma-deposited nitrides grown from NH_3 and SiH_4 have been characterized by infrared (IR) absorption measurements [5]. IR spectra indicated features associated with 1) the Si–N asymmetric bond-stretching vibration at ~850 cm^{-1}; 2) the bond-stretching vibration of the N–H group at ~3350 cm^{-1}; 3) the bond-bending vibration of the N–H group at ~1150 cm^{-1}; and 4) the bond-stretching vibration of the Si–H group at ~2170 cm^{-1}. Beginning at annealing temperatures of about 400°C, hydrogen was evolved from films deposited at 300°C from different source-gas mixtures of NH_3/SiH_4. These studies extended to anneals at 1200°C, all for periods of ~1 min. The loss of H atoms has been fitted to a thermally activated process with an activation energy of ~0.4 eV. The low activation energy is consistent with the H-atom loss being accompanied by the formation of molecular hydrogen. It has been shown that the as-deposited hydrogenated silicon nitride films contain nearest-neighbor Si–H and H–N–Si bonding arrangements. The annealing studies indicated significant decreases in both the Si–H and N–H vibrational amplitudes with increasing annealing temperature [5]. The Si–H feature fell below the level of detection for temperatures of about 600°C, and only N–H features were observed after the 1200°C anneal. A reaction pathway for H-atom evolution that is consistent with the IR results is

$$\text{Si–H}\text{---}\text{H–N–Si} \rightarrow \text{Si}^{\circ} + {}^{\circ}\text{N–Si} + \text{H}_2, \quad (2)$$

where the superscript notation ($^{\circ}$) indicates a neutral dangling bond and the H---H notation indicates nearest-network neighbors. In the temperature range to about 500–600°C, the only significant changes in the IR were decreases in the amplitudes of the Si–H and N–H vibrations, whereas at temperatures greater than about 600°C, there was a systematic increase in the absorbance for the Si–N feature as well. The combination of these results for annealing temperatures >600°C suggests that as H atoms were evolved from the films, new Si–N bonds were formed. This is represented symbolically as follows:

$$\text{Si–H}\text{---}\text{H–N–Si} \rightarrow \text{Si}^{\circ} + {}^{\circ}\text{N–Si} + \text{H}_2, \quad (3a)$$

$$\text{Si}^{\circ} + {}^{\circ}\text{N–Si} \rightarrow \text{Si–N–Si}. \quad (3b)$$

An ~10:1 source-gas ratio of NH_3:SiH_4 gave optimum electrical performance both in thin-film transistors (as-deposited) [30] and in CMOS devices (after a 900°C anneal) [6, 14].

Combining IR and AES data with a statistical model of bonding in amorphous Si:N:H alloys indicates that this source-gas ratio corresponds to a film in which the respective silicon, nitrogen, and hydrogen concentrations ([Si], [N], and [H]) are [Si] ≈ 28 at.%, [N] ≈ 42 at.%, and [H] ≈ 30 at.% [30]. The average number of atomic bonds for this composition is ~2.7, the same as in device-quality plasma-deposited SiO_2 films [31]. Hydrogenated silicon nitride films with this composition have been used as-deposited in amorphous silicon thin-film transistors (TFTs) as the gate dielectrics and yielded optimized electrical performance. For example, electron mobilities of ~1.5 cm^2-V^{-1}-s^{-1} [30, 31] were obtained, indicating a low concentration of charged defects in the nitride gate dielectrics. For both lower and higher source-gas ratios, the film composition changed and the effective channel mobility in the TFTs decreased [31]. The unique aspects of these nitride films that differentiate them from thermally grown nitride films are presented in **Table 4**.

Table 4 Differences in bonding and bond defects between plasma-deposited and CVD nitrides.

Plasma-deposited 300°C	*LP (RT) CVD > 500°C*
[Si], [N], [H] function of source-gas ratio ~30 at.%	No IR-detectable bonded H
As-deposited defect-free bonds/atom ~SiO_2	As-deposited, high defect density
After 900°C anneal gate quality—low defects high reliability in n-MOS	After 900°C anneal, still defective even after N_2, NH_3
Does not require postdeposition oxidation for device quality	Device quality after postdeposition oxidation

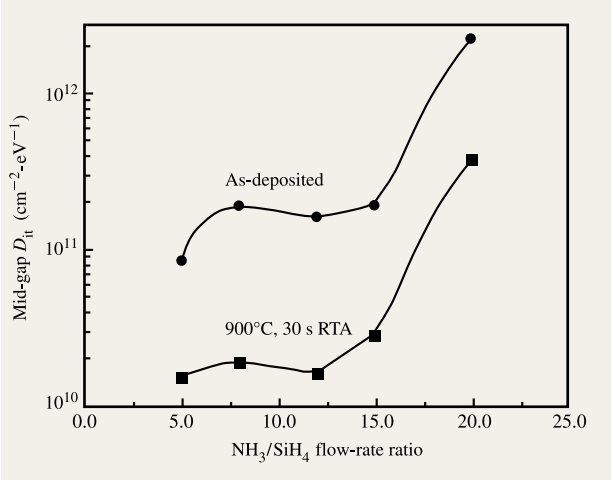

Figure 17

Interface trap density D_{it} vs. flow-rate ratio of NH_3/SiH_4 for ONO structures with ~5 nm of each constituent dielectric layer. Data are shown for as-deposited and 900°C annealed films.

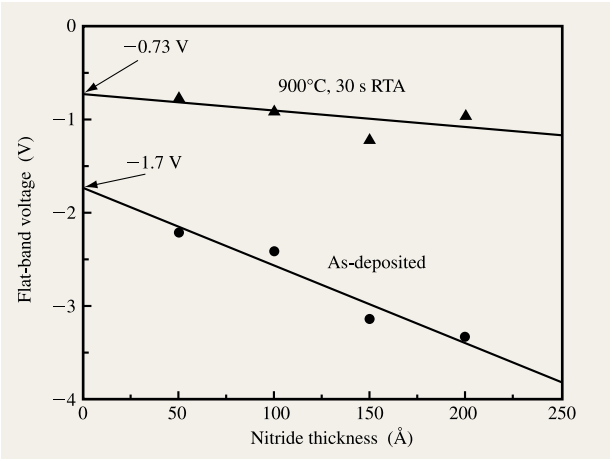

Figure 18

Flat-band voltage (V_{FB}) vs. nitride layer thickness for ONO stacks with 5-nm-thick top and bottom oxide layers and variable nitride layer thickness. Linear fits to the data for as-deposited and 900°C annealed films are included.

• *Electrical performance in n-MOS ONO capacitors*

The initial studies of the plasma-deposited nitride films were performed in devices with stacked ONO gate dielectrics. Three studies were performed. The first identified processing conditions necessary for obtaining low-defect-density films, establishing that postdeposition annealing at 900°C was required [14]. The second addressed fixed positive charge in the ONO stack and showed that this was localized at the internal interfaces between the constituent layers of the ONO stack, rather than in the bulk of the nitride layer, and was reduced significantly after the 900°C anneal [14]. The third study addressed scaled devices with $t_{ox\text{-}eq} \approx 4.5$ nm (~2-nm layers of oxide, nitride, and oxide), demonstrating that reliability in the ONO stack was better than for homogenous oxide dielectrics [32].

Figure 17 displays interface trap density (D_{it}) vs. NH_3/SiH_4 source-gas ratio for metal–insulator–semiconductor (MIS) capacitors with ONO dielectrics (all layers are ~5 nm thick) [14]. The as-deposited values for D_{it} were higher than for metal–oxide–semiconductor (MOS) capacitors with deposited oxides, although both types of devices had been subjected to the same processing temperatures—300°C for the oxidation and deposition steps, and 400°C for the postmetallization anneals (PMAs). In addition, the flat-band shifts (V_{FB}), as shown in **Figure 18**, were significantly more negative in the MOS devices, indicating a higher concentration of trapped positive charge (Q_F) in the vicinity of the Si–SiO_2 interface. However, after a 30-s 900°C RTA in either Ar with a partial pressure of O_2 sufficient to prevent formation and evolution of gaseous SiO at the Si–SiO_2 interface, or only Ar, the values of D_{it} approached those of devices with oxide dielectrics, while the values of Q_F were still about a factor of 2 to 3 higher than for homogeneous plasma-deposited oxides [14]. In the next section, these increased values of Q_F are shown to be associated with fixed charge at the internal ON and NO interfaces of the ONO stack. The best electrical performance after the RTAs was obtained for values of NH_3/SiH_4 flow ratios $R \approx 8$–10, essentially at the same R

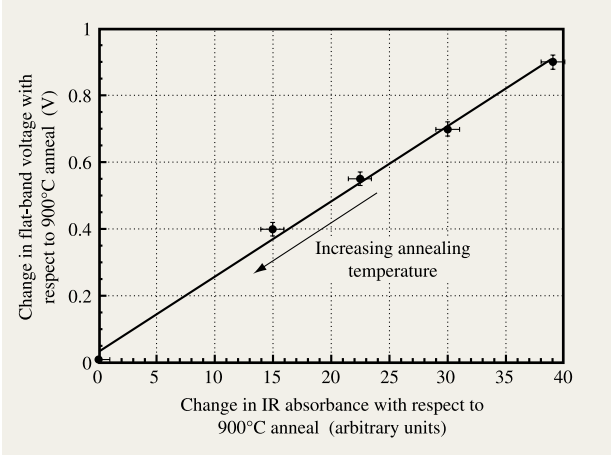

Figure 19

Correlation between IR and V_{FB} device results. Changes in IR absorption at the Si–N bond-stretching frequency referenced to 900°C are plotted on the *x*-axis, and changes in V_{FB} relative to 900°C on the *y*-axis.

values for optimized performance of plasma-deposited nitrides in a-Si TFTs [30, 31]. Breakdown fields in the MOS devices (E_B) were typically in excess of 10 MV/cm before and after the 900°C RTAs.

The second set of experiments, described in [14], demonstrated that defects giving rise to fixed positive charge, shifting V_{FB} to more negative voltages, remained after the 900°C anneal and were at the internal interfaces between the constituent oxide and nitride layers of the ONO stack. This is shown in **Figure 19** by the linear variation of V_{FB} with nitride-layer thickness in ONO stacks in which both O-layer thicknesses were fixed at 5 nm and the N-layer thickness was varied up to 20 nm. Analysis of data from [14] demonstrates that interfacial positive charge densities were reduced at both interfaces from values in excess of 10^{12} cm^{-2}, after a 400°C PMA following film deposition at 300°C, and to the mid-to-low 10^{11}-cm^{-2} range when the processing included a 900°C RTA between deposition and metallization and the 400°C PMA.

Finally, MOS devices with ONO dielectrics having oxide equivalent thicknesses in the range of 4.5–6.0 nm have also been fabricated. The electrical properties of these devices are essentially the same as those discussed above; and, as shown in [32], they display significantly improved reliabilities with respect to devices with thermally grown oxide layers (see **Table 5**).

The differences in the processing requirements for optimized electrical performance in TFTs and MOS devices are explained by combining the temperature-dependent IR absorption with the results of these device studies. This combination of IR data [5] and device data [14] is consistent with the following explanation. Release of bonded H at temperatures between ~400°C and 600°C results in defect generation via the creation of Si and/or N-atom dangling-bond defects, degrading the MOS devices not subjected to the 900°C anneal. Depending on the position of these states within the bandgap of Si, they can be either neutral or charged. Additionally, their charge states can change under an applied voltage bias, as in *C–V* measurements. However, the Si and N dangling-bond defects created by H evolution are annealed out at higher temperatures (~900°C), as evidenced by increased IR absorbance in the Si–N bond-stretching band. The achievement of optimized performance in TFTs, where the postdeposition processing temperatures are no higher than the 300°C deposition temperatures, is consistent with IR results which indicate that 300°C is below the threshold for detectable H-atom release from the Si–H and SiN–H sites.

The explanation for improved performance in MOS devices after the 900°C anneal is supported by the data displayed in Figure 19, which combines IR and device results. Changes in IR absorption data at the Si–N bond-stretching frequency referenced to the 900°C are plotted on the *x*-axis, and the decrease in defects in films annealed at temperatures to 900°C are plotted on the *y*-axis. The origin ($x = 0$ and $y = 0$, respectively) represents the final value of absorbance of the Si–N band at 900°C relative to ~500°C, and the final value of V_{FB} with respect to the more negative values found at temperatures below 900°C. This plot supports a model in which charged Si- and N-atom dangling bonds are the source of the fixed positive charge that shifts the V_{FB} to increasingly larger negative values in films annealed below 900°C. Recombination of these dangling bonds in films annealed at 900°C forms additional Si–N bonds, neutralizing the defects which possess fixed positive charge.

Table 5 Comparison between electrical reliability properties of devices with thermally grown oxides and plasma-deposited and annealed ONO stacked dielectrics ($t_{ox\text{-}eq} \approx 4.5$ nm) [32].

Dielectric layer	Q_{BD} (C-cm^{-2})	*Injected Q* (C-cm^{-2})	V_{th} (V)	D_{it} ($\times 10^{11}$ cm^{-2}-eV^{-1})
Plasma ONO	17	0.17	−0.15	6.5
Thermal oxide	7–10	0.1	−0.35	8.6

• *Electrical performance of n-MOS FETs with ON gate dielectrics*

Figure 20 presents data on the performance of n-MOS FETs incorporating stacked ON gate dielectrics with $t_{ox\text{-}eq} < 2$ nm [17, 33]. The approximate electrical thickness of the gate dielectrics was obtained to ±0.1 nm from analysis of *C–V* data. TEM micrographs of the composite ON structures indicated approximately equal oxide and nitride layer thicknesses. Based on an approximately 1:2 ratio of silicon nitride to silicon dioxide static dielectric constants, this means that a 1.8 ± 0.1-nm ON film is composed of oxide and nitride layers which are each approximately 1.2 ± 0.1 nm thick. The transistor characteristics, drain current (I_D) vs. drain bias voltage (V_D), are plotted in Figure 20 for different normalized gate voltages ($V_G - V_{th}$), where V_G is the gate voltage and V_{th} is the threshold voltage. Since I_D is proportional to the inversion capacitance, the equality of the drain currents for the two devices is consistent with their having approximately the same $t_{ox\text{-}eq}$. The oxide dielectric was formed by a combination of a 300°C remote-plasma-assisted oxidation and deposition, and an *in situ* vacuum anneal at 900°C for 30 s. The data in [17] and [33] indicate the robust nature of the ON gate dielectric; i.e., there is no detectable degradation after an integrated current exposure equivalent to about 1.6×10^4 C-cm^{-2} ($V_G = 2.00$ V). Stressing times have been extended so that more than 6×10^4 C-cm^{-2} have been injected with no changes in FET properties. However, changes in current are produced when gate bias voltages are increased to 2.75 V, or >15 MV/cm, a field that is significantly higher than any anticipated operating gate bias voltage would produce. Finally, the direct tunneling current for the FET with the ON dielectric was reduced by a factor of ~5–7 with respect to that of the FET with oxide dielectric. The reduction in tunneling current for substrate injection shown in **Figure 21** is not an artifact resulting from neglect of the differences in threshold voltage between the oxide and ON devices, since these differences have been taken into account in the plotting of these data. Similar reductions are seen in the tunneling current for oxide-equivalent fields less than about 10 MV-cm^{-1} for gate injection. For fields greater than about 12 MV-cm^{-1}, the tunneling current is larger in the ON stack owing to the reduced tunneling barrier height at the gate electrode dielectric interface [1, 2]. The decreased tunneling for the FET with the ON dielectric was initially attributed to the increased physical thickness of the ON stack compared to that of the oxide. TEM measurements indicated that the physical thickness of the ON stack was 2.3 ± 0.2, whereas that of the oxide was less than 2 nm. This aspect of these ON gate dielectrics is of significance in aggressively scaled CMOS devices, i.e., increases in the ratio of I_D to the tunneling leakage current. However, the actual decrease is less than anticipated from model calculations which have assumed equal tunneling masses of ~0.5 m_0 for electrons in the oxide and nitride layers, where m_0 is the free-electron mass [1, 2]. A discussion of these data [2]

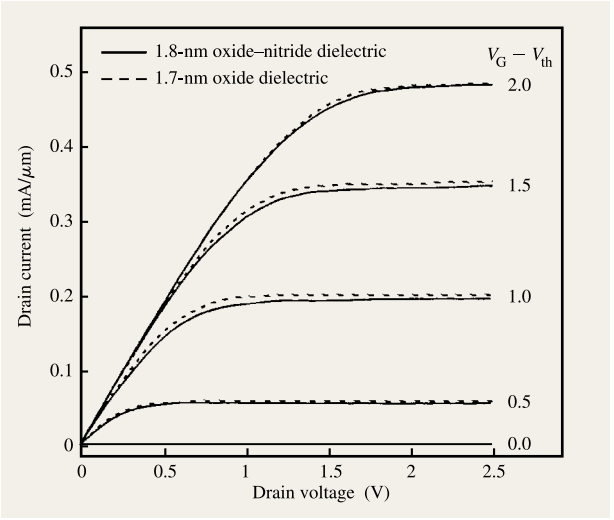

Figure 20

I_D–V_D plots for FETs with ultrathin gate oxide and composite ON gate dielectrics with $t_{ox\text{-}eq} < 2$ nm. Since there are differences in threshold voltage of ~0.2 V, the curves are plotted as a function of $V_G - V_{th}$.

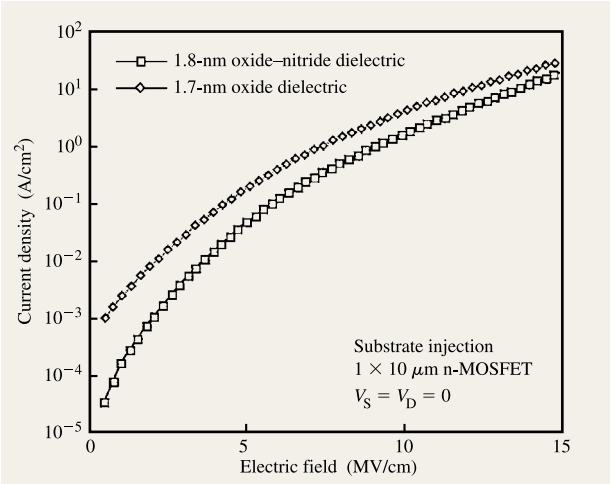

Figure 21

Substrate injection tunneling current vs. electrical field for devices with oxide and stacked ON dielectrics, with $t_{ox\text{-}eq} \approx 1.9$ nm.

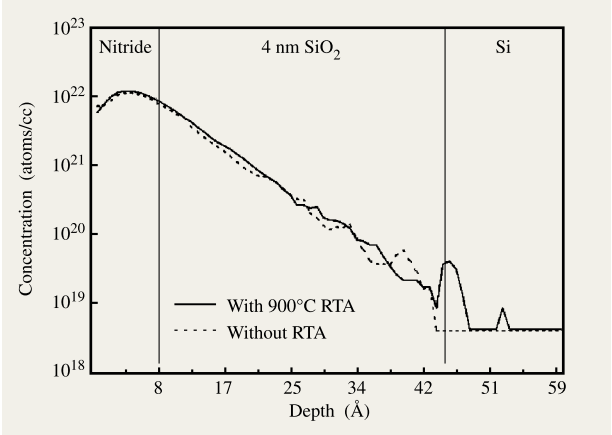

Figure 22

Nitrogen depth profile determined by SIMS analysis for a nitride/oxide (~0.8-nm/4.0-nm) dual-layer gate dielectric before and after a 900°C postdeposition RTA.

suggests that the tunneling mass of electrons in the plasma-deposited and annealed nitrides is indeed less than that in the oxides: ~0.3 m_0, as compared to ~0.5 m_0.

• *Electrical performance of ON dielectrics in p-MOS capacitors and FETs*

This section of the paper is divided into three parts dealing with results from studies of 1) the effectiveness of plasma-deposited nitride layers for stopping B-atom transport out of p^+ polycrystalline silicon gate electrodes during aggressive dopant activation anneals [16, 18, 34]; 2) the effectiveness of plasma-deposited silicon oxynitride alloys in suppressing B-atom transport;[7] and 3) the performance and reliability of p-MOS FETs with stacked ON gate dielectrics[7] [16, 18, 34].

The majority of the P-channel MOS devices with p^+ polycrystalline Si gate electrodes were fabricated on n-type Si(100) substrates doped to 5×10^{17} cm^{-3}; however, some wafers were implanted with phosphorus to increase the channel doping to 1.1×10^{18} cm^{-3}, and these are identified below. Bottom oxides were grown either by thermal oxidation in oxygen with 4.5% HCl at 800°C for thicknesses of 1.5 nm to 4.7 nm, or by remote-plasma oxidation in N_2O at 300°C for an approximate thickness of 0.6 nm. These were followed by RPECVD nitride depositions (N_2/SiH_4 source gases) ranging from ~0.2 nm to ~2.4 nm [6]. The top nitride layer thickness was estimated by AES to be ±0.1 nm [35]; the AES determinations were supported by spectroscopic ellipsometry (SE) measurements and deposition rate extrapolations. Postdeposition anneals of the ON stacked dielectrics were performed in He at 900°C for 30 s [6]. This annealing 1) drives nitrogen to the oxide/substrate interface to form the bottom nitride layer of the resulting nitride–oxide–nitride (NON) stack; 2) reduces the hydrogen concentration in the nitride layers[8] by about a factor of 2 [6]; and 3) minimizes bonding defects in the nitride layer by forming additional Si–N bonds (see above discussion) [5]. Boron implantation into 0.2-μm-thick polycrystalline silicon, at an implant energy of 20 MeV and a level of 5×10^{15} cm^{-2}, was used to form the p^+ gate electrode. This was followed by the deposition of a 200-nm low-temperature oxide (LTO) to prevent the B out-diffusion from the implanted polycrystalline Si during dopant activation anneals ranging from 950°C to 1000°C for up to four minutes. After deposition and patterning of Al for gate, source, and drain contacts, a conventional postmetallization anneal (PMA) in forming gas (N_2/H_2) at 400°C was performed for 30 min. The equivalent oxide electrical thickness, $t_{ox\text{-}eq}$, was determined from *C–V* measurements for capacitors biased in the accumulation region using quantum-mechanical corrections [36]; p-MOS FETs with thermal oxides were fabricated as control devices.

SIMS analysis of an ~0.8-nm/4.0-nm dual-layer nitride/oxide structure was performed (**Figure 22**). The trailing of the nitrogen signal into the oxide is a result of the SIMS analysis method. After annealing at 900°C for 30 s, a distinct nitrogen peak appears at the oxide/silicon interface, showing that N atoms diffuse into the oxide and pile up at the oxide/silicon interface during the anneal, so that the annealed devices have an NON stacked structure. However, in contrast to the postoxidation nitridation process of [16], which achieves monolayer levels of N-atom incorporation for exposure times of 90 s, the level of interface nitridation in Figure 22 is about 100 times smaller, $\sim 5 \times 10^{12}$ cm^{-2}. As shown below, coupled with the blocking of B-atom transport out of p^+ polycrystalline Si gate electrodes, this is sufficient to promote excellent reliability. However, as discussed later, it is insufficient to provide the reduction in direct tunneling current associated with monolayer-level interface nitridation.

Capacitance–voltage (*C–V*) curves for an 0.8-nm ultrathin nitride layer on top of thermally grown oxides indicate a significant improvement for suppression of B-atom penetration to the Si–SiO_2 interface compared to thermal oxide p-MOS devices. **Figure 23(a)** contains the superposition of normalized quasi-static *C–V* curves for capacitors with control oxides, and NON stacked dielectrics in which the deposited nitride layers are ~0.4 nm or ~0.8 nm thick. Oxide thicknesses for these

[7] Y. Wu and G. Lucovsky, submitted to *IEEE Trans. Electron Devices* (1999).

[8] V. Misra, Z. Wang, Y. Wu, H. Niimi, G. Lucovsky, J. Wortman, and J. Hauser, submitted to *IEEE Trans. Electron Devices* (1999).

measurements are in a range from ~4.0 nm to ~4.7 nm. The *C–V* curve for the control oxide is shifted to a more positive voltage by approximately +0.5 V with respect to the capacitor with 0.8-nm top nitride. Based on the calculated value of flat-band voltage as determined from the substrate and gate electrode doping, the large shift of the reference oxide flat-band voltage indicates significant B penetration to the substrate for that device [34, 37]. The flat-band voltage shift of the capacitor with 0.4-nm top nitride film is intermediate, indicating that B-atom transport can be controlled for a fixed thermal budget by simply changing the thickness of the top nitride layer.

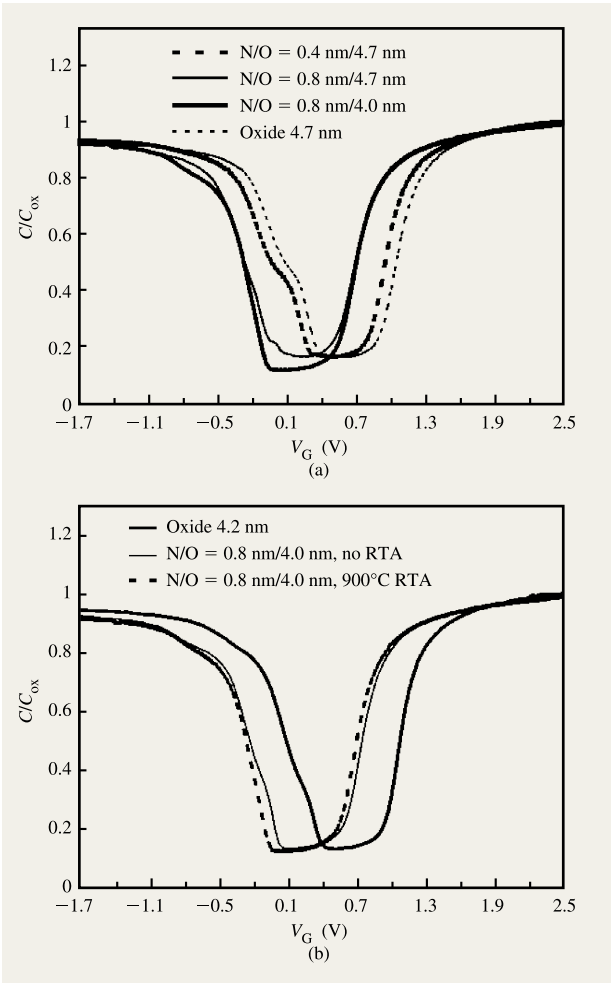

Figure 23

(a) Normalized quasi-static *C–V* curve for a thermal oxide, and 0.4 nm and 0.8 nm top nitride deposited onto thermal oxides. Curves are shifted owing to boron penetration through thin gate material. The dopant activation annealing is 1000°C for 60 s. (b) Normalized quasi-static *C–V* curve for oxide and N/O stack with and without postdeposition treatment. The RTA condition is 900°C for 30 s in a He ambient, and the dopant activation anneal is 1000°C for 60 s.

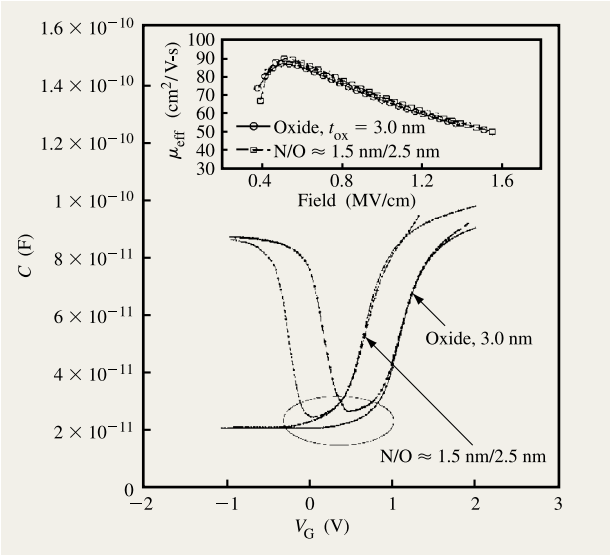

Figure 24

High-frequency (HF) and quasi-static (Q-S) *C–V* curves at the onset of inversion for devices with ON and oxide dielectrics with $t_{ox\text{-}eq} \approx 3.0$ nm. The device with the ON stack has a nitrided Si–SiO_2 interface. The relative separation between the HF and Q-S indicates that the device with the ON stack and nitrided interfaces has a lower interface defect density. The insert shows the effective mobility for these two devices.

In addition, the B-atom blocking capability is improved by performing the 900°C RTA prior to the deposition of the polycrystalline Si gate electrode, as indicated by the *C–V* data in **Figure 23(b)**. Annealing of RPECVD nitride film at 900°C for 30 s prior to the polycrystalline Si deposition, implantation, and activation anneal retards the diffusion of B through the top nitride during the activation anneal. As displayed in Figure 23(b), the *C–V* curve for the stacked dielectric film without 900°C annealing shows a small shift to positive voltage with respect to the film with the postdeposition RTA. This indicates that a small amount of B migrates through the nonannealed nitride films to the substrate. Additionally, this *C–V* curve exhibits some distortion at the onset of the inversion region, presumably due to modification of the channel region potential by B-atom compensation.

Additional benefits are also derived from the top-surface nitride layers in the NON gate dielectrics, as shown in **Figure 24** for devices with an equivalent-oxide thickness ≈3.0 nm. In addition to eliminating a large shift in flat-band voltage, the deviation between the high-frequency and quasi-static *C–V* curves at the onset of

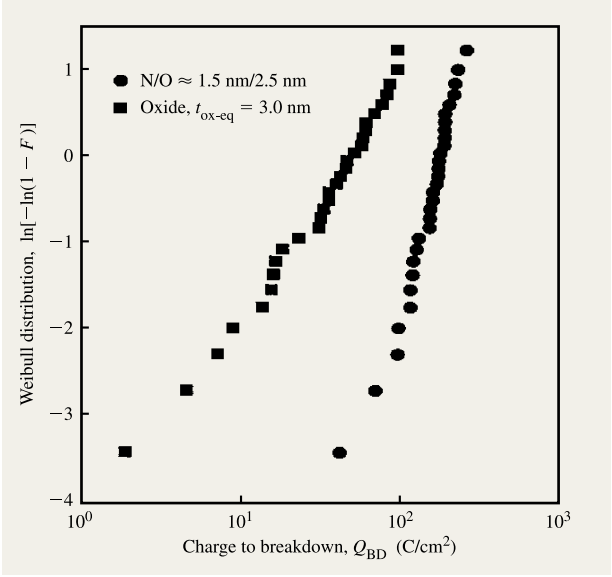

Figure 25

Weibull plots of Q_{BD} for capacitors with ON and oxide dielectrics for $t_{ox\text{-}eq} \approx 3$ nm under constant substrate injection current stressing at a current density of 500 mA-cm^{-2}.

inversion is reduced for the device with the NON dielectric, indicating a reduced interface defect density.

Locating a nitride film on top of an oxide as a diffusion barrier provides other advantages in p-MOS devices. Previous studies have shown that N-atom concentration peaks at the Si–SiO_2 interface for dielectrics grown by oxidation in NO or N_2O. Since B-atom transport is stopped at the Si–SiO_2 interface in these devices, this approach allows an accumulation of B atoms in the bulk oxide, degrading the dielectric reliability compared to oxides grown in O_2 without nitrogen incorporation [16, 34]. By preventing B from diffusing into the bulk oxide layer, devices with NON dielectrics display improved reliability compared to devices with oxides grown in O_2. **Figure 25** shows Weibull plots for gate dielectrics with $t_{ox\text{-}eq} \approx 3.0$ nm under a substrate injection stress of 500 mA-cm^{-2}. The charge to breakdown, Q_{BD}, is improved up to an order of magnitude in the devices with NON dielectrics.

It has been shown above that an 0.8-nm top nitride is effective in stopping boron penetration, even with a thermal budget as high as 1000°C for four minutes [13, 34]. However, as shown in Figure 23(a), some B-atom penetration is observed when the top nitride thickness is reduced to 0.4 nm. It is difficult to explain this thickness dependence of B-atom transport through these nitride layers by conventional diffusion theory ([38] and references therein). A percolation model has been developed to account for boron transport in nitride and silicon oxynitride films.[9] The model is supported by the experimental results of **Figure 26**. Boron penetration to the Si–SiO_2 interface has been studied by changes in flat-band voltage for two different types of oxynitride deposited onto a 2.5-nm oxide. The $(SiO_2)_x(Si_3N_4)_{1-x}$ alloys were formed by an RPECVD technique described in [6]. The SiO_2 fractions, x, in the $(SiO_2)_x(Si_3N_4)_{1-x}$ alloys were 0, 0.3, 0.7, and 1. In the first set of experiments [Figure 26(a)], the thickness of the oxynitride alloy films was fixed at ~0.8 nm and the nitrogen concentration was reduced by alloying; i.e., by increasing x. In the second set of experiments [Figure 26(b)], the areal density of nitrogen atoms was fixed at $\sim 4.5 \times 10^{15}$ cm^{-3} (as in an 0.8-nm Si_3N_4 film), and the oxynitride film thickness was increased as the SiO_2 fraction was increased. Figure 26(a) shows the effectiveness of the boron diffusion barrier formed by 0.8 nm of the $(SiO_2)_x(Si_3N_4)_{1-x}$ alloys. The extent of B-atom penetration is compared by studying the flat-band voltage shifts from the theoretical value determined by the work-function difference between the gate and the substrate Si. In Figure 26(a), the device with an 0.8-nm top nitride film shows essentially no flat-band voltage shift (to ±0.05 V), confirming the results shown in Figures 23(a) and 23(b). By reducing the nitrogen areal density by substituting $(SiO_2)_x(Si_3N_4)_{1-x}$ alloys for the nitride layer, and keeping the physical thickness at 0.8 nm, the flat-band voltage is shifted to more positive voltages, indicating increased B-atom penetration. Similar results apply to the second set of experiments, in which the nitrogen areal density is fixed by increasing the thickness of the $(SiO_2)_x(Si_3N_4)_{1-x}$ alloy films. Reduced B-atom penetration through the second set of films supports a model in which boron transport proceeds via a percolation-like process involving the connectivity of the oxygen atom pathways through the oxynitride alloys. This model is qualitatively different from the model for B-atom transport through oxides as proposed by Fair [38], and will later be discussed in more detail.[9]

It is well established that the performance and reliability of p-MOS FETs with p^+ polycrystalline Si gate electrodes and thermally grown oxide gate dielectrics can be degraded from B-atom penetration into the channel region [39, 40]. In marked contrast, p-MOS FETs with p^+ polycrystalline Si gate electrodes and stacked ON gate dielectrics with nitrided Si–SiO_2 interfaces display improved short-channel performance and reliability. **Figure 27** shows I_D–V_D characteristics with an effective channel length of ~0.5 μm for p-MOS FETs with stacked ON and homogeneous oxide gate dielectrics with the same

[9] G. Lucovsky, Y. Wu, and R. Fair, submitted to *J. Vac. Sci. Technol. B* (1999).

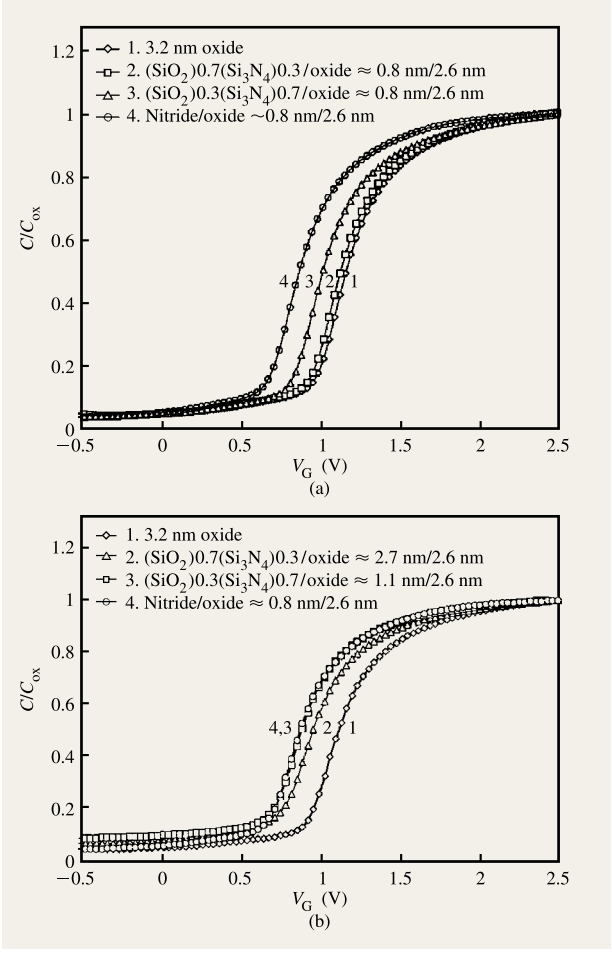

Figure 26

Boron penetration as monitored by changes in flat-band voltage for oxynitride alloys $[(SiO_2)_x(Si_3N_4)_{1-x}$, $(x = 0, 0.3, 0.7,$ and $1)]$ deposited onto a 2.5-nm oxide. The alloys were formed by an RPECVD technique described in [6]. In (a) the thickness of the oxynitride alloy is fixed at ~0.8 nm, and the nitrogen concentration is reduced by alloying. In (b) the areal density of nitrogen atoms is fixed, and the barrier layer thickness is increased proportionally for increased oxide incorporation.

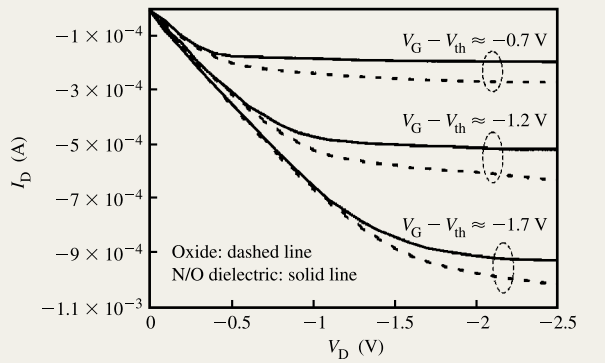

Figure 27

I_D–V_D characteristics for p-MOS FETs with p^+ polycrystalline Si gate electrodes. The solid lines indicate a FET with a 1.9-nm ON dual-layer stack, and the dashed lines indicate a FET with a 1.9-nm control oxide. The effective channel length is 0.5 μm. Since there are differences in threshold voltage of ~0.5 V, the curves are plotted as a function of $V_G - V_{th}$.

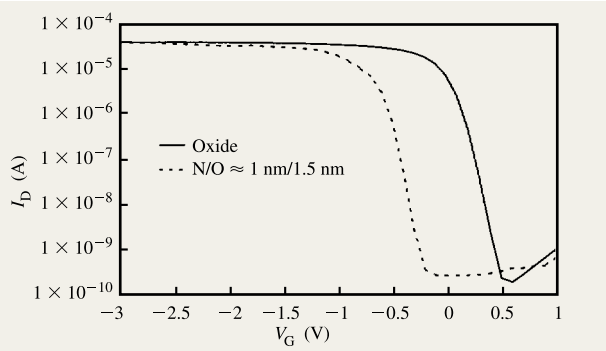

Figure 28

I_D–V_G characteristics for p-MOS FETs (L/W = 0.8 μm/20 μm) with oxide and ON dielectrics for $t_{ox\text{-}eq}$ = 1.9 nm. The increased off-state leakage current for the oxide device is indicative of a larger tunneling current.

$t_{ox\text{-}eq}$ of approximately 1.9 nm. The device with the stacked ON dielectric shows improved (i.e., flatter) saturation characteristics with respect to the device with a homogeneous oxide dielectric. The reduced trend to saturation on the I_D–V_D curves for the device with the oxide dielectric is indicative of an enhanced short-channel effect associated with B-atom penetration into the channel region [16].

Figure 28 shows I_D–V_G characteristics for p-MOS FETs with a channel length of 0.8 μm and an oxide-equivalent thickness of ~1.9 nm. By preventing boron from entering the channel region, the subthreshold slope is improved from 99 mV/decade for a device with an oxide dielectric to 82 mV/decade for a device with the ON stacked dielectric. In addition, the impact of a reduced tunneling current in this sub-2.0-nm region is reflected in an improved transistor off-state characteristic.

The effective mobilities of devices with ON and oxide dielectrics were extracted from I_D–V_G curves in the linear region in the conventional way with V_D = −0.05 V. Almost identical effective channel hole mobilities were obtained for p-MOSFETs with ON and oxide dielectrics

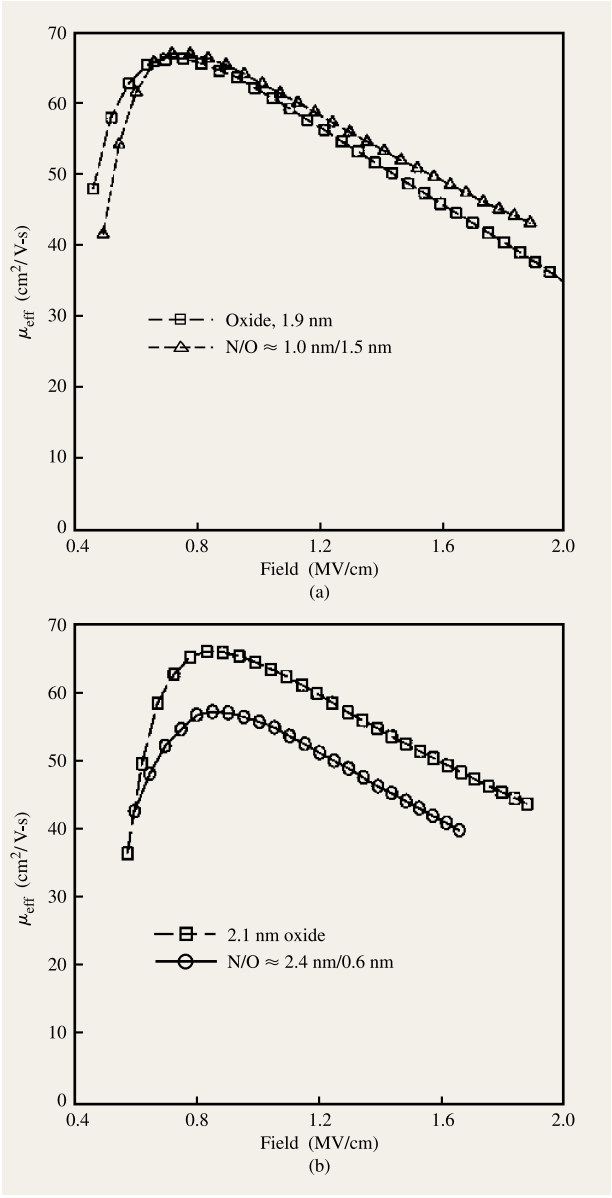

Figure 29

Effective hole mobility vs. effective normal field for p-MOS FETs for oxide and dual-layer ON gate dielectrics with $t_{ox\text{-}eq} \approx 2.0$ nm, with bottom oxide thickness of 1.5 nm (substrate doping 5×10^{17} cm^{-3}); (b) 0.6-nm-thick bottom oxide (substrate doping 1.1×10^{18} cm^{-3}).

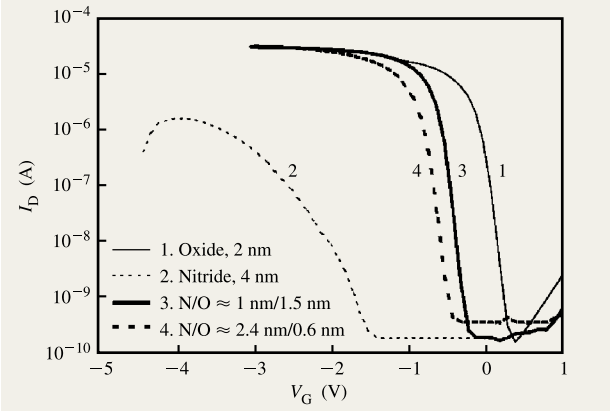

Figure 30

I_D–V_G characteristics for oxide, ON dual layer, and nitride devices with $t_{ox\text{-}eq} \approx 2.0$ nm, indicating the effects of different nitride and oxide interface regions.

for oxide equivalent $t_{ox\text{-}eq} \approx 3$ nm, as shown by the insert in Figure 24.[10] **Figures 29(a)** and **29(b)** show the mobilities for devices with stacked ON dielectrics for scaling $t_{ox\text{-}eq}$ to ~1.9 nm. Devices with two different bottom oxides, ~1.5 nm and ~0.6 nm, were used to monitor the effect of oxide thickness on the channel mobility. As shown in Figure 29(a), the device with a bottom oxide thickness of 1.5 nm displayed a field-dependent channel mobility essentially equal to that of the device with the oxide dielectric. However, when the bottom oxide thickness was reduced to ~0.6 nm, as shown in Figure 29(b), a 10% mobility degradation was found for the device with the ON dielectric as compared to the device with the oxide device. Since the decrease in hole mobility is essentially independent of electric field, this effect is not due to increased surface roughness, but instead is due mostly to a difference in substrate doping. A shift of the peak mobility for the device with the 0.6-nm bottom oxide to higher gate voltages, as indicated by the I_D–V_G plots (see **Figure 30**) is also due mostly to a higher substrate doping density, $\sim 5 \times 10^{17}$ cm^{-3} for the device with the 1.5-nm/1.0-nm ON stack, and $\sim 1.1 \times 10^{18}$ cm^{-3} for the device with the 0.6-nm/2.4-nm ON stack.

The direct substitution of Si_3N_4 for SiO_2 should yield a higher effective dielectric constant than for stacked ON dielectrics. However, it is not possible because of a significantly increased defect density at Si–Si_3N_4 interfaces in both p-MOS and n-MOS devices relative to Si–SiO_2 interfaces in oxide or stacked ON devices[11] [41]. As shown in Figure 30, when a nitride layer is substituted for the ON stack with a monolayer nitrided interface, the nitride curve is shifted in the negative voltage direction by approximately 1 V, and the drive current is degraded by more than one order of magnitude. The soft turn-on of

[10] Y. Wu and G. Lucovsky, submitted to *IEEE Trans. Electron Devices* (1999).

[11] V. Misra, Z. Wang, Y. Wu, H. Niimi, G. Lucovsky, J. Wortman, and J. Hauser, submitted to *IEEE Trans. Electron Devices* (1999).

this device indicates a high density of interface traps, and the threshold voltage shift indicates a large amount of fixed positive charge, of the order of 10^{13} cm^{-2}. Other aspects of device performance are discussed in detail by Misra et al.[12] Figure 30 also shows that when ~0.6 nm of oxide is inserted between the Si substrate and nitride layer, the drive current is improved significantly and is comparable to that of devices with an oxide dielectric. If the thickness of the buffer oxide layer can be further reduced into the 0.3-nm range, this would lead to further reductions of the oxide-equivalent thickness.

Differences in the behavior of interface defects at monolayer nitrided Si–SiO_2 interfaces and Si–Si_3N_4 interfaces have been discussed in the context of constraints imposed by bonding coordination differences between SiO_2 and Si_3N_4. For example, as already noted above, in bulk glasses and thin films, it has been shown that low defect densities in bulk SiO_2 result from an average bonding per atom of 2.7 [31]. This is low enough to match the number of bonding constraints per atom to the network bonding topology to form low-defect-density films [31]. A value of the average bonding per atom of ~3 marks a boundary between device quality and highly defective thin films. A procedure has been developed for extending constraint theory to interfaces between crystalline Si and gate dielectric materials, which also shows that an average bonding per atom of ~3 also marks a demarcation between low defect density and highly defective interfaces [41]. This model provides an explanation for the behavior shown in Figure 30; in particular it demonstrates that insertion of an oxide layer ~0.6 nm thick is sufficient to reduce the average number of bonds per atom at the interface from ~3.5 for a Si–Si_3N_4 interface to ~3 for the Si–SiO_2–Si_3N_4 interface of the aggressively scaled device of Figure 30.

Figure 31 shows the peak-transconductance degradation under channel hot-carrier injection for p-MOS FET devices with the effective channel length $L_{eff} = 0.5$ μm, and ON dual-layer dielectrics and oxide dielectrics with $t_{ox\text{-}eq} \approx 1.9$ nm. The devices were stressed at the peak substrate current for worst-case degradation at a high drain bias of −5 V. Compared with the control oxide, the p-MOS FET with the ON dielectric shows less transconductance degradation during hot-carrier injection, implying a more robust silicon/dielectric interface. This improved interface immunity against hot-carrier stressing is believed to be due to the interfacial strain relaxation by the nitrogen incorporation at the oxide/substrate interface during the postdeposition annealing.

As the gate oxide thickness is reduced into the sub-2.0-nm region, the rapid increase of direct tunneling current becomes a major obstacle for device scaling. **Figure 32** shows a comparison of tunneling current through 100-μm × 100-μm capacitors with a 1.9-nm oxide

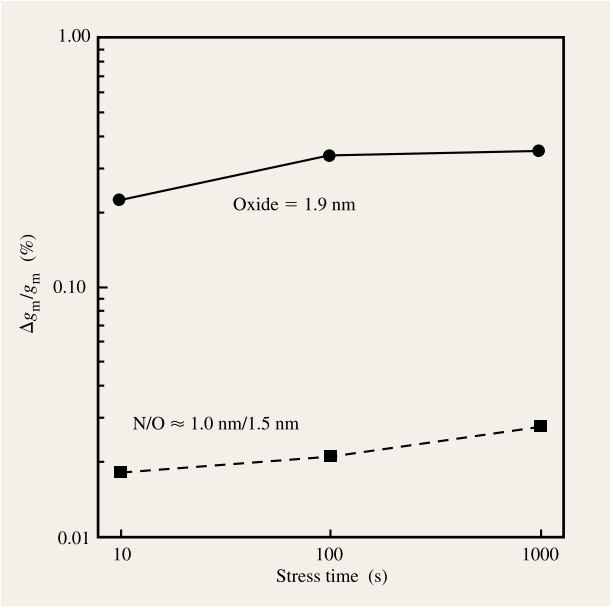

Figure 31

Hot-carrier stressing for p-MOS FETs with ON and oxide dielectrics. The device with the ON stacked dielectric displays increased immunity to hot hole stressing under the following stress conditions: $V_G - V_{th} = -1.7$ V, $V_D = -5$ V.

and with stacked ON devices with O/N thicknesses of ~1.5 nm/1.0 nm and 0.6 nm/2.4 nm that correspond to approximately the same oxide-equivalent thickness. The gate current for devices with NON gate dielectrics is about an order of magnitude lower at a bias voltage of 1 V. In addition, because the physical thickness of ON films is greater than that of an oxide for the same $t_{ox\text{-}eq}$, the equivalent breakdown electrical field is increased from ~13.5 MV/cm for an oxide to ~16 MV/cm for stacked ON devices. Since the physical thicknesses of the two devices with stacked ON dielectrics are significantly different, 2. 5 nm for the 1.0-nm/1.5-nm dielectric and 3.0 nm for the 2.4-nm/0.5-nm dielectric, it is not obvious why the tunneling current has not decreased significantly and differently for the two NON devices. Model tunneling calculations anticipate decreases in tunneling current for fixed $t_{ox\text{-}eq}$ with an increasing N to O thickness ratio [2]; however, as was discussed with respect to the tunneling changes of similar magnitude in Figure 21, the reduction in direct tunneling can vary significantly with the effective mass of the tunneling electrons in the nitride layer on the ON stack [1]. The tunneling current reductions in Figures 21 and 32 are consistent with a tunneling mass of 0.3 m_0 for the electrons in the annealed plasma-deposited nitride films. Experiments are under way to determine the

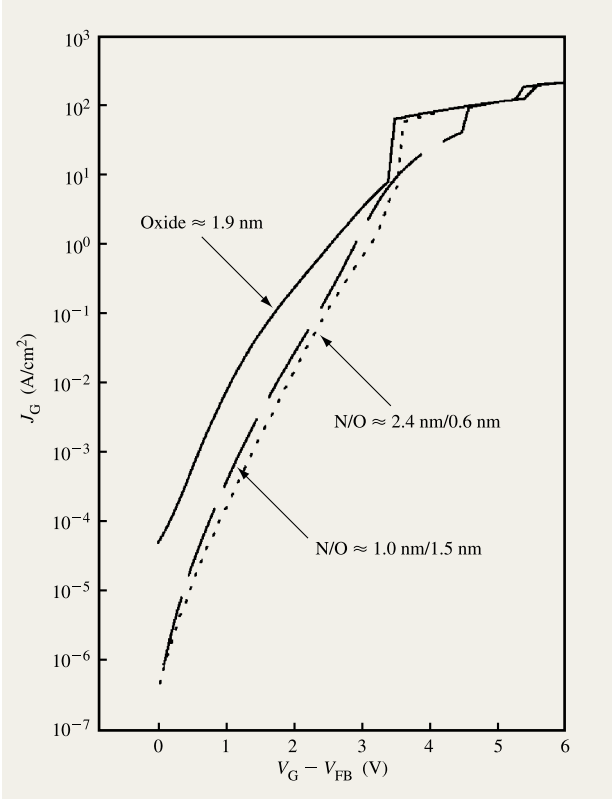

Figure 32

J_G–V_G curves for capacitors with ON stacked and single-layer oxide dielectrics with $t_{ox\text{-}eq} \approx 2.0$ nm. The tunneling currents for devices with ON dielectrics are reduced with respect to the device with the oxide dielectric.

tunneling mass in the Fowler–Nordheim tunneling regime.[12]

Summary

The results presented demonstrate that stacked ON gate dielectrics can play an important role in meeting SIA Technology Roadmap goals [42]. For example, stacked ON gate dielectrics possess all of the potential advantages for gate dielectric nitridation: 1) interface nitridation improves performance and reliability; 2) bulk nitride incorporation allows for increased physical thickness without increased oxide-equivalent thickness and decreased capacitance; and 3) top-surface nitride layers block B-atom transport out of boron-doped p^+ polycrystalline silicon gate electrodes. Preliminary studies have shown that ON gate dielectrics are compatible with elemental and compound metal gate electrodes such as TiN and WN [43].

On the basis of the research reported to date, stacked ON or O-oxynitride alloy structures with physical thicknesses of 2 to 2.5 nm and $t_{ox\text{-}eq} \approx 1.5$–$1.2$ nm could possibly meet SIA Technology Roadmap goals projected to at least the year 2009. The major issues will be the dependence of direct tunneling current, channel mobility, and reliability as a function of $t_{ox\text{-}eq}$ for both n-MOS and p-MOS devices. The feasibility of commercial remote-plasma-processing tools that can be used with 200-mm and 300-mm wafers to form stacked structures of the types discussed above is being addressed by several major tool manufacturers.

Finally, nitride layers or nitrided interfaces may be combined with alternate high-k dielectrics such as Ta_2O_5 and other transition-metal elemental and binary oxides to form stacked gate dielectric structures with $t_{ox\text{-}eq}$ extending to values of 1 nm and below.

Note added in proof

We have recently achieved a breakthrough in reducing direct tunneling currents in devices with nitrided oxides. This has been achieved by combining the interface nitridation process of [13] (see Figure 10) with stacked ON gate dielectrics of [16] (see Figure 32). Devices implementing either monolayer interface nitridation, as in Figure 10, or ON stacks, as in Figure 32, typically display reductions in direct tunneling of about one order of magnitude. The breakthrough has been achieved by including interface nitridation at the monolayer level with ON stacks, and by reducing the oxide-equivalent thickness well below 2 nm.

An analysis of calculations of direct tunneling current for single-layer and stacked dielectrics [1] indicates that tunneling current is approximately proportional to $\exp\{-t[m_{n^*}(E_b - V_{av})]^{0.5}\}$, where t is the physical thickness of the dielectric, m_{n^*} is an electron tunneling mass, and $E_b - V_{av}$ is an effective uniform barrier height for tunneling: E_b is the interfacial barrier height (e.g., the conduction-band offset energy between the Si substrate and the dielectric), and V_{av} is an effective reduction in that barrier height due to a potential drop of about 1 V across the dielectric. A study of direct tunneling in composite oxide/nitride gate stacks prepared by combining low-temperature plasma-assisted oxidation, oxide deposition, and nitride deposition, followed by postdeposition rapid thermal annealing [44], has shown that increases in physical thickness for these O/N stacks are significantly negated by decreases in the tunneling mass and effective barrier height for tunneling. For example, in the thickness range between 1.5 and 2.5 nm, reductions of tunneling associated with ~20% increases in physical thickness are ~10×, as shown in Figure 32, consistent with a reduced band offset (~2.1 eV) and tunneling mass (~0.25–0.3 m_0) for the nitride [1].

[12] G. Lucovsky and E. A. Irene, 1999.

Experiments supported by the extension of constraint theory to Si–dielectric interfaces have indicated high interfacial defect densities, particularly for p-MOS devices, so that nitrides cannot be directly substituted for oxides [41]. Other experiments have shown that controlled nitridation of the Si–dielectric interface provides a significant reduction of direct tunneling; e.g., monolayer nitridation produced by a plasma-assisted process results in reductions of about 10×, as shown in Figure 10(b) [45]. This note added in proof briefly identifies the effects of monolayer interface nitridation in combination with ON stacks in NON structures with monolayer interface nitridation, i.e., interfacial N-atom concentrations $\approx 7 \times 10^{14}$ cm^{-2}. For example, p-MOSFETs prepared in this way with $t_{ox\text{-}eq} \approx 1.6$ nm display direct tunneling currents at 1 V, J_{dt} (1 V), of about 5×10^{-3} A/cm^2. This current density is more than two orders of magnitude smaller than that of devices with single-layer oxides with the same $t_{ox\text{-}eq}$. In addition, these p-MOSFETs exhibit at most a 10% reduction in drive current, and excellent reliability; e.g., stressing for 5000 s at 3 V reduces g_m by about 3% and increases V_{th} by less than 8%. Consistent with the scaling model predictions, similar decreases in tunneling have also been achieved in devices with N-oxynitride stacks with $t_{ox\text{-}eq} \approx 1.3$ nm and J_{dt} (1 V) < 1 A/cm^2. Extrapolating from these results, NON stacks with monolayer interface nitridation will display direct tunneling leakage currents of 1 A/cm^2 for $t_{ox\text{-}eq} \approx 1.1 \approx 1.2$ nm. The implementation of these dielectrics in future generations of CMOS devices will be determined by two factors: the development of manufacturing equipment that can fabricate stacked gate dielectrics with the required nitrogen profiles, and the reliability of these devices as determined by accelerated stress testing.

Acknowledgments

The author wishes to acknowledge research support from the Office of Naval Research, the National Science Foundation, SEMATECH, and the Semiconductor Research Corporation. The author acknowledges the contributions to this research from his former and present graduate students and postdoctoral fellows: A. Banerjee, B. Claflin, S. Hattangady, B. Hinds, K. Koh, D. Lee, Z. Lu, Y. Ma, H. Niimi, C. Parker, P. Santos-Filho, D. Tsu, Y. Wu, H. Yang, and T. Yasuda. Finally, the author acknowledges helpful discussions and collaborations with faculty at North Carolina State University: J. Hauser, N. Masnari, V. Misra, C. Osburn, and J. Wortman.

References

1. H.-Y. Yang, H. Niimi, and G. Lucovsky, *J. Appl. Phys.* **83,** 2327 (1998).
2. E. M. Vogel, K. Z. Ahmed, B. Hornung, W. K. Henson, P. K. McLarty, G. Lucovsky, J. R. Hauser, and J. J. Wortman, *IEEE Trans. Electron Devices* **45,** 1350 (1998).
3. G. Lucovsky, D. V. Tsu, R. A. Rudder, and R. J. Markunas, in *Thin Film Processes II*, J. L. Vossen and W. Kern, Eds., Academic Press, Inc., Boston, 1991, p. 565.
4. G. Lucovsky, H. Niimi, K. Koh, D. R. Lee, and Z. Jing, in *The Physics and Chemistry of SiO_2 and the Si–SiO_2 Interface*, H. Z. Massoud, E. H. Poindexter, and C. R. Helms, Eds., The Electrochemical Society, Pennington, NJ, 1996, p. 441.
5. Z. Lu, M. J. Williams, P. F. Santos-Filho, and G. Lucovsky, *J. Vac. Sci. Technol. A* **13,** 607 (1995); G. Lucovsky, P. Santos-Filho, G. Stevens, Z. Jing, and A. Banerjee, *J. Non-Cryst. Solids* **198–200,** 19 (1996).
6. S. V. Hattangady, H. Niimi, and G. Lucovsky, *J. Vac. Sci. Technol. A* **14,** 3017 (1996).
7. S. C. Song, H. F. Luan, Y. Y. Chen, M. Gardner, J. Fulford, M. Allen, and D. L. Kwong, *IEDM Tech. Digest*, p. 373 (1998).
8. T. P. Ma, *IEEE Trans. Electron Devices* **45,** 680 (1998).
9. X. Guo and T. P. Ma, *IEEE Electron Device Lett.* **19,** 207 (1998).
10. D. R. Lee, G. Lucovsky, M. R. Denker, and C. Magee, *J. Vac. Sci. Technol. A* **13,** 607 (1995); D. R. Lee, C. R. Parker, J. R. Hauser, and G. Lucovsky, *J. Vac. Sci. Technol. B* **13,** 1778 (1995).
11. M. L. Green, D. Brasen, K. W. Evens-Lutterodt, L. C. Feldman, K. Krisch, W. Lennard, H. T. Tang, L. Manchanda, and M. T. Tang, *Appl. Phys. Lett.* **65,** 848 (1994).
12. J. Ahn, J. Kim, G. Q. Lo, and D.-L. Kwong, *Appl. Phys. Lett.* **60,** 2089 (1992).
13. H. Niimi and G. Lucovsky, *Proceedings of the 1998 International Conference on Characterization and Metrology for ULSI Technology*, D. G. Seiler, A. C. Diebold, W. M. Bullis, T. J. Shaffner, R. McDonald, and E. J. Walters, Eds., The American Institute of Physics, Woodbury, NY, 1998, p. 273.
14. Y. Ma, T. Yasuda, and G. Lucovsky, *J. Vac. Sci. Technol. A* **11,** 952 (1993); Y. Ma, T. Yasuda, S. Habermehl, and G. Lucovsky, *J. Vac. Sci. Technol. B* **11,** 1533 (1993); Y. Ma and G. Lucovsky, *J. Vac. Sci. Technol. B* **12,** 2504 (1994).
15. S. V. Hattangady, H. Niimi, and G. Lucovsky, *Appl. Phys. Lett.* **66,** 3495 (1995); S. V. Hattangady, R. Kraft, D. T. Grider, M. A. Douglas, G. A. Brown, P. A. Tiner, J. W. Kuehne, P. E. Nicollian, and M. F. Pas, *IEDM Tech. Digest*, p. 495 (1996).
16. Y. Wu and G. Lucovsky, *Proceedings of the 1998 International Conference on Characterization and Metrology for ULSI Technology*, D. G. Seiler, A. C. Diebold, W. M. Bullis, T. J. Shaffner, R. McDonald, and E. J. Walters, Eds., The American Institute of Physics, Woodbury, NY, 1998, p. 288.
17. C. R. Parker, G. Lucovsky, and J. R. Hauser, *IEEE Electron Device Lett.* **19,** 106 (1998).
18. Y. Wu and G. Lucovsky, *IEEE Electron Device Lett.* **19,** 367 (1998).
19. K. Koh, H. Niimi, G. Lucovsky, and M. L. Green, *Jpn. J. Appl. Phys.* **37,** 709 (1998).
20. G. Lucovsky, S. S. Kim, D. V. Tsu, G. G. Fountain, and R. J. Markunas, *J. Vac. Sci. Technol. B* **7,** 861 (1989).
21. S. S. Kim and G. Lucovsky, *J. Vac. Sci. Technol. A* **8,** 2039 (1990).
22. T. Yasuda, Y. Ma, S. Habermehl, and G. Lucovsky, *Appl. Phys. Lett.* **60,** 434 (1992).
23. E. H. Nicollian and A. Reisman, *J. Electron. Mater.* **17,** 213 (1988).
24. P. Thanikasalam, T. K. Whidden, and D. K. Ferry, *J. Vac. Sci. Technol. B* **14,** 2840 (1996).

25. Y. Ma, T. Yasuda, and G. Lucovsky, *Mater. Res. Soc. Symp. Proc.* **303,** 325 (1993).
26. G. Lucovsky, A. Banerjee, B. Hinds, B. Claflin, K. Koh, and H. Yang, *J. Vac. Sci. Technol. B* **15,** 1074 (1997).
27. B. C. Smith and H. H. Lamb, *J. Appl. Phys.* **83,** 7635 (1998).
28. C. H. Bjorkman, C. E. Shearon, Jr., Y. Ma, T. Yasuda, G. Lucovsky, U. Emmerichs, C. Meyer, K. Leo, and H. Kurz, *J. Vac. Sci. Technol. A* **11,** 964 (1993); C. H. Bjorkman, T. Yasuda, C. E. Shearon, Jr., U. Emmerichs, C. Meyer, K. Leo, and H. Kurz, *J. Vac. Sci. Technol. B* **11,** 1521 (1993).
29. H. Niimi, Ph.D. Thesis, Department of Materials Science and Engineering, North Carolina State University, Raleigh, 1998.
30. G. Lucovsky, Y. Ma, S. S. He, T. Yasuda, D. J. Stephens, and S. Habermehl, *Mater. Res. Soc. Symp. Proc.* **284,** 34 (1993); G. Lucovsky, S. S. He, M. J. Williams, and D. Stephens, *Microelectron. Eng.* **25,** 329 (1994) and references therein.
31. G. Lucovsky and J. C. Phillips, *J. Non-Cryst. Solids* **227,** 1221 (1998), and references therein.
32. Y. Ma, T. Yasuda, and G. Lucovsky, *Appl. Phys. Lett.* **64,** 2226 (1994).
33. C. J. Parker, Ph.D. Thesis, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, 1998.
34. Y. Wu and G. Lucovsky, *Proceedings of the IEEE International Reliability Physics Symposium*, 1998, p. 70.
35. C. C. Chang, *Surf. Sci.* **48,** 9 (1975).
36. S. Lo, D. Buchanan, Y. Taur, and W. Wang, *IEEE Electron Device Lett.* **18,** 209 (1997).
37. L. Han, D. Wristers, J. Yan, M. Bhat, and D. Kwong, *IEEE Electron Device Lett.* **16,** 319 (1995).
38. R. Fair, *IEEE Electron Device Lett.* **18,** 244 (1997).
39. S. Wolf, in *Silicon Procession for the VLSI Era*, Vol. 3, Lattice Press, Sunset Beach, CA, 1995, p. 311.
40. J. Pfiester and F. Baker, *IEEE Electron Device Lett.* **11,** 247 (1990).
41. G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, *Appl. Phys. Lett.* **74,** 2005 (1999).
42. *The National Technology Roadmap for Semiconductors*, Semiconductor Industry Association, San Jose, CA, 1997.
43. B. Claflin, M. Binger, and G. Lucovsky, *J. Vac. Sci. Technol. A* **16,** 1757 (1998).
44. G. Lucovsky, *J. Vac. Sci. Technol. A* **16,** 356 (1998).
45. H. Niimi and G. Lucovsky, *J. Vac. Sci. Technol. B* **17** (1999), in press.

Received March 27, 1998; accepted for publication April 14, 1999

Gerald Lucovsky *Departments of Physics, Electrical and Computer Engineering, and Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina 27695 (gerry_lucovsky@ncsu.edu).* Dr. Lucovsky received his B.Sc. and M.A. degrees in physics from the University of Rochester, New York, in 1956 and 1958, respectively. In 1960 he received his Ph.D. in physics from Temple University, Philadelphia, Pennsylvania. He spent 21 years in industry, working at the Philco Corporation Research Laboratories in Philadelphia and Blue Bell, Pennsylvania, from 1958 through 1965, and at the Xerox Corporation Research Centers in Webster, New York, and Palo Alto, California, from 1965 through 1979. Currently Dr. Lucovsky is a University Professor of Physics at North Carolina State University in Raleigh, where his research activities include deposition and characterization of the thin films for semiconductor device applications, with a focus on low-temperature, low-thermal-budget processing for advanced silicon and compound semiconductor devices.