SAND2011-0910C

Dynamic Negative Bias Stress Instability Effects in Hafnium Silicon Oxynitride and Silicon Dioxide

J. K. Mee[a], R. A. B. Devine[a,b], H. P. Hjalmarson[c] and K. Kambour[c]

[a]AFRL/RSVE, Kirtland AFB, NM 87117
EMail: Jesse.Mee@kirtland.af.mil
[b]EMRTC/NMT, 801 Leroy Place, Socorro, NM 87801
[c]Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

Negative bias temperature instability (NBTI) is an issue of critical importance as the space electronics industry evolves because it may dominate the reliability lifetime of space based assets. Understanding its physical origin is therefore essential in determining how best to search for methods of mitigation. It has been suggested that the magnitude of the effect is strongly dependent on circuit operation conditions (static or dynamic modes). In the present work, we examine the time constants related to the charging and recovery of trapped charged induced by NBTI in HfSiON and SiO_2 gate dielectric devices at room temperature.

Introduction

The issue of reliability in advanced electronics for space based assets is rapidly superseding that which is traditionally assumed to be the most important: radiation effects. The primary reason is that the more advanced technologies have lifetimes which are becoming substantially shorter whereas older technologies had perfectly acceptable reliability lifetimes (> 10 years). Though this may be acceptable in the terrestrial commercial scenario given that the general public accepts the concept of "outmoding", which hides the reliability issue to some extent, this is not an acceptable logic for the military and space avionics.

Negative bias temperature instability (NBTI) is one of a variety of mechanisms [1] leading to failure of microelectronics circuits based upon Si technology. In particular, it is believed to be one of the most important factors in defining reliability lifetime in the more advanced technologies which already/will incorporate newer materials such as the "high-κ" gate dielectrics in field effect transistors. Though the full physics of the bias temperature instability phenomena (NBTI and positive PBTI) remains unclear, it has been recognized and demonstrated that there are dynamic and "static" components which can influence the operation of a micro-circuit in different ways [2]. Accurate modeling of the phenomenon of NBTI must therefore take careful consideration of the dynamic effects related to rapid charge capture and release at the silicon/gate dielectric interface. The dynamic effects, at least in PBTI [3], can occur very rapidly (<1μs) upon the application or removal of a stressing bias at

the gate electrode. As a result, such effects have largely gone unnoticed due to the inadequacies of the measurement equipment to detect such rapid changes.

Given the evolving interest in measurement of dynamic effects, we have acquired a rapid data acquisition system (Keithley 4200 SCS), and begun a study of NBTI induced charging and recovery with sub-microsecond stressing/recovery intervals between microsecond measurements giving us maximum resolution in the early time domain. These initial experiments have been performed on metal-oxide-semiconductor field effect transistor (MOSFET) devices with high-κ dielectric and SiO_2 gate insulators. Our initial results already reveal that in the early phase, and at room temperature, we have a fully recoverable mechanism related to NBTI. Preliminary theoretical calculations have been carried out to confirm the origin of this mechanism.

Experiment Methodology

The high-κ MOSFET devices used in the experiments reported here were primarily p-channel with 250, 300, and 350 nm channel lengths. The gate dielectric stacks were formed of chemically oxidized Si (approximately 1 nm of SiO_2) upon which HfSiO was deposited using the technique of atomic layer deposition (ALD). The stack was then nitrided using the process of post-deposition annealing in NH_3. The metal electrodes were formed of physical vapor deposited (PVD) TaN. A conventional CMOS flow completed the fabrications process. Some additional devices were used from IBM's 130nm process which contained a nominal 3.4 nm SiO_2 gate oxide.

Device characteristics [4] were obtained using the Keithley 4200 SCS system equipped with a new state-of-the-art pulse generator. We have developed two types of experiments; the first monitors rapid charging and recovery mechanics, the second monitors charging under AC stress. Schematic pulse trains for the charging/recovery stress and for the AC test are shown in Figs. 1a and 1b respectively. Following a procedure adopted previously [5], in all cases we performed single point source/drain current (I_{ds}) measurements for a chosen gate-source voltage (V_{gs}) and a chosen drain-source voltage (V_{ds}).

Charging/Recovery Stress Procedure

HfSiON and SiO_2 devices were subjected to a continuous stress stressing bias on the gate followed by a recovery time. More specifically, each device was stressed at V_{bias} (V) while I_{ds} was periodically probed using a 1μs pulse at V_{gs} = -1.0 V, and V_{ds} = -0.2 V. The rise and fall times of the measurement pulse were ~ 20 ns. Although the measurement pulse duration was 1 μs, the source-drain current was actually measured during the time window from 750ns to 880ns. This window ensured that ringing and/or overshoot effects had relaxed and the current actually represented the mean current. With a sampling rate ~ 200 M samples per second, we average 26 individual measurements in the 130 ns measurement window. The frequency of the data acquisitions was logarithmic in time with the first and second points occurring at t = 0 and t = 5 μs. Following the stressing period, which amounted to a total of 100 s, the devices were allowed to recover at a selected voltage V_{rec} (V). As with the charging curve, I_{ds}

measurements were taken starting with μs resolution and going out to 100 s. For the work presented here, the recovery was done at 0 V on the gate/source/drain and body contacts. For the charging curves, we stressed at V_{bias} = -2.0 V.

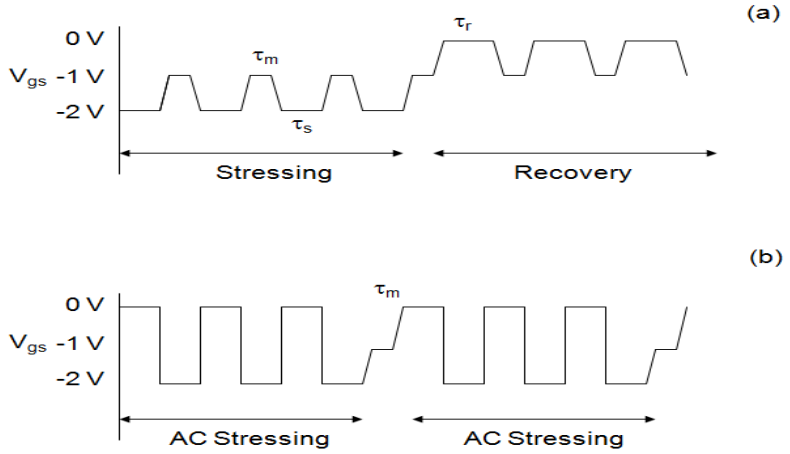


Figure 1. (a) The V_{gs} pulse train for charging/recovery data. (b) V_{gs} pulse train for the AC stress data.

AC Stress Procedure

Additional HfSiON and SiO_2 devices were subjected to an AC stressing bias on the gate with a pulse top of 0V and a pulse bottom of -2 V (50% duty cycle). During this stressing the source, drain and body contacts were held at 0 V. Seven frequencies were used in the range 2.5 MHz to 1 Hz. Some additional devices were stressed with a constant bias which provided a reference DC curve. Upon completing a stressing period, which will contain 'n' number of stressing cycles related to frequency and duration of the AC stress, a 1 μs pulse at -1.0 V is applied to the gate and an I_{ds} measurement is extracted. Again, V_{ds} = –0.2 V which ensured the devices operated in the linear regime [4]. Note the length of the stressing periods should be an integral number of single stressing cycles. Otherwise, it is possible to create a scenario whereby a single stressing cycle is longer than the requested stress time.

In all cases, both the HfSiON and SiO_2 devices were stressed at 25 °C. Previous attempts to measure these dynamics have been approximate because it is hard to extract the data without significantly perturbing the system. We are optimistic that the 1 μs measurement pulse, τ_m, that we have incorporated here has minimized the impact of charging variation during the data acquisitions; however, it is believed that even at this level we may still be offset by very energetically and or spatially close traps which capture and release faster that we can measure.

Interpretation of Single Point Measurement

We are primarily interested in the device threshold voltage, V_{th}, as the relevant parameter for the MOSFET degradation. Usually such a determination requires acquisition of a full $I_{ds}(V_{gs})$ characteristic [4] for fixed V_{ds}, but such measurement requires a finite time during which the stressing voltage is removed or changed. We have suggested previously [5] that to a good approximation one can use a single point measurement, I_{ds} (V_{gs} fixed, V_{ds} fixed) to obtain an estimate of the V_{th}. In the linear regime ($|V_{ds}| << |V_{gs} - V_{th}^{o}|$),

$$I_{ds} = [W/2L]\, C_{ox}\, \mu[\{V_{gs} - (V_{th}^{o} + \Delta V_{th})\}V_{ds} - 1/2V_{ds}^{2}] \ , \qquad [1]$$

where W is the channel width, L the channel length, C_{ox} the gate dielectric stack capacitance, and μ is the mobility of the inversion channel carriers (holes). This can be reduced to the simplified expression for ΔV_{th} as a function of I_{ds}/I_{ds}^{o}:

$$\Delta V_{th} = [V_{gs} - V_{th}^{o}](1 - I_{ds}/I_{ds}^{o}) \ . \qquad [2]$$

This method requires certain approximations with respect to electric field dependent mobility, all of which have been justified experimentally [5,6]. A greater issue with the approach revolves around the uncertainty in determination of the initial current value, I_{ds}^{o}. With previous equipment (e.g. HP 4156), our first current measurement was not at time zero but rather at approximately 0.5 – 1 second. Our initial intuition was to make a linear fit to the first few $I_{ds}(t)$ points and extrapolate back to t = 0. The introduction of the new Keithley 4200 SCS pulse card has given us the ability to measure I_{ds} at t = 0, and to monitor $I_{ds}(t)$ with microsecond resolution. We have since learned (Fig. 3) that the early time behaves much more like an exponential. This means that previous methods introduced a large underestimation of the I_{ds}^{o} value that translates into an underestimation of the full magnitude of the NBTI effect.

Experimental Results and Discussion

It is suspected that the magnitude and rate of charge trapping would be much greater in HfSiON than in SiO_2. This suspicion is based upon the assumption that there is a high defect density at the Si/dielectric interface in HfSiON as compared to the 'renowned" SiO_2/Si interface. The curves shown in Fig. 2 are the result of room temperature continuous stress experiments out to 300 seconds for both high-k and standard SiO_2. As anticipated, SiO_2 devices degrade significantly less than HfSiON for a nominally equivalent stressing voltage on the gate electrode and equivalent physical thickness of dielectric. We furthermore verified our data interpretation methodology by performing full $I_{ds}(V_{gs})$ measurements and by acquiring a single I_{ds} point during each measurement cycle. For the former, we apply a staircase waveform to the gate and take an I_{ds} measurement at the top of each V_{gs} step. The result is a complete $I_{ds}(V_{gs})$ characteristic in ~ 10 μs. We can then extract $\Delta V_{th}(t)$ using classical "exact methods" [4], or from a single point on this curve by applying Eqn. 2. As seen in Fig. 2, ΔV_{th} determinations using both methods track well within experimental error.

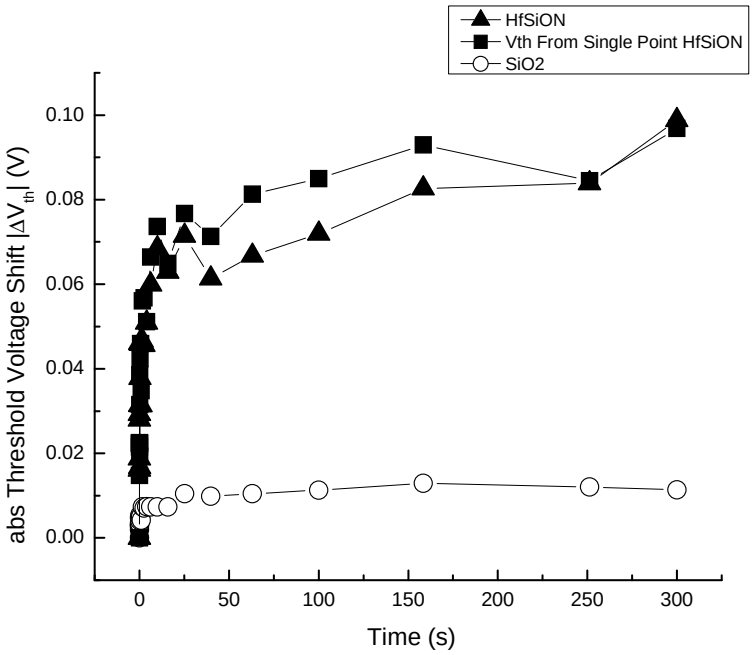


Figure 2. Room temperature continuous stress measurements at V_{gs}=-2.0 V for SiO_2 (circles), and HfSiON (Squares, triangles). Two methods of ΔV_{th} determination were employed for the HfSiON data; (triangles): ΔV_{th} determination from complete $I_{ds}(V_{gs})$ curve, (squares): ΔV_{th} determination from a single point on the $I_{ds}(V_{gs})$ curve using Eqn. 2 with $I_{ds}(t)$ measured at V_{gs} = - 1.0 V.

As observed in Fig.2, charging clearly begins very rapidly upon application of a stressing bias to the gate of a MOSFET. Using very short duration stressing pulses followed by rapid single point measurements, we have examined the magnitude of this effect in multiple devices. Fig. 3 shows the charging and recovery behavior of the threshold voltage as a result of room temperature stressing on HfSiON technology – the effects are emphasized by the logarithmic time plot for the total stress or recovery time.

All three curves in Fig. 3a and 3b are the result of threshold voltage degradation under identical stress conditions on the same device. A 5 minute recovery time was allotted in between stressing experiments (Fig. 3a). Note that the overall stressing times are short enough that apparently these samples do not enter the permanent degradation regime generally associated [7] with NBTI. Clearly a significant amount of V_{th} degradation occurs before 1 second. As mentioned previously, this information would have been lost if this data was collected using previous techniques [5]. Examination of Fig. 3 reveals that the time constants for charging are different than those for recovery. More specifically, it appears that the recovery dynamic is more rapid than the charging dynamic. This asymmetry has very interesting implications for the reliability lifetime characterization because it means that a

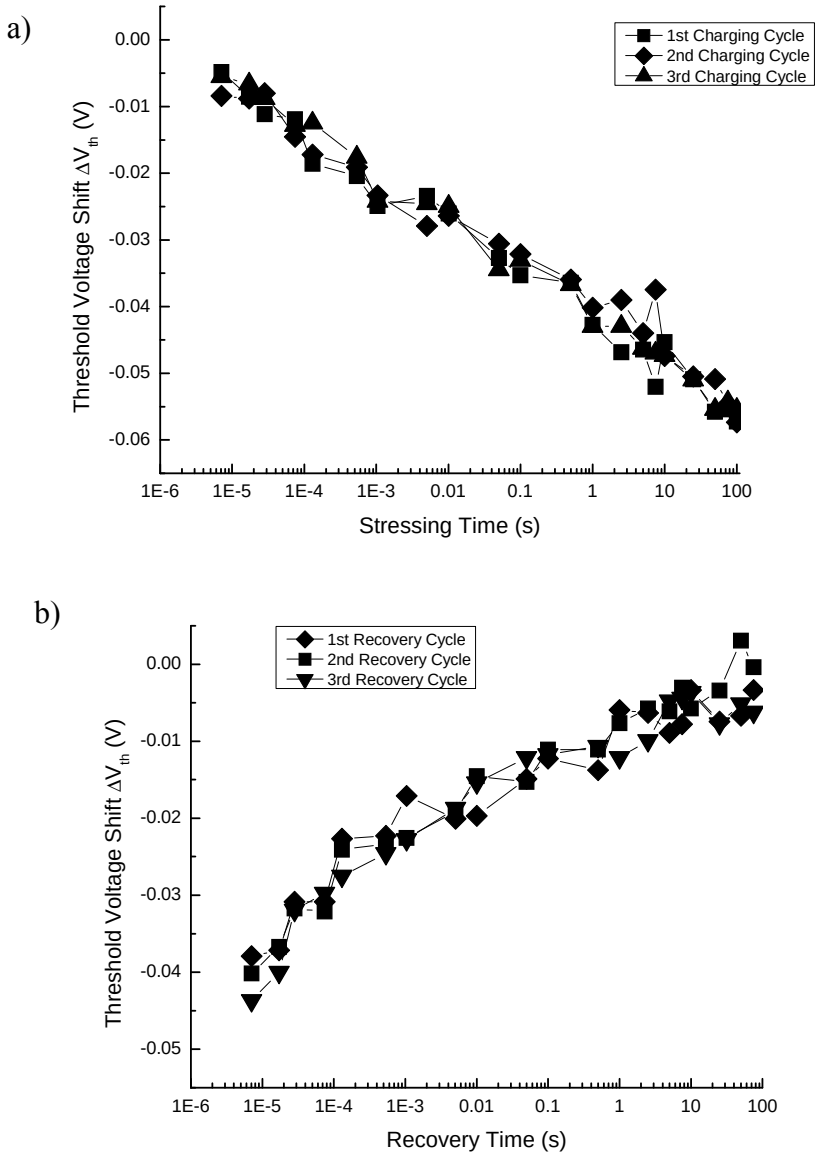
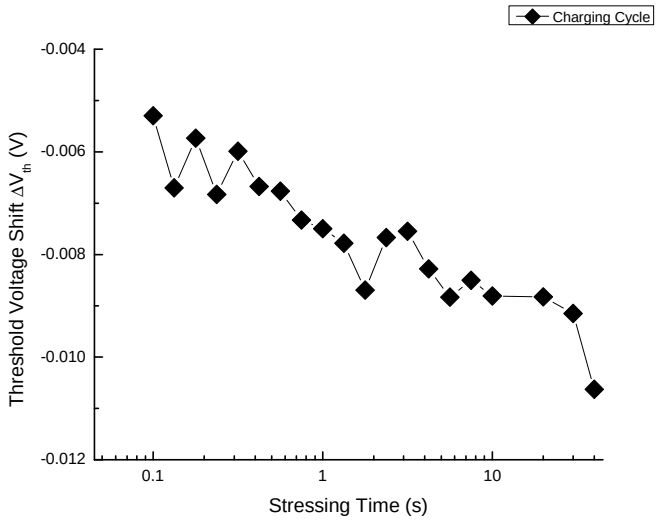


Figure 3. Room temperature continuous stress measurements on HfSiON devices for a) charging at V_{gs}=-2.0 V and b) recovery at V_{gs}=0 V. $\Delta V_{th}(t)$ is calculated by applying equation (2) to single I_{ds} measurements at V_{gs}=-1.0V.

device which is stressed with an AC single pulse at 50% duty cycle should show no net threshold voltage degradation under NBTI stressing. Note that these results are at 25^{o}C; we have not yet determined how curves will behave under elevated temperature stressing.However, one research group [3] has already suggested that the early time behavior is temperature independent – at least for PBTI.

a)



b)

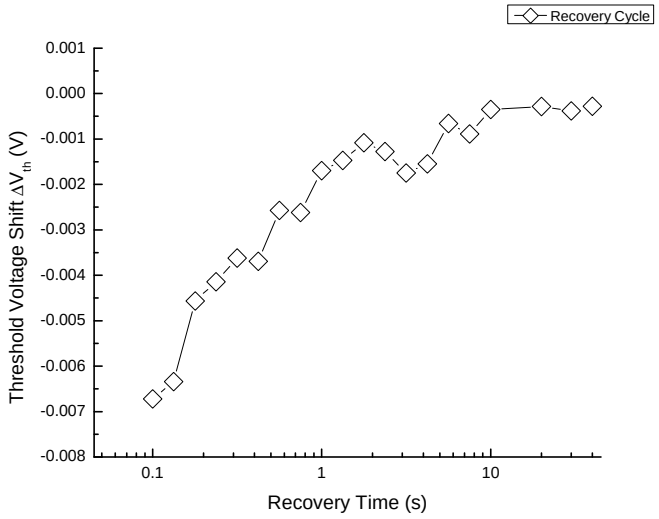


Figure 4. Room temperature continuous stress measurements on SiO_2 devices for a) charging at V_{gs}=-2.0 V and b) recovery at V_{gs}=0V. As with the HfSiON curves, $\Delta V_{th}(t)$ is calculated by applying equation (2) to single I_{ds} measurements at V_{gs}=-1.0V.

Results for SiO_2 are shown in Fig. 4 and are similar to those of the high-κ samples but with obvious variation in the magnitude of the effect. Again, the recovery dynamic appears to be the faster than the charging dynamic. Note that the time base for the SiO_2 experiments is less resolved in the early time domain. The scatter in the sub 1 ms time regime on these devices can be associated with slower charging times for SiO_2.

In light of the observed charging and recovery time constants, we performed a series of NBTI stress experiments in which AC stress was applied at different frequencies for chosen

total stress times, and the associated threshold voltage shift then measured (Fig. 1b). These experiments should have simulated the effect of alternating stress which we anticipated would be non-zero and cumulative [2]. In Fig. 5, we show the results of AC stressing at 7 different frequencies ranging from 2.5 MHz to 1 Hz, and include the continuous stress measurement (DC) for HfSiON devices. Similar data for SiO_2 based devices is shown in Fig. 6. Measurement points are equally spaced in linear time to reduce confusion related to the minimum stressing period. The Keithley pulsed system in fact chooses a stress time equal to an integral number of stress pulses so that the actual point in time at which the AC stress stopped and V_{th} measurement began corresponded to the negative maximum of the stressing pulse (see Fig. 1).

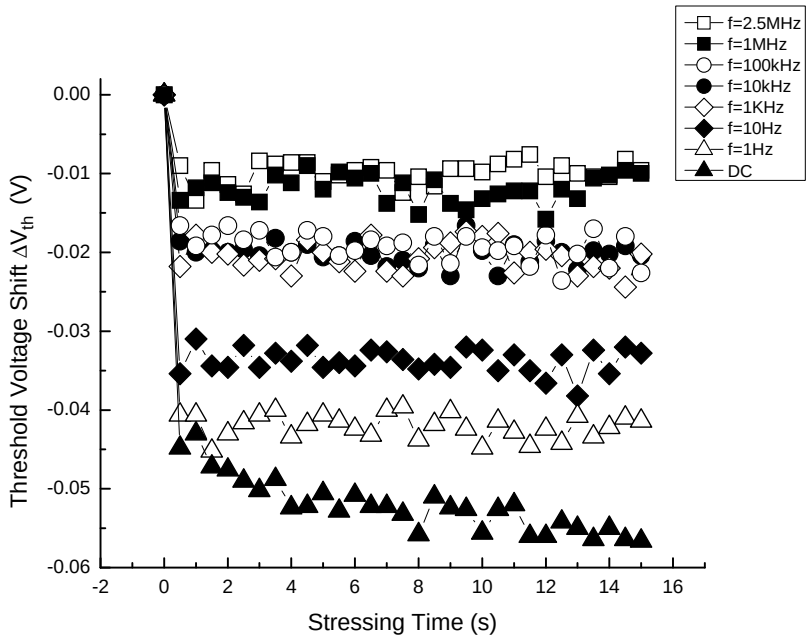


Figure 5. Room temperature AC stress measurements on HfSiON devices. The AC pulse top was 0 V, the pulse bottom was -2.0 V. Frequencies ranged from 2.5 MHz to 1 Hz. The DC curve (Full, up triangles) is shown for reference.

As seen in Fig. 5 the "net" threshold voltage shift varied significantly with the frequency of the AC stressing pulses. However, it would appear that for both HfSiON and SiO_2 based devices, there is no accumulation of threshold voltage shift with total stressing time prior to measurement. This result appeared to us to be rather confusing initially but became clearer. Irrespective of the pulse frequency, for each pulse there was a more or less short relaxation time followed by an equal charging time and then finally followed by rapid measurement of I_{ds}. In fact, it appears that the ΔI_{ds} is induced by the final stressing pulse of the sequence determined by the total stressing time; the other pulses comprising the total pulse chain appear irrelevant. The difference in amplitude of ΔV_{th} shown in Figs. 5 and 6 as a function

of frequency is then explained. For a 1 Hz pulse there is an effective stressing time of 0.5 seconds while for a 1 MHz pulse the stressing time is 5 x 10^{-7} seconds. From Figs. 3 and 4, one clearly understands that very short stressing times induce almost negligible threshold voltage shifts while much longer times (e.g. 0.5 seconds) induced threshold voltage shifts almost comparable to those observed for a continuous, DC bias, situation.

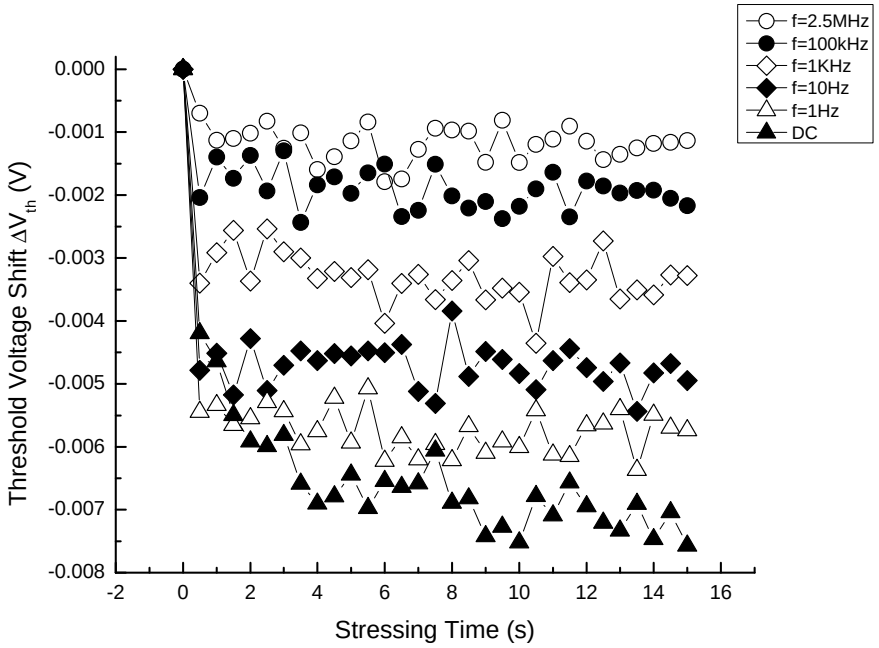


Figure 6. Room temperature AC stress measurements on SiO_2 technology. In this case the frequencies ranged from 2.5 MHz to 10 Hz. Again, the DC curve is shown for reference.

The data presented above reveals the presence of charge trapping and detrapping phenomena in both SiO_2 and HfSiON based devices which can take place at room temperature and can be readily cycled backwards and forwards. Cyclic bias stressing for the frequencies we have used (maximum 2.5 MHz) indicate that there is no net NBTI induced shift in the threshold voltage of the device, at least at room temperature. Preliminary theoretical modeling has been carried out to probe the origin of the observed NBTI effect.

Theoretical Calculations and Discussion

The role of defects in the gate layer was investigated by performing calculations of the threshold voltage shifts for some simple models. The basic physics involves tunneling between the Si substrate and defects in the oxide region. These calculations were performed on a representative one-dimensional structure. In principle, these calculations can mimic the

transient gate voltages applied during the measurements of the threshold voltage shifts. Such calculations are used to obtain the density of trapped charge in the HfSiON layer. From the trapped charge density, the threshold voltages shifts are also obtained.

The simplified bandstructure we have assumed is shown in Fig. 7. This figure shows the conduction and valence bands of the constituent materials. Proceeding from the left, this figure shows the heavily doped Si region that serves as the gate electrode in these calculations. The oxide layer is composed of a 3 nm-thick HfSiON layer and a 1 nm-thick SiO_2 layer. To the right is the lightly doped p-type Si substrate In these calculations, the gate contact has an n-type density of 10^{19} cm^{-3} and the substrate has a p-type density of 10^{16} cm^{-3}.

The transient electrical effects are computed using the radiation effects on semiconductors (REOS) [8] program to solve the kinetic equations for the electrons, holes, and the defect densities. The tunneling current is obtained from a reaction of the form

$$T^0 + p \leftrightarrow T^+ \qquad [3]$$

in which the holes p in the substrate tunnel to neutral traps T^0 in the HfSiON layer to produce positively-charged traps T^+. These positively-charged traps contribute to a threshold voltage shift.

The tunneling reaction leads to a rate equation for the trapped charge:

$$\mathrm{d}[T^+] / \mathrm{dt} = k_f [T^0][p] - k_r[T^+] \qquad [4]$$

in which

$$k_f = \alpha \exp(-\beta | x_f - x_i |) \qquad [5]$$

and

$$k_r = n_{th} k_f . \qquad [6]$$

In these equations, the physical parameters have the values $\alpha = 10^{13}$ sec^{-1} and $\beta = 5 \times 10^7$ cm^{-1} [8]. However, for the present calculations, the case of non-resonant tunneling must also be included. In these cases, the energy of the holes is conserved by emission and absorption of phonons. The absorption process causes the trapping rate to become much reduced.

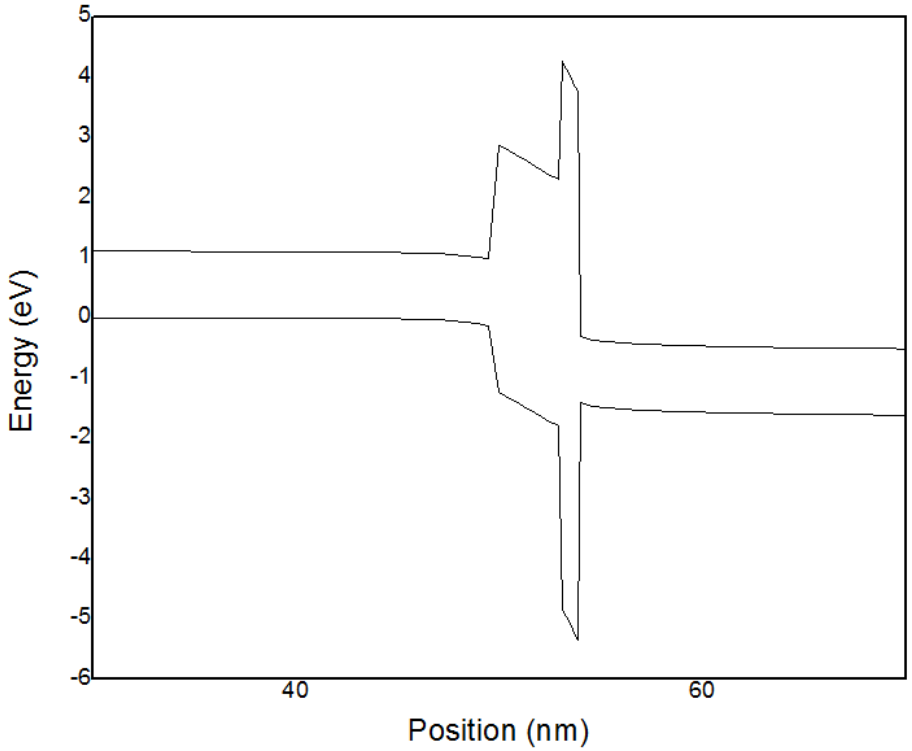


Figure 7: Shows the conduction and valence bands as a function of position for the gate-stack structure.

The kinetics explored in the experiments can be understood by using these equations. The filling of the traps is governed by the forward rate constant k_f, and the release is governed by the reverse rate constant k_r. Both of these quantities depend strongly on the tunneling distance $d = | x_f - x_i |$. For these calculations, a uniform distribution of traps in the HSiON is assumed. The varying distance d leads to a distribution of forward and reverse rates defined by the distance of the trap from the Si substrate.

The computed shift in threshold voltage as a function of time is shown in Fig. 8. This voltage shift is caused by the increase in trapped holes in the HfSiON layer that have tunneled from the Si substrate into the empty traps. The characteristic shape of the threshold voltage is in good agreement with the data shown in the experimental section. The shift in threshold voltage as the holes are released shows similar but faster kinetics.

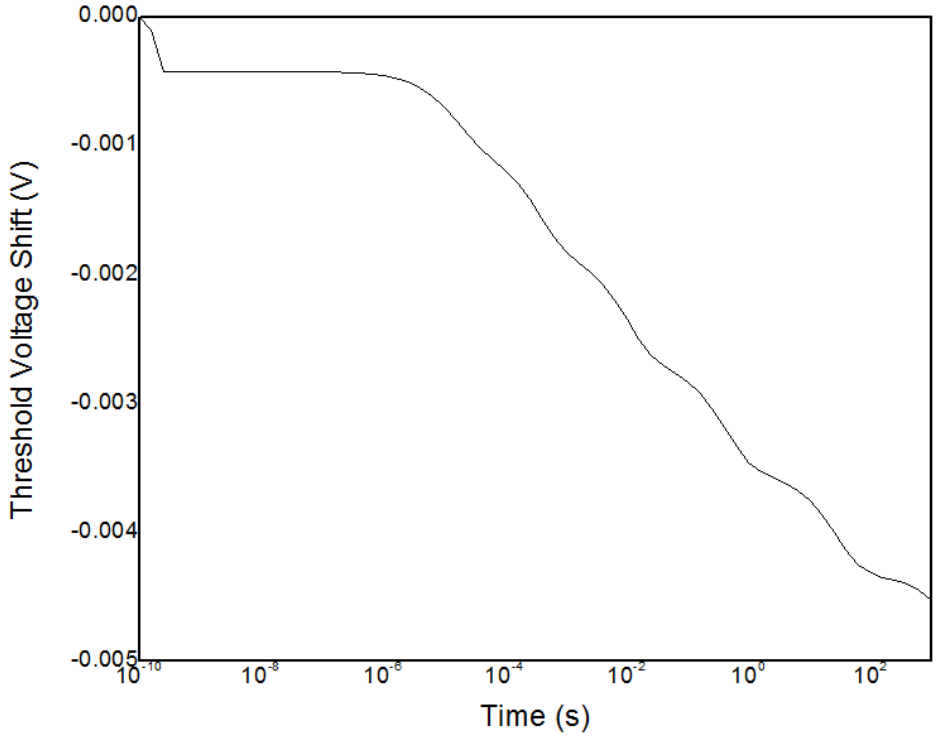


Figure 8: Shows the shift in threshold voltage as the traps are filled in the HfSiON layer.

The kinetics can be understood by doing a simple calculation for traps located at one location a distance *d* from the substrate. The trapped hole density at a time τ_f for traps at a distance *d* from the substrate can be written as

$$[T^+(\tau_f)] = T^0 (1 - \exp(-k_f \tau_f)) \qquad [7]$$

in which

$$k_f = k_0 \exp(-\beta d). \qquad [8]$$

In the calculations shown in Fig. 8, each sheet of traps contributes to the total trap density. The shift shown in this figure is the sum of these contributions at the various trap locations. The fact that there is a distribution leads to the log(t) behavior seen in the data and the calculations.

The results of calculations shown in Fig. 8 are to be compared with the experimental values shown in Fig. 3a. Although the absolute magnitude is not reproduced, the time dependent behavior is relatively respected. To understand the data obtained from the pulse train experiments, one can once again assume that only a sheet of traps is involved. First one can assume that a filling pulse has been applied for a time τ_f to produce filled traps governed

by the previous equations. If now the gate pulse is reduced for another period of time τ_r, the trapped hole density becomes:

$$[T^+(\ \tau_r)] = [T^+(\ \tau_f)]\ \exp(-k_r\ \tau_f) + [T^+(\infty)] \qquad [9]$$

This expression follows because the kinetics are now governed by the reverse rate, but at long times the trap density falls to a value determined by the new gate bias (in this case 0 V). In these formulae, the rates are k_f and k_r, respectively. Combining these expressions, the final form for the filling of these traps with holes and then releasing the holes is obtained:

$$[T^+(\ \tau_r)] = T^0(1 - \exp(-k_f\ \tau_f))\ \exp(-k_r\ \tau_r) + [T^+(\infty)] \qquad [10]$$

The factor $[T^+(\infty)]$ is the trap density at long times. To understand this expression, it is useful to assume that the time durations are small. Also, one can assume that the long time density is very small. Then

$$[T^+(\ \tau_r)] = T^0(1 - \exp(-k_f\ \tau_f))\ \exp(-k_r\ \tau_r)$$

$$= T^0 k_f\ \tau_f\ (1 - k_r\ \tau_r) \qquad [11]$$

This expression shows that pulse times that are longer will lead to larger shifts in qualitative agreement with the data. Further calculations for distributions of traps are being performed.

Summary

Experimental data confirms that at very short times, reversible charging and discharging can occur leaving no residual threshold voltage shift. The direction of the threshold voltage shift is consistent with the accumulation of positive charge. It is probable that in most experiments performed to date, this term has been ignored because of instrumental limitations. The preliminary calculations support the hypotheses that holes are indeed the positively charged species are involved in the trapping and they originate in the silicon substrate. This conclusion follows from the successful use of the well-known mechanism of hole tunneling to explain the data. The next phase of these experiments will involve measurements as a function of temperature while the theoretical effort will center upon a more quantitative physics mechanisms.

Acknowledgments

This work has been partially supported by the Defense Threat Reduction Agency (DTRA – Dr. L. Palkuti) through a contract with North Carolina State University. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. M. White and J. B. Bernstein, JPL Publication 08-5 2/08 (2008)
2. G. Chen, K.Y. Chuah, M. F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin and D. L. Kwong, Proc. IRPS 196-202 (2003)
3. D. Heh, C. D. Young and G. Bersuker, IEEE Electron Dev. Lett., **29** 180-182 (2008)
4. S. M. Sze,"Physics of semiconductor devices" (Wiley, N.Y. 1981) Chapt. 8
5. J. K. Mee and R. A. B. Devine, J. Appl. Phys. **107** 024511 (2010)
6. S. C. Sun and J. D. Plummer, IEEE Trans. Electron. Devices **27** 1497 (1980)
7. D. K. Schroder and J. A. Babcock, J. Appl. Phys. **94** 1 (2003)
8. H. P. Hjalmarson, R. L. Pease and R. A. B. Devine, IEEE Trans. Nucl. Sci., **55** 3009 (2008)
9. T.R. Oldham, A. J. Lelis, and F. B. McLean, IEEE Trans. Nucl. Sci., **33** 1203 (1986)