Published in final edited form as:

IEEE Trans Electron Devices. 2017 December ; 64(12): 5099–5016. doi:10.1109/TED.2017.2764804.

Local field effect on charge-capture/emission dynamics

Kin P. Cheung [Senior Member, IEEE], Dmitry Veksler [Member, IEEE], and Jason P. Campbell

Engineering Physics Division, National Institute of Standards and Technology, Gaithersburg, MD 20899 USA

Abstract

Charge-capture/emission is ubiquitous in electron devices. Its dynamics often play critical roles in device operation and reliability. Treatment of this basic process is found in many text books and is considered well understood. As in many electron device models, the individuality of immobile charge is commonly replaced with the average quantity of charge density. This has worked remarkably well when large numbers of individual charges (ensemble) are involved. As device geometries become very small, the ensemble “averaging” becomes far less accurate. In this work, the charge-capture/emission dynamic of Metal-Oxide-Semiconductor-Field-Effect-Transistor (MOSFET) is re-examined with full consideration of individual charges and the local field in their immediate vicinity. A dramatic modification of the local band diagram resulted, forcing a drastic change in emission mechanism. The implication is that many well-understood phenomena involving charge capture/emission will need to be reconsidered. As an example, this new picture is applied to the random telegraph noise (RTN) phenomenon. When the screening of a trapped charge by a polar medium such as SiO₂ is quantitatively accounted for in this local field picture, a new physically sound RTN emission mechanism emerges. Similarly, the dynamics of post-stress recovery of Negative-Bias-Instability of p-channel MOSFET can be more rationally explained.

Index Terms

MOSFET; local field; trapped charge; emission; NBTI; RTN

I. Introduction

Charge-capture/emission is a fundamental process in electron devices. Its effect is beneficial at times and detrimental at other times. It is the basic operation principle of some nonvolatile memory [1, 2], including exotic new ones based on novel materials [3]; the likely source of resistance fluctuation in resistive memory [4] and the primary reason for many performance degradation modes of Metal-Oxide-Semiconductor-Field-Effect-Transistors (MOSFETs), such as random telegraph noise (RTN) [5-7]. Thus, to understand a broad range of electron devices, one must understand the dynamics of charge-capture/emission.

For the case of a MOSFET, the most important charge-capture happens at defect sites in the dielectric layer of the device. The defect sites are spatially distributed in three dimensions and so are the trapped charges. The standard approach to treat trapped charges is to think of them as a thin sheet (two dimensions) of uniform charges at the centroid location along the

thickness direction of the dielectric layer [8]. The thin sheet of uniform charges produces a uniform modification of the electrostatic field across the entire device and a uniform band-bending in the semiconductor near the interface, as illustrated in fig. 1. Thus, the flat-band voltage shift, V_{FB} , due to trapped charges can be expressed for this simple case as [8]:

$$\Delta V_{FB} = -\frac{1}{C_{ox}} \int_0^{x_{ox}} \frac{x}{x_{ox}} \rho(x) dx = -\frac{Qx_c}{\epsilon\epsilon_0 A} \quad (1)$$

where C_{ox} is the oxide capacitance, x_{ox} is the oxide thickness, $\rho(x)$ is the location dependent charge density, Q is the total trapped charges, x_c is the charge centroid location within the dielectric, ϵ is the dielectric permittivity, ϵ_0 is the permittivity of free space, and A is the area of the MOS structure. Adoption of this approach greatly simplified the treatment of the charge-capture/emission dynamics and therefore the operation of many electron devices. This paradigm has consequently dominated all discussions of MOSFET charge-capture/emission in the literature for decades.

The practice of approximating a collection of immobile charges as a uniformly-charged body with no local character has been a common practice in the modeling of transistors. Until recently, this practice proved a quite capable construct to understand ionized channel dopants in larger area devices. As device dimensions scaled down such that individual ionized dopants have deterministic influence on device characteristics, it became necessary to treat the highly-localized potential associated with each individual dopant ion. This resulted in the, now commonly accepted, fluctuating surface potential in the channel concept [9-12]. Since trapped charges in the oxide are electrostatically quite similar to ionized dopants in the substrate, they too should cause localized surface potential distortions in the channel.

Fig. 2 is an illustration of a highly-localized surface potential in the n-channel MOSFET channel. This is a modification of the dopant only picture presented by Asenov [13] – each localized peak corresponds to an ionized dopant. In the middle of the channel, a peak corresponding to a dielectric trapped electron near the interface has been added. This peaked surface potential from the trapped electron is the basic idea behind the “hole-in-the-inversion-layer” model for the explanation of the amplitude of RTN in MOSFETs [5, 14-19]. In this model, the trapped charge creates a localized region of reduced (a “hole”) or zero (a cored-out “hole”) inversion charge density in the inversion layer, leading to a reduction in drain current. This model presents a more physical and quantitative understanding of RTN amplitude in highly scaled devices. However, the effect of the local surface potential distortion on the emission kinetics of the trapped charge was, surprisingly, not considered. In other words, charges are assumed to emit to a non-distorted surface potential even though the model assumes a highly-distorted surface potential locally. While the transistor is biased in inversion during the RTN measurement, these “holes” produce *local* band bending more akin to depletion or, more likely, accumulation.

In this paper, which is an expansion of a recent presentation [20], we explore what this means to the charge emission process and reconcile experimental observations, using RTN

as one of the examples. In general, we note that there is strong reason to believe that charge emission proceeds in a fashion quite different than conventional wisdom. We also make a strong case for large multi-phonon relaxation consistent with the highly flexible amorphous SiO₂ network. This combination forces the consideration that the charge emission (de-trapping) process, typically used to explain much device physics, does not proceed as expected. In many cases, it is quite conceivable that while the origin of the charge capture process may be the conduction band, the emission final state is likely located in the valance band, or vice versa. These observations are consistent with a wide body of literature and have implications not only to RTN, but also to a variety of other device degradation/wear-out mechanisms including the negative bias temperature instability.

II. Results and Discussion

A. The Local-Charge Model

Fig. 3 schematically illustrates the electrostatics associated with a trapped electron at the oxide/channel interface. The schematic shows several band diagrams “snapshots” transitioning from outside the “hole” (inversion), to the center of the “hole” (strong accumulation). If the trapped electron were to emit to the inversion layer, as commonly assumed, it must tunnel laterally through a potential barrier of the oxide by a distance given by the radius of the “hole”. As tunneling is a strong function of barrier thickness, this is possible only when the “hole radius is well below the trap depth (distance from interface). This means the “hole” radius must be well below 1 nm, a condition very difficult to achieve even in very strong inversion [14-19]. As most RTN experiments are done at or near threshold, this is an extremely low probability process. We see immediately that the common treatment of trapped electron emission kinetics is substantially more complicated. Quantum mechanically, emission occurs through the shortest distance or thinnest barrier. However, the distance/barrier defined in this trapped electron situation corresponds to a region where the surface potential is strongly accumulated. Thus, the conventional understanding of electron emission from a dielectric trap to the conduction band must be re-envisaged.

Fig. 4 further illustrates the difference between the conventional sheet charge model and the localized charge model near the trapped charge. Trap sites inside the dielectric capture charge from the inversion layer through tunneling (right-hand side). The resulting (trapped charge) band diagrams are on the left-hand side. The top diagram is the conventional sheet-charge approximation model. The sheet charge uniformly distorts the oxide field, leading to weaker inversion everywhere in the device. The conventional emission models treat the problem of electrons emitting under this band-bending condition. The bottom diagram shows the local electrostatics associated with one of the trapped electrons. The strong local field leads to accumulation in the semiconductor region adjacent to the trapped electron. Electron emission in this band-bending condition is clearly different than what is typically assumed.

B. Trap Relaxation Energy

This somewhat simple explanation of the emission conditions has large implications to the current understanding of charge capture and emission in modern devices. Using the local-

charge band diagram in fig. 4 to explain the charge-capture/emission process leading to RTN in small transistors reveals an immediate problem. The electron emission process corresponding to the local-charge model (lower left-hand side) is too efficient – tunneling is to an unbounded energy region in the semiconductor with a high density (continuum) of empty states that are ready to accept charge. This makes the emission time orders of magnitude shorter than the capture time so that conventional measurement windows of RTN will “miss” the transitions. Note that both capture and emission of charge involve tunneling processes. The distance of the defect site from the interface affects both capture and emission time constants and the range of time constants from different defects can be very large. For each defect, measures of the capture time constant to emission time constant ratio are not too large. For example, to cover the 10s and 1ms time constants simultaneously requires, at the least, 100s of measurement at 0.1ms sampling rate, or 1 million data point pairs per data set. A more severe limitation is that when inspecting a 100s data set, one cannot see the 10ms break in the data. Thus, the meaning of measurement will “miss” the transition is that most of the RTN measurements in the literature would fail to capture these transitions. Throughout this work, phenomena are deemed “unobservable” to indicate that typical experimental measurements are insufficient to resolve this behavior.

At first glance, the sheet-charge model (upper left-hand side) appears to present a better explanation of RTN. However, this elastic tunneling picture cannot explain many RTN observations. This forces the consideration of an inelastic picture involving relaxation of the energy level of the trapped charge [21, 22]. Emission must be considered from the energy level of the trapped electrons that have relaxed through a multi-phonon process [23], not as depicted on the left-hand-side of figure 4.

Upon charge localization (capture) inside a polar medium (dielectric), the dipoles surrounding the charge will rotate to screen the electric field of the charge and lower the overall energy. This is basic physics and the details of this multi-phonon relaxation process were discussed by Henry and Lang [23] as is depicted in fig. 5. An electron tunnels into a defect state. It can tunnel back out, or it can be captured by overcoming an energy barrier E_B representing the configuration difference between the defect with unrotated surrounding dipoles and the defect with rotated surrounding dipoles. After crossing over to the defect complex with the proper configuration, the defect complex is in a highly excited vibrational state. Through the emission of many phonons, the complex loses energy, $E_R + E_b$, and relaxes to the lowest vibrational energy state. E_R is the relaxation energy reflecting the energy difference between the initial and final state. (Note that relaxation energy is sometimes defined differently [24].)

After multi-phonon relaxation, the trapped charge is at an energy level below the original energy level of the defect when it was empty. With the energy of the trapped electron below the Fermi level, emission to empty states becomes a phonon-assisted tunneling process which is less probable. The larger the energy difference from the empty states, the slower the emission time constant. To reconcile with experimental emission data and using the sheet-charge model, both Kirton and Uren *et al.* [21] and Palma [22] are forced to conclude that the relaxation energy E_R must be in the 20 to 150 meV range. Whether such a relaxation energy range makes physical sense was never given proper consideration.

The relaxation energy can be bounded by considering the solvation energy of an electron in water. Water molecules are polar and they are free to rotate (hydrogen bonding neglected). Thus, the solvation energy of an electron in water represents the ultimate relaxation energy. A consensus [25, 26] of 3.3 eV has recently been reached for the solvation energy of an electron in water. Furthermore, the solvation energy of an electron at the surface of water is 1.6 eV, in agreement with the physical picture of the electron being only half screened. SiO₂ molecules in the dielectric layer are, obviously, not free to rotate. However, the estimated 20 to 150 meV of screening is strikingly small compared to the 3.3 eV in the freely rotatable water case. This comparison indicates that the RTN derived, 20 to 150 meV, relaxation energy should correspond to a highly rigid dielectric environment (very limited dipolar rotations). This notion is in stark contrast to the well-known high degree of flexibility present in the SiO₂ network [27-29].

Experimentally, there are two independent reports that directly measure the relaxation energy in SiO₂. The first report details the self-trap energy of a hole measured by Lu and Sah [30]. The second report details relaxation energy of an inelastic-tunneling electron measured by Takagi *et al.* [31]. Both experiments found the relaxation energy of a charge in the SiO₂ dielectric to be ~1.5 eV – a value much more consistent with a flexible network. In other words, nearly half of the electric field by a trapped charge in the dielectric is screened by the rotation of the dipoles in SiO₂. The unscreened portion is still very large and will produce a strong local field effect.

Recall that the 20 to 150 meV relaxation energy E_R was deduced by the need to reconcile with measured emission time constants of RTN in the sheet-charge model [21, 22]. With E_R at 1.5 eV, the emission kinetics are very difficult to explain using the conventional sheet charge model. Fig. 6A illustrates the energy landscape involved. The shaded arrows indicate the electron-emission route. To access the empty states, the trapped charge must absorb enough phonons to overcome a large energy barrier of ~ 1.5 eV. The probability is simply too low to reconcile with experiment [21, 22]. Note that calculation of emission probability often invokes the detailed balance principle, which requires the electron to follow the capture route in reverse. Such a route will have even lower probability. Emission could occur to the gate electrode as indicated by the dotted arrow, but such scenario is incompatible with the many early studies showing RTN in thicker gate oxide devices (> 3 nm) as this tunneling pathway is effectively closed off due to the large barrier [32].

Armed with a better grasp of the relaxation energy, it is possible to get a quasi-quantitative estimate of the strength of the accumulation in region 1 of fig. 3. The electric field at the oxide-semiconductor interface created by a trapped-charge in the dielectric can be estimated using a modified Coulomb equation:

$$E = \frac{ss'q_0}{4\pi\epsilon\epsilon_0d^2} \quad (2)$$

where d is the distance of the charge from the interface, s is the bulk dielectric screening factor (dipole rotation), q_0 is the unit charge, ϵ is the relative permittivity of the oxide (3.9), and s' is the gate screening factor, which is given by:

$$s' = \frac{x_{ox} - d}{x_{ox}} \quad (3)$$

.In this simple formalism, there is an assumption that at a finite distance from the charge ($d > 0$), the equation is applicable. Fig. 7 plots the electric field at the interface as a function of distance to the trapped-charge for 40 nm (Fig. 7a) and 2 nm (Fig. 7b) oxide thicknesses. Since the bulk dielectric screening factor (s) is somewhat variable we calculate the induced electric field at the interface for the two extreme screening cases. In the case that the dipoles are far away from the interface, s is taken as one minus the ratio of the SiO₂ electron relaxation energy to the free electron relaxation energy ($1 - (1.5 \text{ eV}/3.3 \text{ eV}) \cong 0.55$). In the case that the dipoles are located closer to an interface and are less free to move, s is taken as one minus the ratio of $\frac{1}{2}$ of the SiO₂ electron relaxation energy to the free electron relaxation energy ($1 - (0.75 \text{ eV}/3.3 \text{ eV}) \cong 0.77$). This follows from the preceding bulk/interface solvation energy arguments.

As seen in fig. 7, the electric field created by a trapped-charge in the dielectric is a strong function of the location of the charge. At very short distances, the uncertainty in s is bounded by the bulk and surface values. At longer distances ($>0.6 \text{ nm}$) the difference becomes unimportant. Overall, this simple exercise gives a crude estimate as to the field induced by the electric field at the interface and defines an applied field range where the one needs to account for differences in bulk and interface screening. Although we note that the bulk and interface screening differences introduce only small modifications to the field.

Depending on the device design, the operation field varies from 3 MV/cm to 6 MV/cm. For most RTN experiments, the applied field is lower, ranging from 0.5 MV/cm to 2 MV/cm. In this lower field measurement range, the near interface trapped-charge completely overwhelms the applied field, turning inversion into extremely strong accumulation. As the location of the trapped-charge extends deeper into the oxide, the effect weakens rapidly. It should be noted that depth more than 1 nm is not often encountered because reduced tunneling probability make these locations too slow to respond in most experiments. Note that since RTN is often measured sub-threshold, even the charge at 1 nm distance can turn inversion into accumulation. Overall, this simple exercise gives a crude estimate as to the field induced by the electric field at the interface and defines an applied field range where the one can safely assume that the trapped charge dominates the electrostatics.

C. RTN Time Constants

Returning to the emission problem, with the local charge model, the available empty states for electron emission are in the valence band as indicated by the shaded arrows in fig. 6B. In this case, emission becomes very efficient if the band diagram remains conventional (unconfined in the lateral directions). If electrons are captured from the first sub-band of the quantum-confined inversion layer ($\sim 0.2 \text{ eV}$ above the bottom of the silicon conduction

band), the energy of the post-relaxation trapped-electron would be ~ 0.2 eV below the top of the silicon valence band. Locally this state is in strong accumulation. Capturing a hole from the accumulation layer is energetically extremely favorable. This would lead to very short emission times which would again render RTN unobservable (capture time to emission time ratio is too large) in most experiments. Thus, it appears that the inclusion of strong multi-phonon relaxation (with more physically reasonable relaxation energies) poses difficult problems for both models.

The above discussed difficulty is the result of the assumed conventional band diagram that, in strong accumulation, has a high density of available holes (empty states) - a consequence of the sheet-charge model in which the band-bending is uniform across the transistor. Recognizing that with the local-charge model there exists a strong quantum confinement associated with the highly-localized field, this difficulty can be removed. In addition to the vertical confinement, the local field adds confinement in the other two lateral dimensions. With confinement in all three dimensions, the first sub-band not only shifts away from the top of the valence band, it is also further split into distinct levels.

As an example, if one compares the energy difference between the ground state of a $1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m}$ transistor to the “hole” size (arising from local confinement) with a 2 nm radius. Using the simple quantum-well equation, the ratio of the ground state energies scales as the well size squared $(1000\text{ nm}/4\text{ nm})^2 = 62,500$. This simple calculation reveals that the energy spacing increases substantially such that the first few energy levels are very far from the band edge. Furthermore, one would expect the density of states to decrease by the same factor. With the first few energy levels much further away from the band edge. Access to these levels requires the absorption of more phonons. Fig. 6b shows the quantized energy levels in the accumulation layer. Since the tunneling rate depends on both the barrier (height and thickness) and the density of available tunneling states, the 3-D confinement very effectively suppresses the electron emission probability and makes RTN observable in an experimentally relevant measurement window. Furthermore, even though the electron is ~ 0.2 eV below the valence band edge, the emission still requires phonon assisted tunneling to reach the empty states as illustrated in fig. 6b. This is consistent with the observed temperature dependence on capture and emission time constants [33].

Now that the local-charge model has resolved the electron emission difficulty encountered by the sheet-charge model, the objective of the rest of the paper is to support the proposed model by applying the local-charge model to a few well-studied RTN observations.

Experimentally, most of the reported RTN time constants have the following behavior: as gate bias decreases, the capture time constant increases and the emission time constant decreases [33]. The charge-capture model shown in the right-hand side of fig. 4 explains the gate bias dependent *capture* time constant. For the local-field model to work, it needs to somehow also account for the gate-bias dependent *emission* time constant. The “hole” radius r from the “hole-in-the-inversion-layer” model [19], already includes gate and inversion layer screening. Thus, dielectric screening s , can be included with only a slight modification:

$$r = \sqrt{\frac{sq}{Q} \cdot \frac{x_{ox} - d}{x_{ox} + X_Q}} \quad (4)$$

where q is electron charge, Q is inversion charge density, d is the distance of the trapped electron from the interface, and X_Q is the distance (in equivalent oxide thickness) of the inversion layer from the interface (see inset of fig. 8). For gate overdrives slightly above threshold, the inversion layer distance is crudely approximated as 0.4 nm [34]. Eqn. (4) then allows one to calculate the hole radius for case of screening by the gate, the inversion layer, and the dielectric. Fig. 8 shows the hole radius as a function of inversion charge density for the trapped charge distance $d = 0$ (red triangles) and $d = 0.3$ nm (black diamonds) in a 40 nm thick oxide. For $d = 0$, the rigid substrate limits dipole rotation which is approximated as $s = 0.77$. For $d = 0.3$ nm, dipole rotation is easier which is approximated as $s = 0.55$.

In general, the factor $(x_{ox} - d)/(x_{ox} + X_Q)$ has very small impact except when oxide is very thin where the impact can be easily calculated using d and X_Q . The main impact of distance from the interface is the decrease in the induced electric field.

As the gate bias decreases, the “hole” radius r increases. That is, the lateral quantum confinement of the accumulation layer decreases, leading to a rapid increase in the density of empty states near the top of the valence band as shown in fig. 8. The result is a higher tunneling probability and emission rate of the trapped electrons. Furthermore, the lower gate bias also decreases the vertical quantum confinement at the charge-capture step (right-hand side of fig. 4), which lowers the first sub-band toward the conduction band edge. In other words, charge capture through tunneling is to a defect site at lower energy. After the 1.5 eV relaxation, the energy level of the trapped electron is now further below the valence band edge and therefore able to access the higher density of empty states. Combining these two factors results in a drastic increase in the density of empty states available for tunneling with decreasing gate bias. This is equivalent to a rapid increase in emission rate – consistent with the experimental results [33].

The general trend of lower gate bias leading to larger “hole” size has been observed by many groups [5, 14-19]. Also note that (4) clearly illustrates the link between the distance of the trapped-charge from the oxide/semiconductor interface and the “hole” radius. Coupled with the location’s effect on the local field at the interface and the effect on tunneling probability, one should expect the observation of a wide range of time constants. Obviously, one cannot discuss all these cases in detail here. The important conclusion is that the deep relaxation and the associated modifications to the local field are unavoidable for SiO₂ and possibly many oxides.

It should be mentioned that the idea of a local field effect was briefly considered by Uren *et al.* [35] to explain the observed anomalous RTN - some rapidly switching RTN signal in the drain current of a MOSFET disappears for some period and then reappears. While they did not elaborate, they argued that when a charge is captured at a defect site, the nearby defect sites will experience a Fermi level shift such that the probability of this near defect site to

capture charge is reduced. If this is valid, the anomalous RTN, at least the simple two-level cases, can be explained. They estimated that at least 75 to 100 meV (a few kT where k is the Boltzmann constant and T is temperature in Kelvin) of Fermi level shift is required for the neighboring defect site to experience suppressed charge-capture. They claimed, without justification, that for such a large shift, the neighboring defect sites must be within 2 nm of the trapped electron. They estimated that the probability of finding a neighboring defect site at such short distance is about 40 times too small to explain the observed occurrence of anomalous RTN, they therefore discarded the explanation.

Within the confines of the local-field model, “shifting the Fermi level of a neighboring defect site” can be replaced by a change in band-bending. Referring to fig. 3, the band-bending within the “hole” is quite large. The boundary of the “hole” is where the inversion charge density changes rapidly. Intuitively, more than a few kT shift in band edge is required for that to happen. Thus, the “distance” Uren *et al.* needed to explain their anomalous RTN is simply the radius of the “hole” which can be found using equation (4), or fig. 8. Since Uren *et al.*’s RTN measurements were taken at drain currents of about 10 nA (gate oxide = 40 nm), it is safe to assume that their transistor was at or below threshold. The inversion charge density for their experiment should therefore be somewhere in the shaded region of fig. 8. To account for Uren *et al.*’s 40 $\times$ probability miss match, the “hole radius needs to be 6.3 $\times$ larger than the 2 nm they assumed. As can be seen, the 12.6 nm radius is within the shaded area and therefore the local-field model can consistently explain the anomalous RTN without introduction of further assumptions.

Using 3-D atomistic simulations, Asenov *et al.* [17] found that the “hole” radius is larger for thicker gate oxides. For transistors with a 7nm gate oxide, they found that 2 nA/ μ m of drain current (below threshold) corresponded to a “hole” radius of approximately 11 nm – in agreement with fig. 8.

Given the size of the “hole” at these low inversion charge density situations, it is quite possible that more than two defects can be entangled by the local-field and more complex anomalous RTN can be explained. One should remember that the defects are not all at the same distance from the interface. The 3-D consideration produces a rich possibility of complex RTN phenomena.

D. NBTI

So far, we have used RTN as an example to demonstrate how the local-field model of trapped charge works and showed that RTN explained with the local-field model is more consistent with experimental facts. Since charge-capture/emission is central to so many phenomena associated with the electron device, this local charge model likely plays a role in other well-known phenomenon such as the negative-bias-temperature-instability (NBTI). The application of a negative bias and elevated temperature stress to modern p-channel MOSFETs evokes a wear-out mechanism which manifests as a threshold voltage shift [36]. This threshold voltage shift recovers partly upon the cessation of the stress [37-39]. This threshold voltage shift and recovery is often linked in part to a hole capture and emission behavior respectively [36-39]. It is widely agreed that hole capture occurs below the Si valence band. Emission of the trapped hole is often assumed to proceed back to the Si

valence band. However, like in the RTN picture, the local-charge model leads to different conclusions.

Fig. 9A shows the band diagram of hole-capture during NBTI stress of a pMOSFET. Fig. 9B shows the recovery in the sheet-charge model and fig. 9C shows the recovery in the local-charge model. It should be note that we have assumed cold hole injection during stress, which is the real NBTI mechanism [40] instead of the high-field stress effects that confuse the issue in many modern studies [41].

The most notable characteristic of NBTI is the post-stress recovery, which is the emission of the trapped-hole when the gate bias is lowered to threshold voltage. It is also known that the application of a higher recovery bias, V_R , significantly slows the recovery rate [42]. From Fig. 9B, the sheet-charge model band diagram indicates that recovery requires phonon assisted tunneling into the valence band. The energetics for a different V_R are the same and cannot explain the bias dependent recovery rate. From fig. 9C, the local-charge model band diagram indicates that the post relaxation trapped hole requires phonon assisted tunneling into the conduction band. The effect of V_R is in its impact on the inversion charge density and therefore the “hole” radius. Lowering V_R (less negative) rapidly increases the density of states available for tunneling as shown in fig. 8. Thus, the local-charge model naturally explains the observed V_R dependent recovery.

III. Conclusions

In summary, the sheet-charge approximation that has dominated charge-capture/emission in electron devices is re-examined in this work. The effects of the local-field from individual trapped charges are considered. The magnitude of the energy relaxation associated with a charged defect complex due to dipole screening is pinned down to be $\sim 1.5\text{eV}$, based directly from two independent experiments and is consistent with the flexible network nature of SiO_2 . It is shown in this work that once the full magnitude of energy relaxation of the trapped charge is accounted for, the local-field model offers a much more physically sound mechanism to charge-emission kinetics, using RTN and NBTI in MOSFETs as examples. The new mechanism in each case differs from the traditional one drastically. The implication is that many well-known phenomena in electron devices involving charge capture/emission need to be reconsidered.

References

1. Wegener HAR, Lincoln AJ, Pao HC, O'Connel MR, Oleksiak RE. The variable threshold transistor, a new electrically-alterable, non-destructive read-only storage device. IEEE Int Electron Dev Meeting (IEDM). 1967:70.
2. Goel N, Gilmer D, Park H, Diaz V, Sun Y, Price J, Park C, Pianetta P, Kirsch P, Jammy R. Erase and retention improvements in charge trap flash through engineered charge storage layer. IEEE Electron Dev Lett. 2009; 30(3):216–218.
3. Choi MS, Lee G-H, Yu Y-J, Lee D-Y, Lee SH, Kim P, Yoo J, Hone WJ. Controlled charge trapping by molybdenum disulphide and graphene in ultrathin heterostructured memory devices. Nature Communications. 2013; 4:1624.
4. Ielminia D, Nardi F, Cagli C. Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories. Appl Phys Lett. 2010; 96 053503-053503-3.

5. Yau LD, Sah C-T. Theory and experiments of low-frequency generation-recombination noise in MOS transistors. *IEEE Trans Electron Dev.* 1969; 16(2):170–177.
6. Ang DS, Pey KL. Evidence for two distinct positive trapped charge components in NBTI stressed p-MOSFETs employing ultrathin CVD silicon nitride gate dielectric. *IEEE Electron Dev Let.* 2004; 25(9):637–639.
7. Campbell JP, Cheung KP, Suehle JS, Oates A. Negative-bias temperature instability induced electron trapping. *Appl Phys Lett.* 2008; 93(3):033512–1.
8. Wolters DR, van der Schoot JJ. Kinetics of charge trapping in dielectrics. *J Appl Phys.* 1985; 58(2): 831–837. Taur, Y., Ning, TH. *Fundamentals of Modern VLSI Devices.* Cambridge University Press; 1998. p. 87 See also
9. Nishinohara K, Shigyo N, Wada T. Effects of Microscopic Fluctuations in Dopant Distributions on MOSFET Threshold Voltage. *IEEE Trans Electron Dev.* 1992; 39(3):634–639.
10. Hagiwara, T., Yamaguchi, K., Asai, S. *Tech Digest, Symp VLSI Technology.* Honolulu, HI, USA: 1982. Threshold Voltage Deviation in Very Small MOS Transistors Due to Local Impurity Fluctuations; p. 46
11. Tang X, De VK, Meindl JD. Intrinsic MOSFET parameter fluctuations due to random dopant placement. *IEEE Trans Very Large Scale Integration (VLSI) Systems.* 1997; 5(4):369–376.
12. Asenov A. Random dopant induced threshold voltage lowering and fluctuations in sub 50 nm MOSFETs: a statistical 3D ‘atomistic’ simulation study. *Nanotechnology.* 1999; 10(2):153.
13. Brown A, Asenov AR, Roy G, Cheng B, Alexander C, Riddet C, Kovac U, Martinez A, Seoane N, Roy S. Simulation of statistical variability in nano-CMOS transistors using drift-diffusion, Monte Carlo and non-equilibrium Green’s function techniques. *J Comput Electron.* 2009; 54(8):349–373.
14. Reimbold G. Modified 1/f trapping noise theory and experiments in MOS transistors biased from weak to strong inversion - Influence of interface states. *IEEE Trans Electron Dev.* 1984; 31(9): 1190–1198.
15. Ohata A, Toriumi A, Iwase M, Natori K. Observation of random telegraph signals: Anomalous nature of defects at the Si/SiO₂ interface. *J Appl Phys.* 1990; 68(1):200–204.
16. Simoen E, Dierickx B, Claeys CL, Declerck GJ. Explaining the amplitude of RTS noise in submicrometer MOSFETs. *IEEE Trans Elect Dev.* 1992; 39(2):422.
17. Asenov A, Balasubramaniam R, Brown AR, Davies JH. RTS amplitudes in decananometer MOSFETs: 3-D simulation study. *IEEE Trans Elect Dev.* 2003; 50(3):839.
18. Sonoda K, Ishikawa K, Eimori T, Tsuchiya O. Discrete Dopant Effects on Statistical Variation of Random Telegraph Signal Magnitude. *IEEE Trans Elect Dev.* 2007; 54(8):1918.
19. Southwick, RG., Cheung, KP., Campbell, JP., Drozdov, SA., Ryan, JT., Suehle, JS., Oates, AS. Physical model for Random Telegraph Noise amplitudes and implications; *Tech Digest, IEEE Silicon Nanoelectronics Workshop (SNW);* Honolulu, HI, USA. 2012.
20. Cheung, KP., Veksler, JP., Campbell, D. Rethinking Charge trapping and detrapping dynamic without the sheet-charge approximation; 47th IEEE Semiconductor Interface Specialists Conference (SISC); San Diego, CA. Dec. 2016 (Attached as supplementary material)
21. Kirton MJ, Uren MJ. Capture and emission kinetics of individual Si:SiO₂ interface states. *Appl Phys Lett.* 1986; 48(19):1270–1272.
22. Palma A, Godoy A, Jimenez-Tejada JA, Carceller JE, Lopez-Villanueva JA. Quantum two-dimensional calculation of time constants of random telegraph signals in metal-oxide-semiconductor structures. *Phys Rev B.* 1997; 56(15):9565–9574.
23. Henry CH, Lang DV. Nonradiative capture and recombination by multiphonon emission in GaAs and GaP. *Physical Review B.* 1977; 15(2):989LP–1016.
24. Alkauskas A, Yan Q, Van de Walle CG. First-principles theory of nonradiative carrier capture via multiphonon emission. *Phys Rev B.* 2014; 90:075202.
25. Siefertmann KR, Liu Y, Lugovoy E, Link O, Faubel M, Buck U, Winter B, Abel B. Binding energies, lifetimes and implications of bulk and interface solvated electrons in water. *Nature Chemistry.* 2010; 2(4):274–279.
26. Abel B, Buck U, Sobolewski AL, Domcke W. On the nature and signatures of the solvated electron in water. *Phys Chem Chem Phys.* 2012; 14(1):22–34. [PubMed: 22075842]

27. Mozzi RL, L R, Warren BE. The structure of vitreous silica. *J Appl Crystallography*. 1969; 2(4): 164–172.
28. Grimley DI, Wright AC, Sinclair RN. Neutron scattering from vitreous silica IV. Time-of-flight diffraction. *J Non-Crystalline Solids*. 1990; 119(1):49–64.
29. Tucker MG, Keen DA, Dove MT, Trachenko K. Refinement of the Si–O–Si bond angle distribution in vitreous silica. *J Phys: Condensed Matter*. 2005; 17(5):S67.
30. Lu Y, Sah CT. Thermal emission of trapped holes in thin SiO₂ films. *J Appl Phys*. 1995; 78(5): 3156–3159.
31. Takagi, S., Yasuda, N., Toriumi, A. Tech Digest, Int Electron Dev Meeting. DC, USA: 1996. Experimental evidence of inelastic tunneling and new I-V model for stress-induced leakage current; p. 323-326.
32. Heiman FP, Warfield G. The effects of oxide traps on the MOS capacitance. *IEEE Trans Electron Dev*. 1965; 12(4):167–178.
33. Ralls KS, Skocpol WJ, Jackel LD, Howard RE, Fetter LA, Epworth RW, Tennant DM. Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/f) Noise. *Phys Rev Lett*. 1984; 52(3):228LP–231.
34. King Y-C, Fujioka H, Kamohara S, Chen K, Hu C. Dc electrical oxide thickness model for quantization of the inversion layer in MOSFETs. *Semiconductor Science and Technology*. 1998; 13(8):963–966.
35. Uren MJ, Kirton MJ, Collins S. Anomalous telegraph noise in small-area silicon metal-oxide-semiconductor field-effect transistors. *Physical Review B*. 1988; 37(14):8346LP–8350.
36. Schroder DK. Negative Bias Temperature Instability: What do we understand? *Microelectron Rel*. 2007; 47:841–852.
37. Lin, HC., Lee, DY., Ou, SC., Chien, CH., Huang, TY. Impacts of hole trapping on the NBTI degradation and recovery in PMOS devices; Tech Digest, Int Workshop Gate Insulator (IWGI); Tokyo, Japan. 2003. p. 76-79.
38. Denais M, Huard V, Parthasarathy C, Ribes G, Perrier F, Revil N, Bravaix A. Interface trap generation and hole trapping under NBTI and PBTI in advanced CMOS technology with a 2-nm gate oxide. *IEEE Trans Dev and Mat Reliab*. 2004; 4(4):715–722.
39. Ang DS, Wang S, Du GA, Hu YZ. A Consistent Deep-Level Hole Trapping Model for Negative Bias Temperature Instability. *IEEE Trans Dev and Mat Reliab*. 2008; 8(1):22–34.
40. Reisinger H, Vollertsen RP, Wagner PJ, Huttner T, Martin A, Aresu S, Gustin W, Grasser T, Schlunder C. The effect of recovery on NBTI characterization of thick non-nitrided oxides. *IEEE Int Integrated Reliability Workshop Final Report*. 2008
41. Campbell JP, Cheung KP, Suehle JS, Oates AS. The Role of High-Field Stress in the Negative Bias Temperature Instability. *IEEE Trans Dev And Mat Reliab*. 2010; 10(4):482–491.
42. Parthasarathy CR, Denais M, Huard V, Ribes G, Vincent E, Bravaix A. New Insights into Recovery Characteristics Post NBTI Stress. *IEEE Int Reliability Phys Symp*. 2006

Fig. 1.

Trapped charges in the oxide layer is approximated as a sheet of uniform charges at the centroid location, x_C , producing a uniform electrostatic field and a uniform band-bending in the semiconductor.

Fig. 2.
Peaky surface potential in the nMOSFET channel due to random dopant and electron trapped in oxide.

Fig. 3.

A cored-out hole in the inversion layer surrounds the trapped electron. Three regions are identified to show the corresponding band bending.

FIG. 4.

On the right-hand side defect sites capture electron from the inversion layer. The post-capture band diagrams are on the left-hand side. The top diagram is the conventional sheet-charge model. The trapped sheet charge leads to a uniformly weaker inversion across the entire device. The bottom is the local charge model. Each trapped electron causes the local semiconductor strongly accumulated.

Fig. 5.

The electron tunneled into a defect state has a finite probability of overcoming a configuration barrier E_B to be captured into the excited vibrational state of the defect complex with the proper dipole orientations. It then emits multiple phonons to reach the lower vibrational state of this complex, losing energy E_R .

Fig. 6.

A: Band diagram after the capture of an electron by a defect in the sheet-charge model. Shaded arrows indicate thermally excited emission to the semiconductor side. Dotted arrow indicates thermally excited emission to the gate side. B: Band diagram in the local-field model. Thermally excited emission involves a much smaller barrier.

Fig. 7.

The electric field at the oxide/semiconductor interface due to a trapped-charge in the oxide as a function of the trapped-charge location from the interface. a: 40 nm oxide; solid circle is the surface trapped-charge case; open circle is bulk trapped-charge case. b: 2 nm oxide; open circle is surface trapped-charge case; open circle is bulk trapped-charge case. Gate screening is included in all cases. Solid lines: surface to bulk linear transition over 0.6 nm distance is assume.

Fig. 8.

Solid lines are calculated “hole” radius as a function of inversion charge density using equation (4) for the trapped charge at the interface (triangle) and at ~0.3nm from the interface. Dotted line: density of state ratio to 2nm “hole” radius for the charge at the interface case.

Fig. 9.

(a) Band diagram depicting the capture of cold holes in the gate dielectric of a pMOSFET under NBTI stress, (b) sheet-charge model band diagram showing that de-trapping requires phonon assisted tunneling, and (c) the local-charge model band diagram showing de-trapping requires phonon assisted tunneling.