Anomalous accelerated negative-bias- instability (NBI) at low drain bias

K. P. Cheung
National Institute of Standards & Technology, Gaithersburg, MD USA

We observed at very low drain bias an anomalous acceleration of Negative-bias-instability at room temperature, as if the channel temperature has been raised significantly. The channel width and channel length dependent of this acceleration suggest that in addition to the conventional self-heating effect that raises the lattice temperature, there is indication that hot holes thermalized at a temperature (T_e) higher than the lattice temperature (T_L) is the cause of this anomaly. Analysis of the frequency dependent of drain bias modulation, as well as the wave shape dependent of the modulation support this explanation.

Self-heating accelerated channel-hot-carrier-degradation (CHD) and bias-temperature-instability (BTI) is a current concern for advanced technology nodes [1-3]. While there are disagreements on the severity of the problem, consensus is that there is an elevated lattice temperature (T_L) in the channel resulting from inhibition of thermal energy removal in advanced geometries such as FinFet or Gate-all-around (GAA) Fet, leading to accelerated CHD and BTI. The observed acceleration of degradations so far seems to entirely attributable to the increase in T_L, even though theoretical works [4, 5] suggest that non-equilibrium effect such as a carrier temperature T_e higher than T_L exists at the drain end of the channel and can lead to an even more severe degradation. This may be deal to most studies are done at stress conditions that produced strong "conventional" degradation that make it difficult to identify the non-equilibrium effect.

In this work, we choose a stress condition that should have weak expected conventional degradations to search for the non-equilibrium effects. From the thermal argument, a planar MOSFET (power devices excluded), particularly a standalone one in nanoscale, should not have notable self-heating accelerated degradation [1]. We choose normal NBTI gate bias but no elevated temperature to suppress normal NBTI. We choose low to very low drain bias to suppress CHD. Under these conditions, we observed a clearly anomalous acceleration of NBI. We argue that elevated electron temperature (T_e) may be the culprit. If true, this is perhaps the first example of reliability impact of high T_e. Naturally, we expect what we observed applies to advanced geometry as well.

The p-MOSFET used in our experiment are standalone 2μ x 50nm devices with 1.6nm (EOT) nitride gate oxide and polysilicon gate. All stresses are done using -2V gate bias. Stress-measure-stress sequence was done at fixed cycle: 600s stress, 4s wait, 5s measure, 4s wait, 5s measure, 4s wait, 5s measure. So, the total interval is 627 seconds. Checking the three measurements (V_t extraction: V_G sweep from -0.8V to 0V, V_D = -0.6V) found that they agree to within measurement noise (indicated by the error bars in fig. 1a). The average of the three is used as the measured result. All experiments were carried out at room temperature, apart from the standard NBTI reference run.

Fig. 1a shows the threshold voltage (V_t) shift as a function of stress time for various drain bias. The fitted lines have been forced to have zero intercept, meaning that V_t shift after 1 second stress is within the measurement noise. In other words, there is no sudden rise in V_t shift due to filling of existing traps as found in many reports in the literature. Fig. 1b shows the V_t shift as a function of drain bias after 7200 seconds of stress. A monotonic rise starting at zero bias is evident.

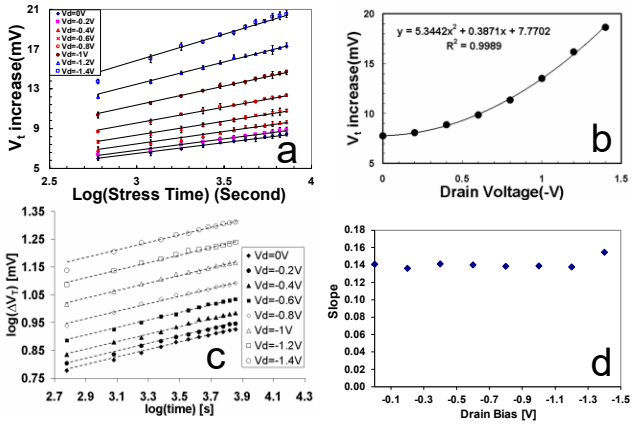

Fig.1 **a**. V_t shift as a function of stress time for various drain bias; **b**, V_t shift after 7200s stress as a function of drain bias; **c**, data in a replotted in log-log scale, **d**, slopes of the plots in c.

NBTI and CHD are intertwined and studies [2, 6, 7] showed that normally the overall degradation decreases with increasing drain bias until a turn-around point at which the increase in CHD overtakes the decrease in NBTI and degradation will increase with drain bias. The reason is at drain bias below -1V the population of energetic hot holes is so low that CHD is negligible. Since drain bias also reduces the oxide field in the channel, a reduction in NBTI as well as the overall degradation results. Contrary to these observations, our data showed only accelerated degradation even at very low drain bias. Note that our devices are from a production technology of a major foundry and very well behave. It is very unlikely that unusually bad devices are the culprit. Indeed, unusually bad devices that degrade easily even at very low drain bias means high density of preexisting defects. The fact that all lines cross zero (the fast, early degradation commonly reported is absent) in figure 1a indicates the quality of the devices studied here were superior to most of the devises reported in the literature. Thus, we call our result anomalous.

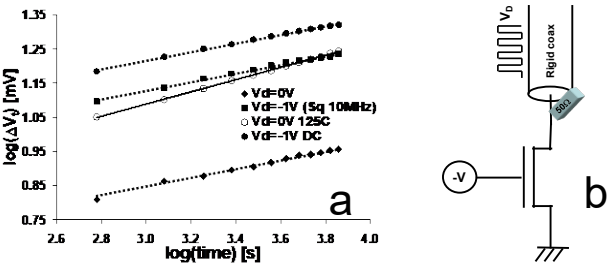

Fig. 2 **a**, V_t shift vs stress time for various conditions; **b**, measurement set up for high-speed drain modulation.

Adding to the anomaly is fig. 1c where the degradations are plotted in the more common log-log fashion and the slope of the fitted lines for all drain bias are 0.14 as shown in fig. 1d, lower than expected for NBTI or CHD [1-3]. To show that this is not an artefact of our experiment, a reference NBTI measurement (V_G = -2V, source/drain grounded, wafer heated to 125°C using hot chuck) was performed and the slope was found to be 0.18 as

shown in fig. 2a. The 0.18 slope is consistent with the theoretical value when a finite delay is used in the measurement [1].

Also plotted in fig. 2a is a modulated rain bias case using a square wave at 10 MHz (0.5 ns rise/fall time). Note that this is not an AC stress with modulated gate bias. Instead, this is a "pulsed" drain bias. To ensure that the drain bias waveform is not distorted, a 50Ω terminated probe as shown in fig. 2b was used. This probe was verified to be good to 25 GHz using time-domain reflectometry measurement. As can be seen, the 10 MHz drain modulated case produced similar level of degradation as the 125°C zero bias case, as if the -1V modulated drain bias is heating the transistor to 125°C. However, the expected temperature rise is less than 10°C [1] for the 0.38 mW/μm power dissipation in this case. If CHD is indeed negligible as expected, then this large degradation is anomalous.

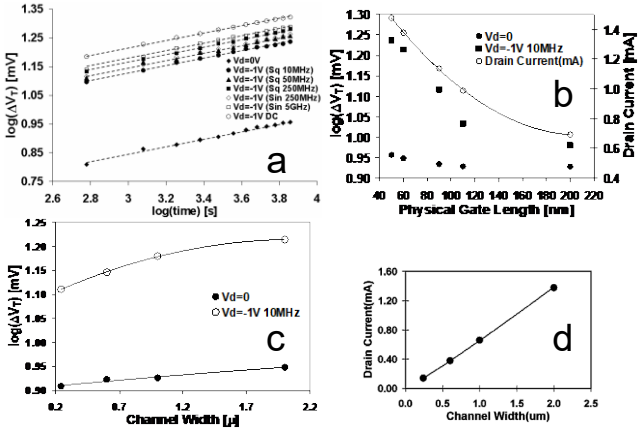

Fig. 3 **a**, V_t shift vs stress time for various drain bias frequency and wave shapes; **b**, V_t shift after 7200s stress and drain current vs channel lengths; **c**, V_t shift vs channel widths; **d**, drain current vs channel widths.

Fig. 3a shows more drain bias modulation cases involving different frequencies and wave shapes. For the square wave modulations, increasing the frequency from 10 MHz to 50 MHz to 250 MHz results in a commensurate increase in degradation, toward the level from DC bias. This again is counter to the expectation from heat removal argument [1] which says lower frequency modulation leads to higher peak temperature and therefore higher degradation. We note that if CHD were responsible for the degradation (instead of being negligible as we argued), the degradation could increase with frequency if gate voltage is modulated. Since gate voltage is steady and at least double the drain voltage at all time, this is not the case.

Sine waves were also used to modulate the drain bias at higher frequencies up to 5GHz (fig. 3a). We note that the degradation due to 250MHz sine wave modulation is very close to the degradation due to 50MHz square wave modulation. From this observation, we reason that the 5 GHz sine wave modulation should be like a 1GHz square wave modulation. The progression from 10MHZ to 1GHz square wave is asymptotically approaching the DC bias level.

Fig. 3b shows that log(degradation) tracks the drain current density (fixed channel width), suggesting that degradation increases exponentially with drain current density. Fig. 3C shows saturation of increases in log(degradation) with channel width. This means degradation increases with power dissipation even though the current density is unchanged (Drain current tracks the channel width as shown in fig. 3d).

Note that even though the degradations measured here are small, ranging from 8mV to 19mV, it is not insignificant given that the stress time is only 7200 seconds and the drain is only at operation voltage. For advanced nodes, 19mV is nearly 10% of V_t.

How to explain this collection of experimental results? We start with the channel width and channel length dependent behavior. The phonon mean-free-path for silicon is ~300nm [8] so the channel widths are all large enough that the thermal diffusion picture is valid (Fourier equation applicable). From narrow width to wide width the heat source changes from short line source to a long line source. For a uniform heat generation rate, the temperature will be higher for longer line. So, the channel width effect may be interpreted as conventional self-heating effect like those reported in the literature [1-3] (Similar to single fin vs multi fins in Finfets). Comparing fig. 3b and c, we see that the degradation change is significantly smaller for the channel width data set than the channel length data set. For the small power dissipation rate of 0.38mW/μm, we may assume that the thermal effect for the narrow width device is negligible and the entire range of degradation in the channel width data set is a measure of the conventional self-heating contribution, which translates to about 6mV out of the 19mV V_t shift.

Since the frequency dependent degradation counters the conventional self-heating picture, it further reinforces the notion that the larger degradation fraction is not due to conventional self-heating (T_L increase). In conventional self-heating heat diffusion is by acoustic phonons and T_e is the same as T_L. In nanoscale hot spot creation, local T_e and "temperature" of optical phonons (T_r) can rise much faster [9] and to a much higher level [4, 5], all riding on top of the slower responding conventional lattice temperature (T_L). How much higher than T_L obviously depends on both the drain current and drain bias. Both T_e and T_r provide a source of hot holes that can accelerate degradation. Also obvious is that at the very low drain bias level, both effects should be very weak. However, even a weak increase in hot holes population would impact the degradation rate.

With higher T_e and T_r the frequency dependent degradation can be explained as follows: As the drain bias is turned on, holes flow from source to drain. As they travel across the channel, they gain energy from the electric field and, at the same time, loss part of what they gain to phonon scattering. Upon arrival at the drain, some of these energetic carries lose their energy by creating optical phonons and the rest will thermalize extremely rapidly [9] to T_e by carrier-carrier scattering. The created optical phonons convert to acoustic phonons rapidly. Nevertheless, a steady state higher optical phonon population ($T_r > T_L$) can be expected.

A well-known argument [1] is that higher phonon scattering will lower the average carrier energy and therefore reduce the phonon generation rate. This argument will be important for the discussion later. For now, we point out that a T_e much higher than T_L will still result at steady state even with this negative feedback mechanism.

Since the optical phonon to acoustic phonon conversion is rapid, T_r follows the drain bias closely, meaning that $T_r \sim T_L$ at the end

of the low drain bias cycle. When drain bias rises in the next cycle, T_r rises from a low value. Thus, every time there is a rising edge, there is a transient increase in T_e due to reduced optical phonon scattering as illustrated in fig. 4. Higher frequency means more rising edges, leading to more degradation. Eventually, frequency is high enough that T_r no longer falls with drain bias, one would approach the DC bias condition in agreement with data in fig. 3a.

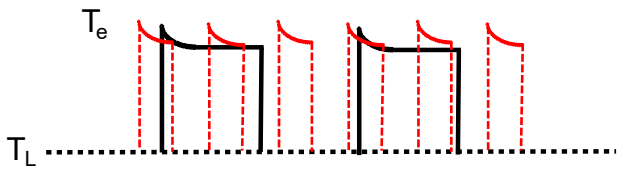

Fig 4. Profile of T_e and T_L in two frequencies (different color) showing the transient higher T_e at the rising edge.

What role does the T_L play in this scenario? T_L is too slow to follow the modulation at 10MHz or higher frequency. It is known that at low frequency, lower frequency leads to higher peak temperature [1]. That argument does not work here because at high frequency T_L does not respond fast enough. Any residual T_L modulation will be too small to show up as frequency dependent effect.

Substrate temperature is known to affect impact ionization rate in silicon and, at low drain bias the effect is positive, meaning that higher temperature leads to higher impact ionization rate [10]. (Note that, at high drain bias the opposite is true.) As impact ionization requires carrier energy higher than the drain voltage at low drain bias, the physics is very similar to what we are dealing with here – high energy carriers are needed to overcome the injection barrier into the SiO_2 layer for degradation to occur. It was suggested [10] that the high energy tail of the carrier distribution (at T_L) entering the channel preserves its shape after acceleration. This means higher T_L leads to higher T_e. Since even after acceleration the mean energy is still not high enough at low drain bias, it is the high energy tail of the T_e distribution that really matters. Consequently, higher T_L leads to higher degradation rate even when the dominant mechanism is T_e.

The reason why sine wave drain bias produces smaller degradation (~5x) than square wave can then be understood by the extremely fast thermalization of hot holes. It literally means that time spent at high voltage is what counts. Sine wave spend much less time at peak voltage.

Finally, what is the explanation for the low exponent of the degradation log-log plots? We do not have one at this point. Perhaps the fact that the channel is not uniformly heated may have something to do with it. However, we do notice that even when drain bias is zero at room temperature, the slope has the same low value.

In summary, we observed at very low drain bias an acceleration of Negative-bias-instability at room temperature, as if the channel temperature has been raised significantly. Many of the observed properties of the degradation is counter to the prevailing understanding of self-heating effects, thus are anomalous. The channel width and channel length dependent of this acceleration suggest that in addition to the conventional self-heating effect that raises the lattice temperature, there is indication that hot holes thermalized at a higher temperature is causing more degradation. Analysis of the dependence of drain bias modulation frequency, as well as the dependence on modulation wave shape further suggest non-equilibrium phonon distribution (optical phonon "temperature" higher than lattice temperature) also plays a role. The use of a regular planar MOSFET at low drain bias facilitated the study of the effect of carrier temperature because conventional self-heating effect is minimized, and CHD is also suppressed. The effect of high carrier temperature, if we are right, should exist in other geometries such as FinFets and GAA Fets.

Acknowledgement: Most of the data were collected by Dr. Yun Wang while he was a student at the Rutgers University.

[1] M. A. Alam et al, TED, 66(11), 4556-4565(2019).
[2] D. Son et al, EDL 40(9), 1354(2019).
[3] P. Paliwoda et al, TDMR 19(2), 249(2019).
[4] A. Abramo et al, JAP 76(10) 5786(1994).
[5] P. A. Childs et al, JAP 79(1), 222(1996).
[6] N. K. Jha et al, IRPS2005, pp524-528.
[7] P. Chaparala et al, Microelectr. Reliab. **45**(1), 13(2005).
[8] E. Pop et al, Proc. IEEE 94(8), 1587(2006).
[9] S. V. J. Narumanchi et al, Heat and Mass Transfer **V42**(6): 478-491(2006).
[10] B. Eitan et al, JAP 53(2), 1244(1982).