Chapter 4

Characteristics and Physical Mechanisms of Positive Bias and Temperature Stress Induced Drain Current Degradation in HfSiON nMOSFETs

4.1 Preface

Unlike SiO_2 CMOS where NBTI in pMOSFETs is considered to be a dominant reliability constraint [4.1], PBTI in high-k nMOSFETs dictates device lifetime [4.2]. PBTI in HfSiON nMOSFETs arises from trap charging and creation in the high-k layer [4.3]-[4.7]. Various PBTI models have been proposed to explain the evolution of the V_t shift although the measured results in literature are not consistent. For example, Zafar et al. found that PBTI induced V_t shift increases with stretched exponential dependence on stressing time and is saturated after prolonged stressing. In their model, they assumed high-k traps have a continuous distribution in cross-sections to derive the observed time dependence. Creation of additional high-k traps during stressing is ignored and the V_t shift is solely due to trapping of electrons at pre-existing high-k traps. On the other side, Shanware et al. showed that the V_t shift has log-time dependence and no saturation. They attributed the V_t shift to electron tunneling into high-k traps continuously distributed in space and could explain the log-time dependence. They ignored trap creation during stressing, too.

In this chapter, we use a fast transient technique to characterize PBTI induced drain current degradation. With this measurement setup, we are

able to monitor the drain current evolution over seven decades of time (from 10^{-3}s to 10^{4}s). As will presented in Section 4.4, the drain current degradation exhibits two stages. The first stage has log-time dependence and the second stage follows power law dependence on stress time. The onset time of the second stage is related to stress voltage, stress temperature and also device process condition. A degradation model is proposed to account for the experimental observations. In addition, high-k trap growth rate is characterized by using the recovery transient technique developed in Chapter 2 and Chapter 3 [4.8][4.9]. Charge pumping measurement is also performed for comparison. The impact of processing on the two-stage degradation behavior is finally discussed.

4.2 Devices

The gate stack in our measured devices consists of a poly-Si gate electrode, HfSiON as the high-k layer with physical thickness of 2.5nm, and an interfacial SiO_2 layer with thickness of 1nm~1.4nm. The transistors have an equivalent oxide thickness (EOT) of 1.4nm~1.8nm, a gate length of 0.08~100μm, and a gate width of 0.16~100μm. Detailed process conditions and device characteristics can be found in [4.2][4.10].

4.3 Two-Phase Measurement

Since post-stress HfSiON CMOS exhibits a large recovery effect in the millisecond range as demonstrated in previous chapters, a switching delay in a conventional method will lead to significant underestimation of degradation and thus an erroneous interpretation [4.6][4.8][4.11]. Therefore, throughout

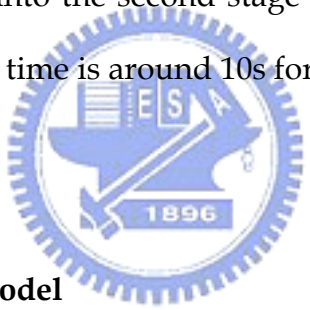
this chapter, again, a transient measurement technique illustrated in Fig. 4.1(a) is employed to measure PBTI induced drain current degradation [4.6][4.8]. Details on the system have been described in Chapter 2. I_d rather than V_t is monitored because measurement of V_t requires a sweep in V_g, thus consuming a longer measurement time. The waveforms applied to the gate and drain are depicted in Fig. 4.1(b). Measurement phases are inserted into stress phases on a log-time base. The measurement bias is V_g/V_d =1.2/0.2V and the measurement time (t_{meas}) is chosen to be 50μs such that it is long enough for integrating reliable signals as well as short enough to avoid introducing additional stress.

4.4 Results and Discussions

4.4.1 Two-Stage Drain Current Degradation

As shown in Fig. 4.2, the drain current degradation initially evolves relatively slowly. After a certain stress time, denoted by τ_c in Fig. 4.2, accelerated degradation is observed. The acceleration was also observed in literature (Fig. 4 and Fig. 5 in [4.5]) where the authors attributed the imperfect log-time dependence to a non-uniform initial trap distribution. The drain current degradation versus stress time before τ_c ("the first stage" hereafter) and after τ_c ("the second stage") is plotted in Fig. 4.3(a) and 4.3(b) respectively. The stress conditions are V_g=2V~2.4V, and T=100C. The second stage degradation is obtained by subtracting the extrapolation of the first stage degradation from the measured ΔI_d. Notably, the first stage ΔI_d (Fig. 4.3(a)) has a log(t) dependence while the second stage ΔI_d (Fig. 4.3(b)) exhibits a power law time-dependence with a universal power factor of 0.35 for the

stress V_g of interest. Stress temperature effect is also characterized in Fig. 4.4 and Fig. 4.5. Three points are worth noting: (i) At a higher stress temperature, the I_d degradation enters the second stage earlier, or, a smaller τ_c. (ii) The first stage current degradation has negative stress temperature dependence, i.e., a smaller drain current degradation at a higher stress temperature. Nevertheless, the second stage shows an opposite trend, a positive temperature effect. A crossover of the drain current degradations at T=25C and 125C is noticed in Fig. 4.4 and a larger stress V_g gives an earlier crossover. The opposite temperature dependence implies that the dominant degradation mechanisms in the first and the second stages are different. (iii) The degradation is driven into the second stage earlier at a higher stress V_g. For example, the cross-over time is around 10s for V_g=1.8V and 1s for V_g=2.2V in Fig. 4.5.



4.4.2 The Degradation Model

The log(t) degradation rate in the first stage suggests that charging [4.5] and concomitant discharging [4.6] of pre-existing high-k traps dominate the first stage I_d degradation. A higher V_g induces a larger electron density in the inversion channel readily for trapping, thus causing a more severe I_d degradation. The negative temperature dependence can be explained as follows. Since the high-k charge de-trapping rate increases with temperature [4.6], a higher temperature results in a smaller net charge trapping rate and thus a smaller drain current degradation. On the other side, new high-k traps are created during stress. At a certain stress time (the aforementioned "corner time") the additionally created high-k trap density reaches a level

comparable to or even more than pre-existing ones. Charging and discharging of the pre-existing traps are then no longer the limiting process. Thus, the drain current degradation is dictated by trap generation which exhibits power-law stress time dependence [4.12]. Furthermore, larger stress V_g and higher temperatures lead to faster high-k trap generation in the second stage (Fig. 4.2 through Fig. 4.4) because of larger carrier fluence and energy [4.12][4.13] and thus an accelerated thermo-chemical reaction for trap creation [4.14]. As a result, the device I_d degradation is driven into the second stage earlier (or a smaller τ_c) at higher stress V_g and/or temperature.

4.4.3 High-k Trap Growth Rate

To characterize high-k trap growth, two techniques are employed: a recovery transient technique [4.8][4.9], and a charge pumping (CP) technique [4.12]. The characterization procedures of these two techniques are shown in Fig. 4.6. The devices are firstly subject to PBTI stress followed by a discharging step to empty the high-k traps. The discharging step is necessary or the residual trapped charge generated by PBTI stress would de-trap during the recovery transient or CP measurement, giving rise to incorrect results.

(i)Recovery Transient Technique

After the discharging, a moderate V_g (=1.2V) is applied for 0.2s to fill high-k traps with electrons and the drain current is measured at V_g/V_d=0.25V/0.1V immediately after filling[1]. A recovery V_g as low as around V_t is selected to magnify the effect of the trapped charges on the

[1] The "filling" and "recovery" phases here are exact the same as the "stress" and "measurement" phases in Chapter 3.

channel conductance (thus on I_d). To make the around-threshold I_d accurately collectable in a short integration time, a device with a very large width-to-length-ratio (W/L=100μm/0.08μm) is chosen. Fig. 4.7 is the temporal evolution of the actual I_d. The transimpedance gain needed to make this current level (~μA) differentiable from the background noise should be larger than ~10mV(resolution of the digital oscilloscope)/ 1μA=10^4(Ω). The feedback resistor in our circuit is 100KΩ. For a measurement time of 10^{-4}s, the voltage gain of the operational amplifier is about 10^5, which is sufficient to meet well the virtual ground condition of the circuit. For μA currents, the measurement system can respond reliably with 10^{-4}s, and the slope for the two curves in Fig. 4.7 can be confidently claimed different.

Fig. 4.8 shows the normalized drain current recovery transients before stress, after 1s stress and after 2000s stress. The device dimension is W/L=100μm/0.08μm. A large transient in the milli-second range is observed, suggesting an underestimate of the stress effect from conventional measurement. All curves in Fig. 4.8 have log(t) dependence. The recovery rate, i.e., the slope in Fig. 4.8, for "t=1s" is about the same as for the fresh device and increases considerably for "t=2000s". In Chapter 3, we have shown that the drain current recovery is caused by de-trapping of charges spatially distributed in the high-k layer [4.8][4.9]. The threshold voltage shift as a result of high-k trapped charge emission can be approximated as

$$\Delta V_t(t) = \int_0^x \frac{qN_{HK}^V(x,0)}{\varepsilon_{HK}}(T_{HK} - x)\{1 - \exp[-t/\tau(x)]\}dx \propto \frac{qN_{HK}^V T_{HK}}{\varepsilon_{HK}\alpha_k}\log(t) \qquad (4.1)$$

where N_{HK}^{V} is the volumic trap density in the high-k gate dielectric, ε_{HK} the dielectric constant, T_{HK} the thickness of the HfSiON layer. Other variables have their usual definitions. The corresponding drain current change is then readily obtained as

$$\Delta I_d(t) \propto G_m \Delta V_t(t) \propto \frac{qN_{HK}^{V} T_{HK}}{\varepsilon_{HK}\alpha_k} \log(t) \qquad (4.2)$$

where $G_m=dI_d/dV_g$ is obtained from measurement. It is possibly questioned that even if there would be no per se ΔV_t, lower G_m would lead to smaller extracted linear V_t value as well as to lower I_d. Fig. 4.10 indicates the fresh and post-stress I_d-V_g curves in both a linear and in a semi-log scale. The G_m value indeed changes with stress time, but in the recovery measurement (Fig. 4.8) I_d was measured at a specific stress time and G_m is a constant in the measurement period. Measured ΔI_d and measured G_m are used to extract ΔV_t, and thus the corresponding trap density. For another stress time, ΔI_d and G_m measurements were repeated to obtain a new high-k trap density. The stress induced G_m variation is already considered in the calculation of trap density. In Eq. (4.2), the slope of the ΔI_d-log(t) plot is linearly proportional to the high-k trap density. Therefore, we can extract the high-k trap density from the recovery slope and the result is shown in Fig. 4.11. Our result shows that the high-k trap density in the first stage (τ_c ~ tens of seconds) is dominated by pre-existing traps. The high-k trap density, however, increases drastically in the second stage.

Moreover, as reported in literatures [4.6][4.8] and summarized in Chapter

3, individual electron detrapping from the high-k dielectric can be observed directly in a small area device, which is manifested by a discrete evolution of the recovery drain current rather than a continuous log(t) increase with time. Fig. 4.12 compares the pre-stress and post-stress recovery drain current evolutions in a small area device (W=0.16μm, L=0.08μm). Each current jump in the current evolution accounts for a single charge detrapping. Apparently, the post-stress device has more current jumps in the same measurement period than the pre-stress device (from 1 jump in a fresh device to 5 jumps after 100s stress). This result, again, provides evidence of high-k trap generation during PBTI stress.

(ii) Charge Pumping Technique

A two-frequency CP method is also used to characterize high-k trap generation [4.12]. The experimental procedure is described in Fig. 4.6(b). The PBTI stress condition is identical to that used in the recovery transient technique Fig. 4.6(a). The high-k trap density is obtained from the difference between CP results at two frequencies,

$$N_{HK} = \frac{1}{WLq}\left[\frac{(I_{CP}@2kHz)}{2kHz} - \frac{(I_{CP}@1MHz)}{1MHz}\right]. \quad (4.3)$$

Note that N_{HK} in the above equation denotes the number of traps per unit area. The choice of 2kHz as the lower CP measurement frequency is somewhat arbitrary[2] and the N_{HK} reflects only the relative high-k trap density.

[2] First, a theoretical calculation gives a direct tunneling time of $\tau \sim 10^{-6}$s through a 1.4nm-thick IL (the thickest IL in this work) based on the equation $\tau = \tau_0 \exp(\alpha_{ox} T_{ox})$ [4.15]. Second, extensive SONOS experimental results confirm the above calculation. For example, measured

Fig. 4.13 shows normalized high-k trap growth in various stress conditions. A larger stress V_g (Fig. 4.13(a)) and/or a higher stress temperature (Fig. 4.13(b)) result in a larger trap generation rate and drive the device into the second stage earlier. This is consistent with Fig. 4.5. Fig. 4.14 compares the additionally generated high-k trap density measured from the recovery transient technique and from the CP technique. The N_{HK}^{V} from the recovery slope method is multiplied by the thickness of the high-k layer in Fig. 4.14. Note that the areal high-k trap density extracted from the CP method is about 3 times lower than that from the recovery transient technique. The reason is that high-k dielectric traps are distributed in space. The inverse of the frequency used in CP measurement determines the depth into the high-k dielectric the electron tunneling front can reach. Thus, only high–k traps which has a tunneling time shorter than the inverse of the CP frequency can contribute to the CP current. In other words, only a small portion of high-k dielectric traps can be detected by the CP method. Despite the absolute value of the trap density, the trap growth rate from the two methods follows the same power law time-dependence with an exponent of ~0.33, which is close to that in Fig. 4.3(b).

4.4.4 The Impact of Processing

Appropriate nitrogen incorporation into Hf-silicate can substantially improve the gate dielectric electrical reliability [4.2]. To evaluate the nitrogen effect on HfSiON nMOSFET degradation, two devices "optimal" vs.

---

tunneling time is ~10^{-4}s for a 1.8nm-thick bottom oxide (Fig. 3 in [4.16]). Third, IMEC group successfully used 2.5kHz in CP to characterize high-k trap density in their devices with IL=1nm (EOT=1.9nm) [4.12].

"control", which have different nitrogen profiles in the HfSiON gate dielectric, are compared. The high-k film in the "optimal" sample remains in amorphous while that in the "control" becomes crystallized after 1100C annealing of source/drain dopant activation [4.2]. The PBTI and hot-carrier lifetime and the Weibull plot of time-to-breakdown are improved as indicated respectively in Fig. 4.15, Fig. 4.16, and Fig. 4.17 with other parameters, such as V_t and gate leakage current, not adversely affected. High-k bulk trap generation is suppressed in the "optimal" samples. Detailed comparisons and discussions can be found in [4.2]. The two devices have almost identical I_d-V_g characteristic, as shown in the inset of Fig. 4.18. After PBTI stress, the normalized trap density from the CP method for these two samples is shown in Fig. 4.18. A large trap density difference between high-frequency and low-frequency CP measurement in the control sample suggests considerable high-k trap creation during stress. On the other side, the optimal device shows better robustness against high-k trap generation. Correspondingly, Fig. 4.19 shows the drain current degradation versus stress time in the two samples. Again, the optimal device shows smaller degradation. We also evaluate the stress temperature effect in the optimal sample (Fig. 4.20). An opposite temperature dependence between the first stage and the second stage is still observed. The cross-over time in the sample is about 10^4 s, much longer than ~20s in the control sample.

4.5 Summary

HfSiON gate dielectric nMOSFETs exhibit two-stage drain current degradation in NBTI stress. The first stage drain current degradation is

attributed to the charging of h pre-existing high-k traps and the second stage degradation is mainly caused by new high-k trap creation. The degradation rate and stress temperature effect in the two stages have been characterized. Larger stress gate voltages and/or elevated temperatures shorten the time required to enter the second stage. High-k trap density is characterized by two different methods: the recovery transient technique developed in Chapter 3, and the two-frequency charge pumping technique. The trap growth rate follows power-law stress time dependence in the second stage. Finally, an optimized nitrogen profile in HfSiON layer is demonstrated to have better stress immunity and thus prolongs the time to enter the second stage. An accurate PBTI lifetime extrapolation strategy should be built upon the physics considering both degradation stages.

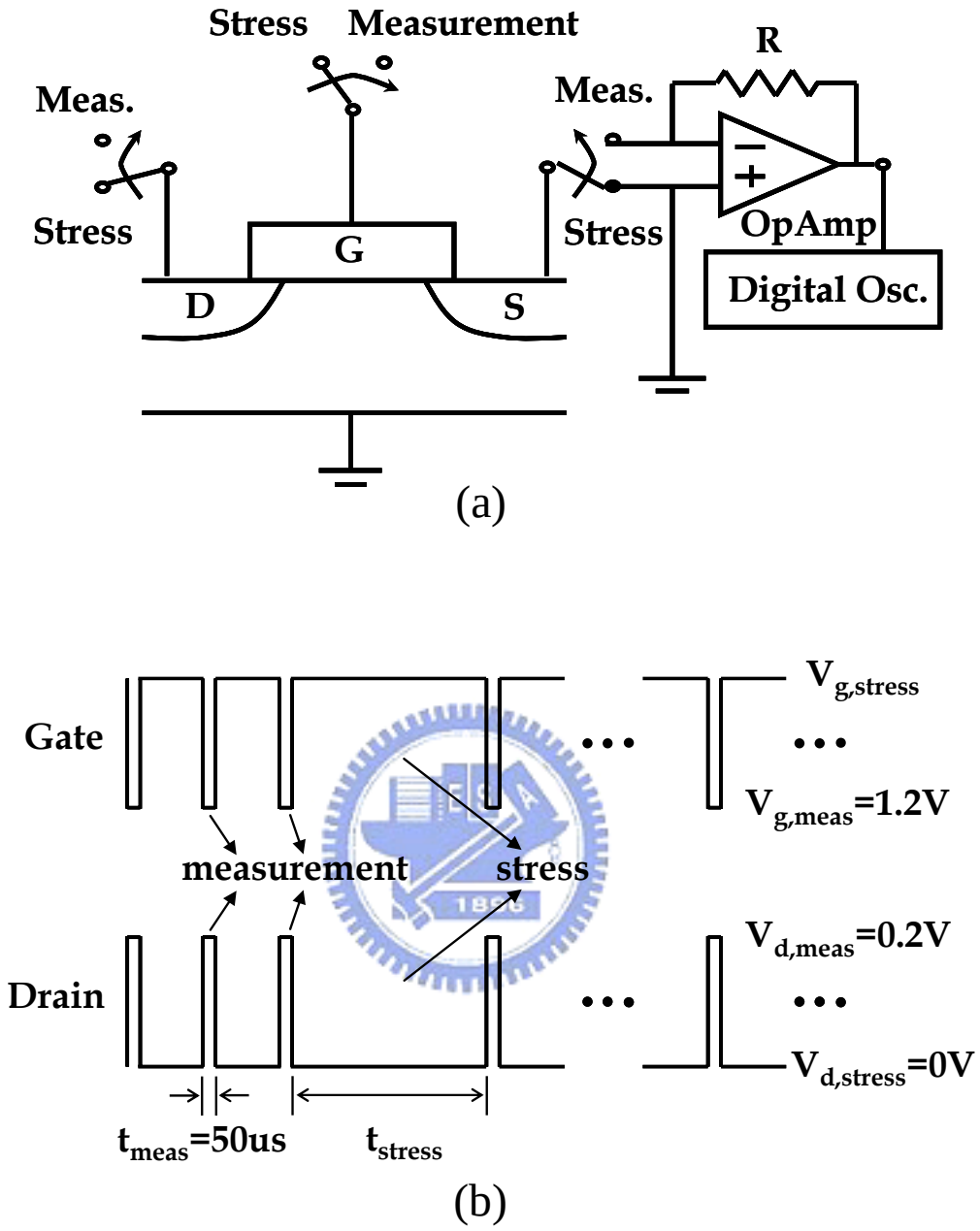


Fig. 4.1 (a) Setup for fast transient measurement as described in Chapter 2. (b) Pulses applied to gate and drain for stress and measurement. The stress time (t_{stress}) is scheduled on a log-time base and the stress gate voltage ($V_{g,stress}$) varies. The measurement condition is $V_{g,meas}$=1.2V and $V_{d,meas}$=0.2V. A measurement time (t_{meas}) of 50 μs is chosen such that it is long enough to integrate signals reliably and short enough to avoid additional degradation.

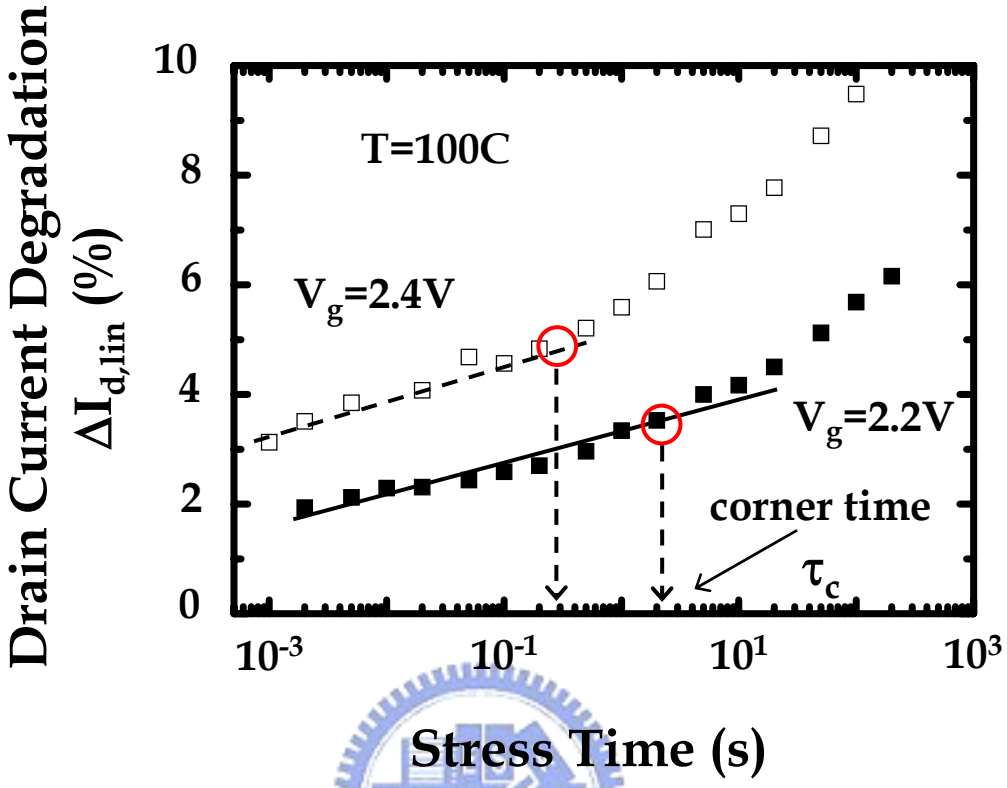


Fig. 4.2 Linear drain current degradation as a result of PBTI stress at V_g=2.2V and 2.4V, T=100C. Two-stage degradation is observed. The onset time of the accelerated degradation is denoted by τ_c (corner time). The first stage degradation is the degradation before τ_c, and the second stage is after τ_c.

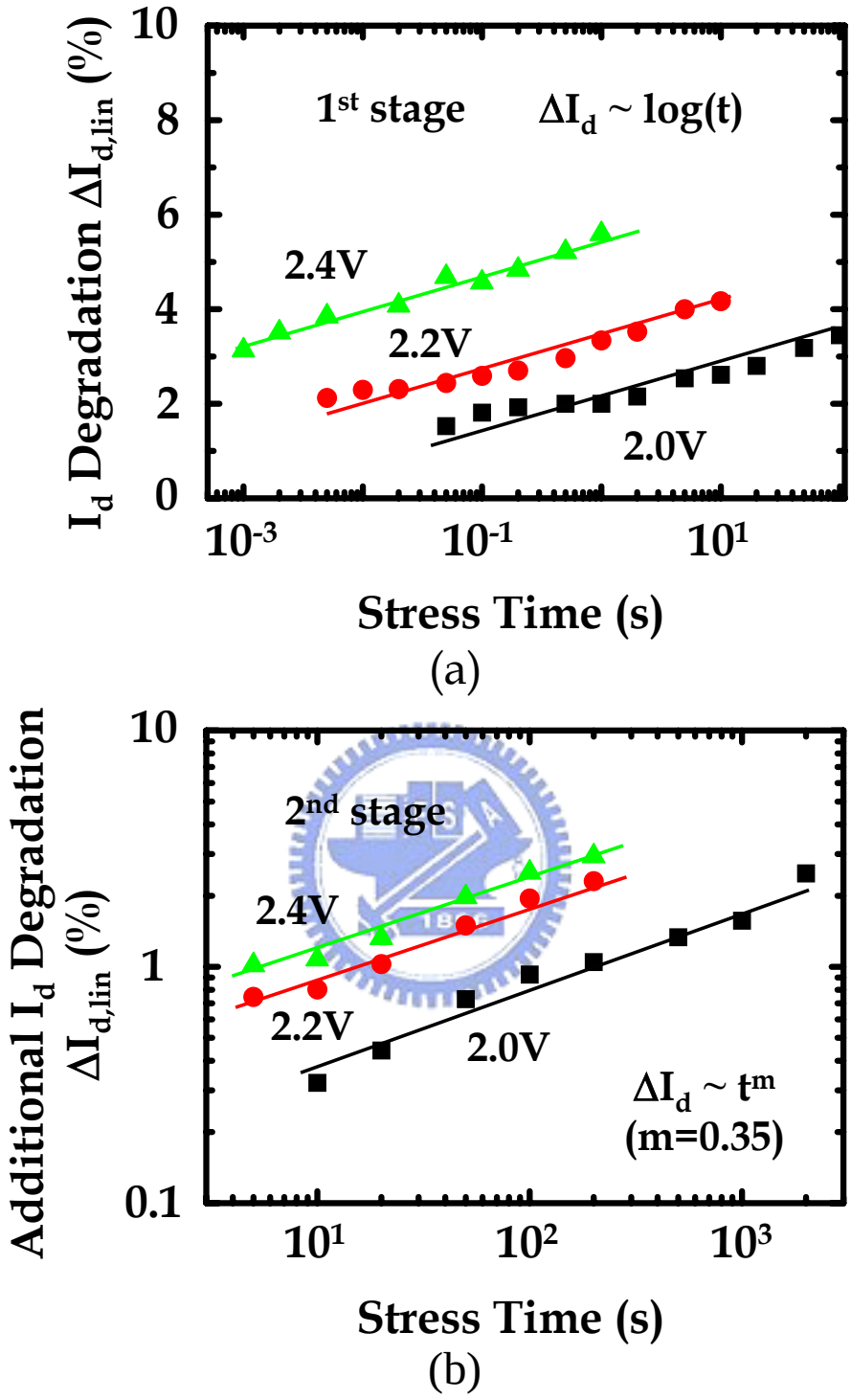


Fig. 4.3 The drain current degradation rate in the first stage (a) and in the second stage (b). The second stage degradation is obtained by subtracting the extrapolation of the first stage degradation from the measured ΔI_d.

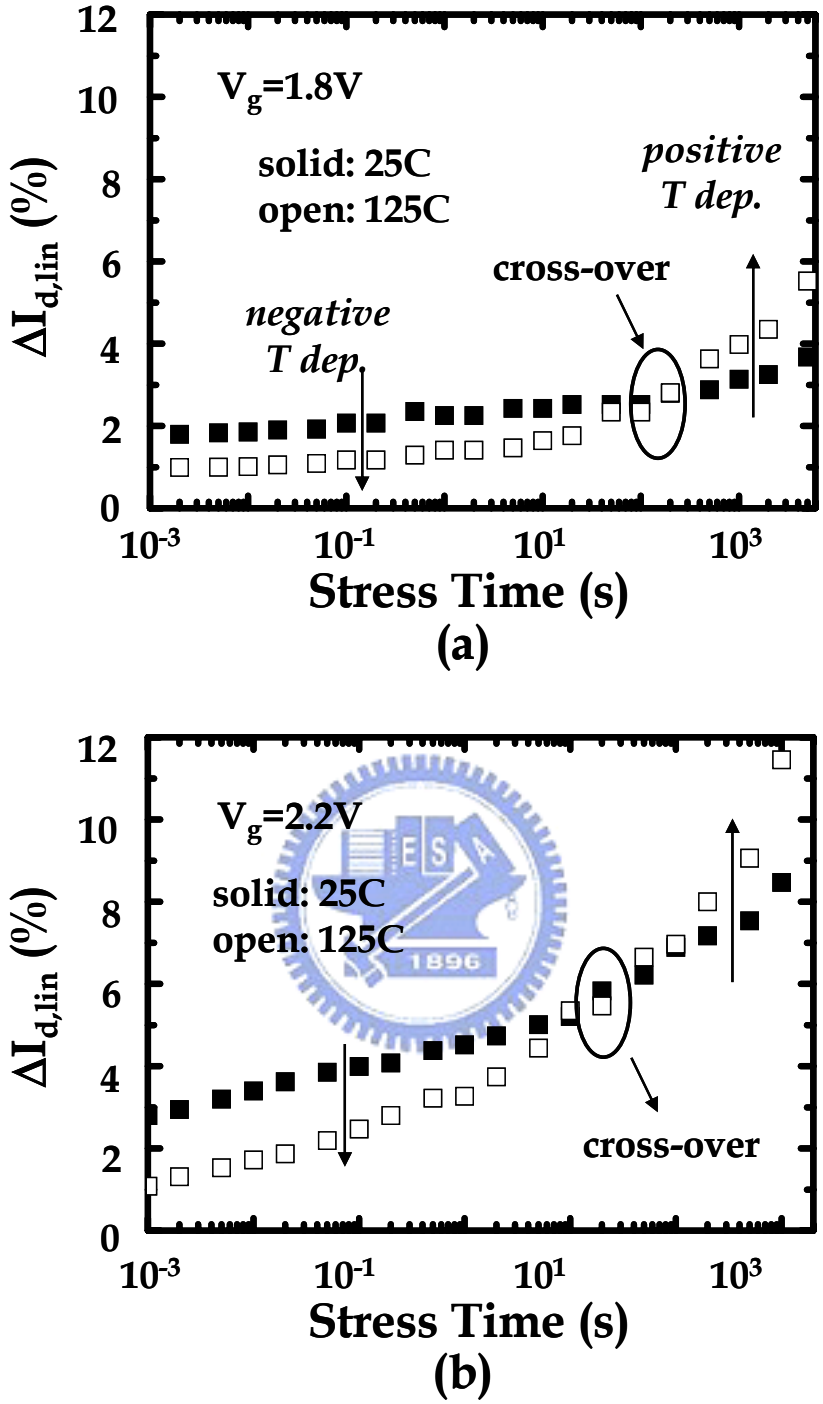


Fig. 4.4 Stress temperature effect on I_d degradation for V_g=1.8V (a) and 2.2V (b). A cross-over of the I_d degradation at T=25C and 125C is noticed and a larger stress V_g gives rise to an earlier crossover.

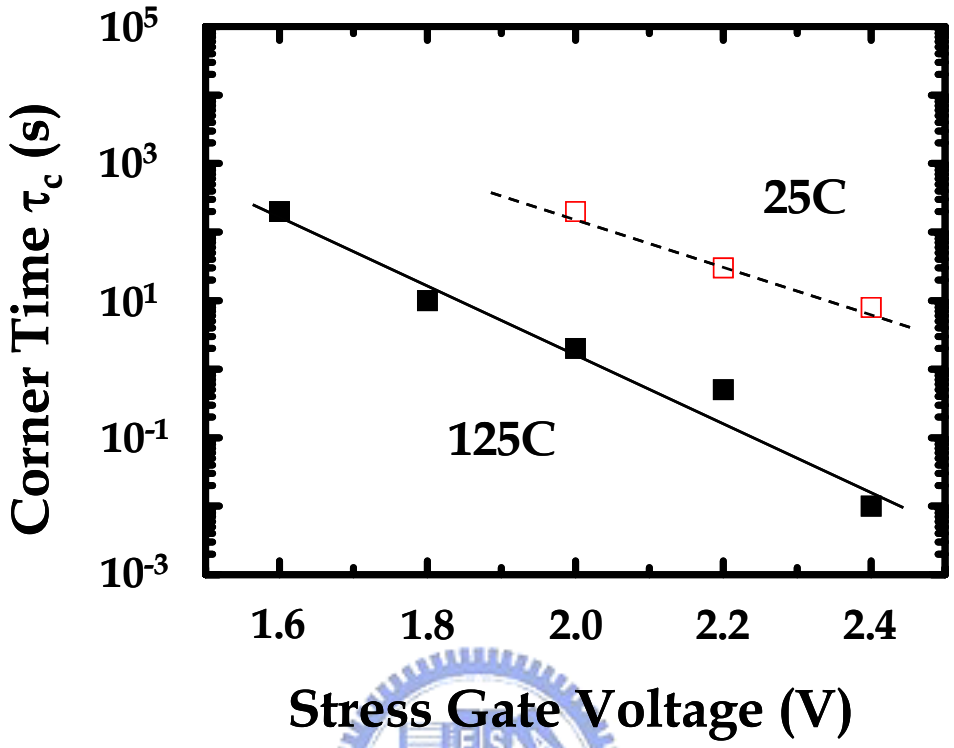


Fig. 4.5 The corner time τ_c versus stress V_g at T=25C and 125C. The I_d degradation is driven into the second stage earlier (a smaller τ_c) at a higher stress V_g and temperature.

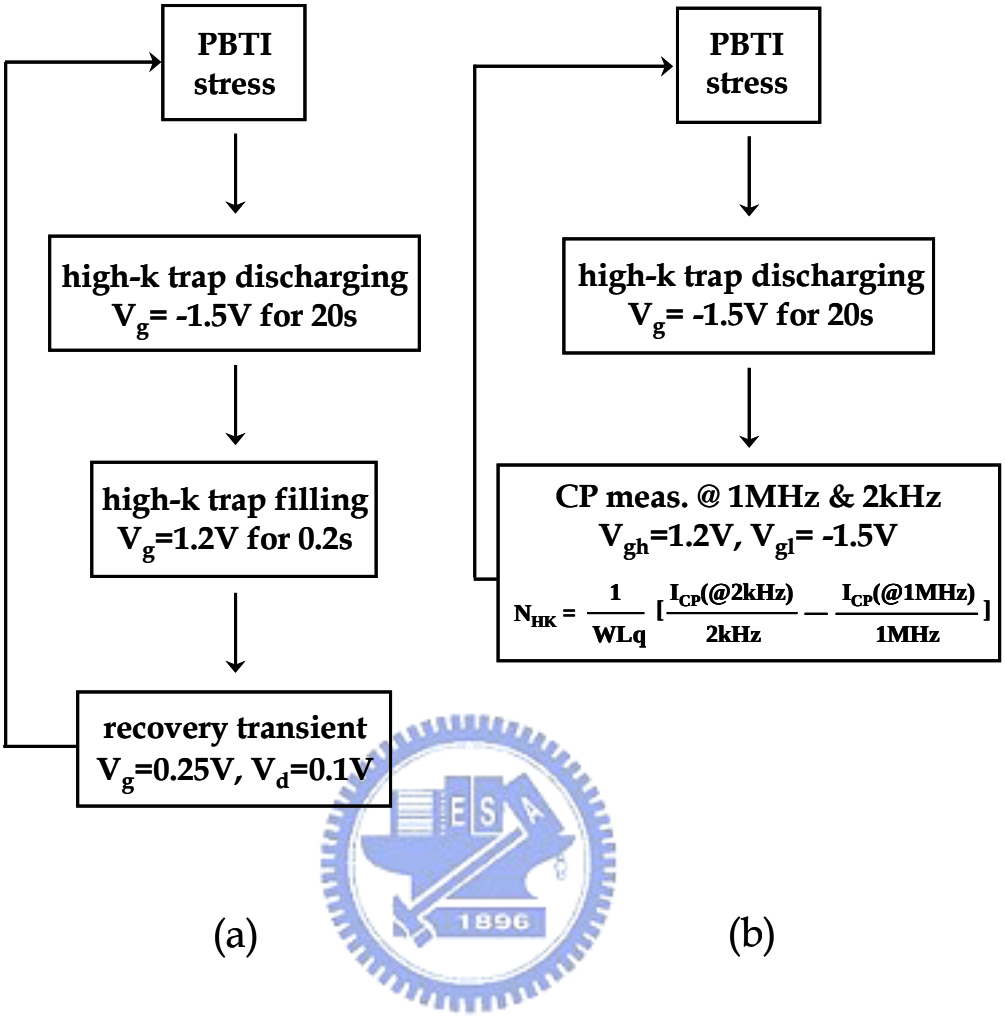


Fig. 4.6 The characterization procedure of the two methods for high-k trap density extraction. (a) A drain current recovery transient technique. (b) A two-frequency charge pumping technique.

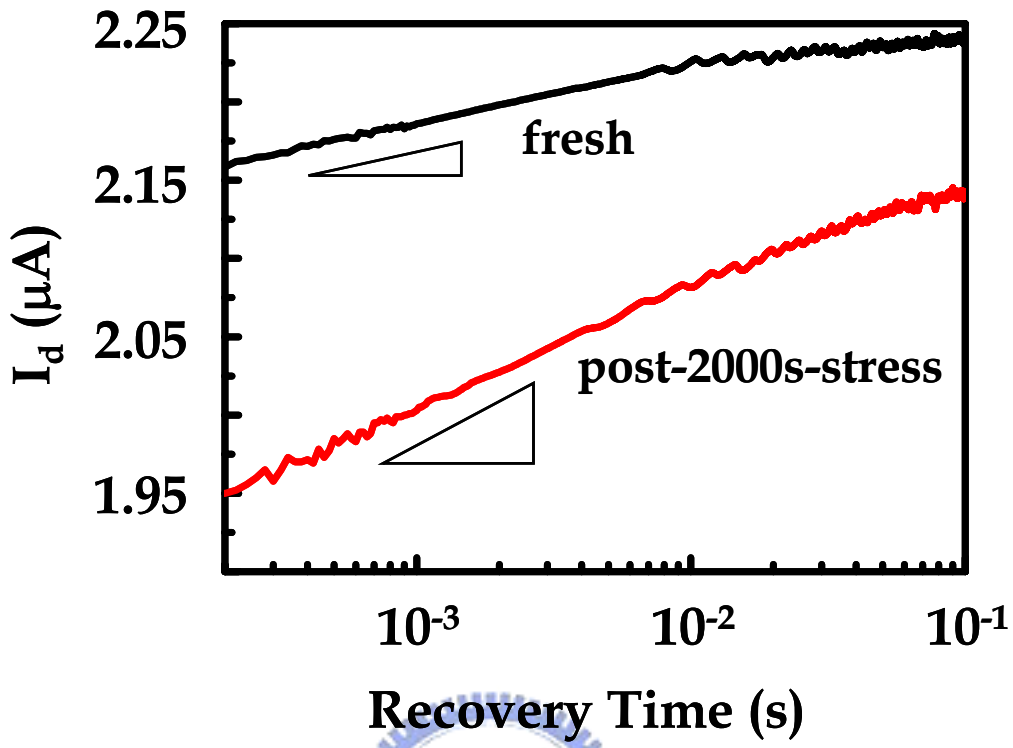


Fig. 4.7 Drain current recovery transients measured at V_g/V_d=0.25V/ 0.1V in a fresh and in a post-2000s-PBTI-stressed device.

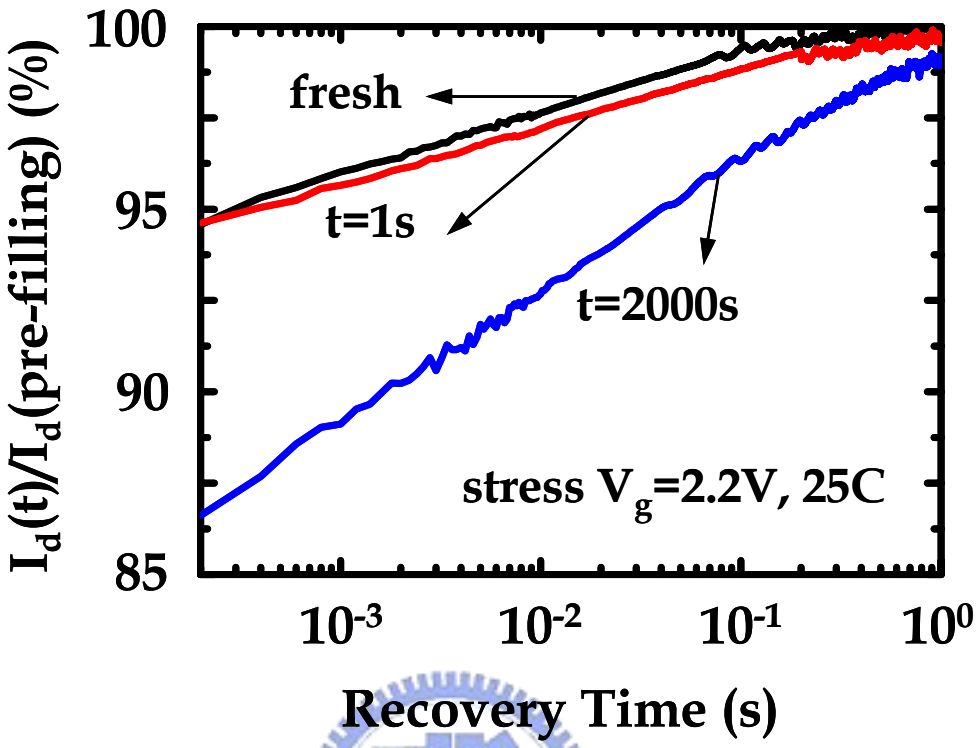


Fig. 4.8 Normalized drain current recovery transients measured at V_g/V_d=0.25V/0.1V for different PBTI stress times: t=0s, 1s, and 2000s. Devices dimension is W/L=100μm/0.08μm. The increased slope implies additional traps are generated during stress.

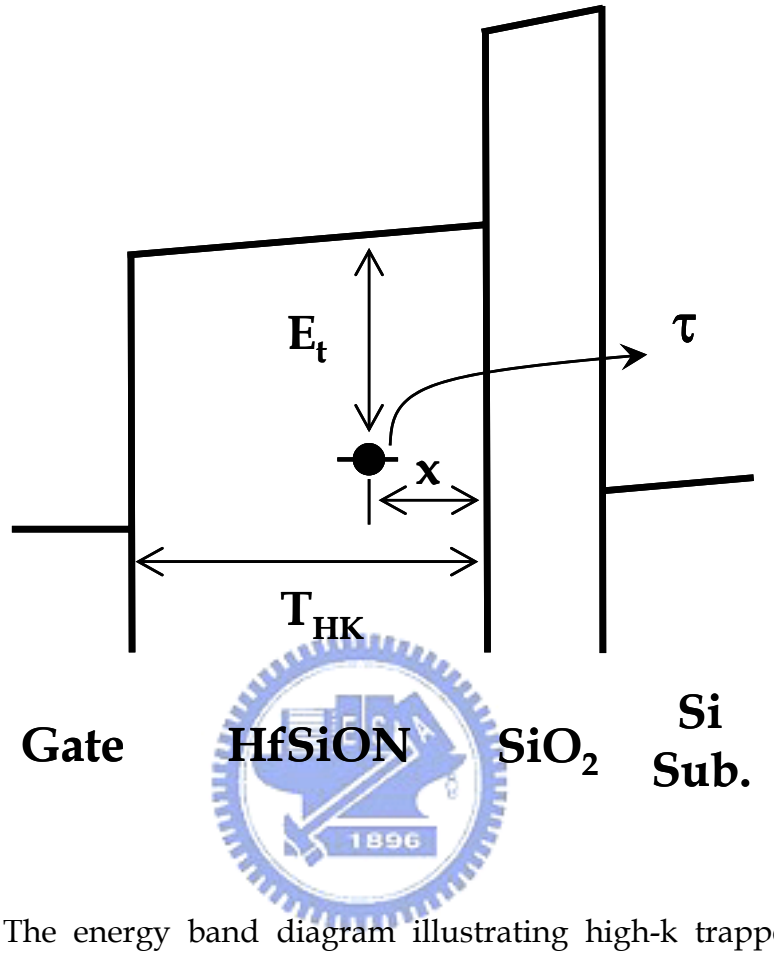


Fig. 4.9 The energy band diagram illustrating high-k trapped charge de-trapping during drain current recovery transient.

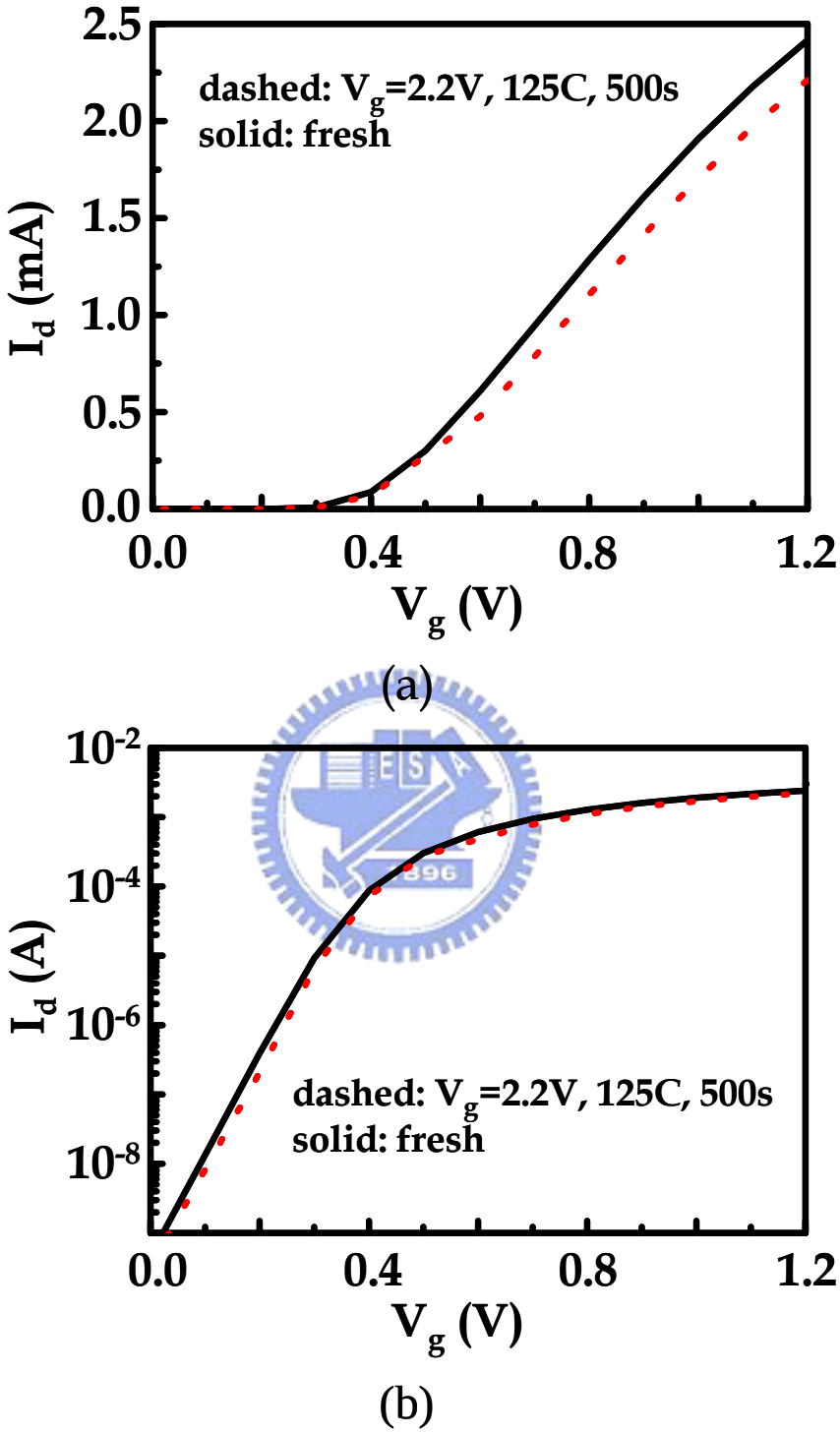


Fig. 4.10 The pre- and post-stress I_d-V_g characteristics in a linear scale (a) and in a semi-log scale (b).

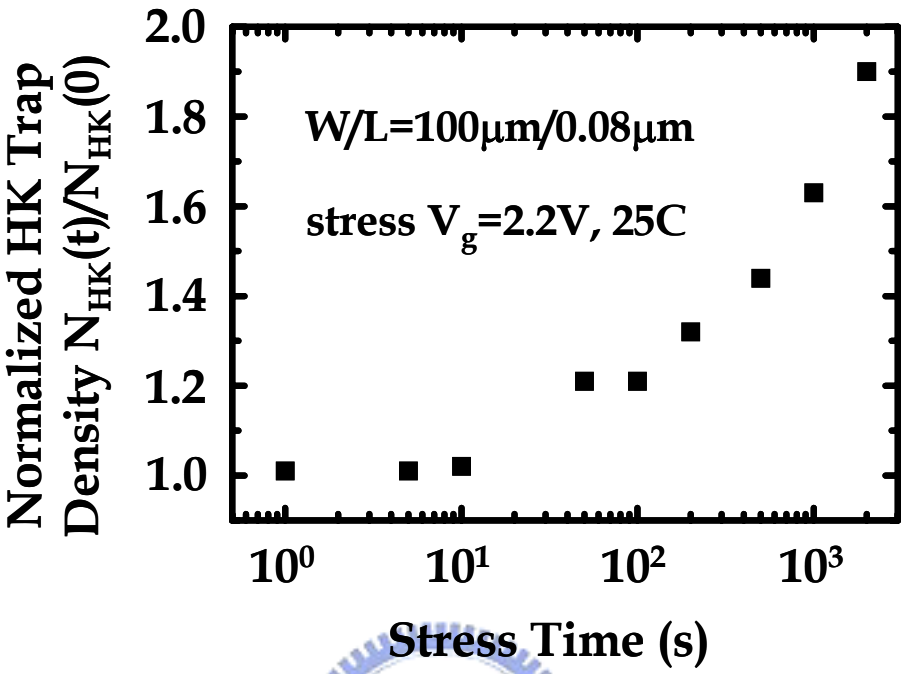


Fig. 4.11 High-k trap density versus stress time from the recovery transient technique. The trap density is normalized to the initial high-k trap density.

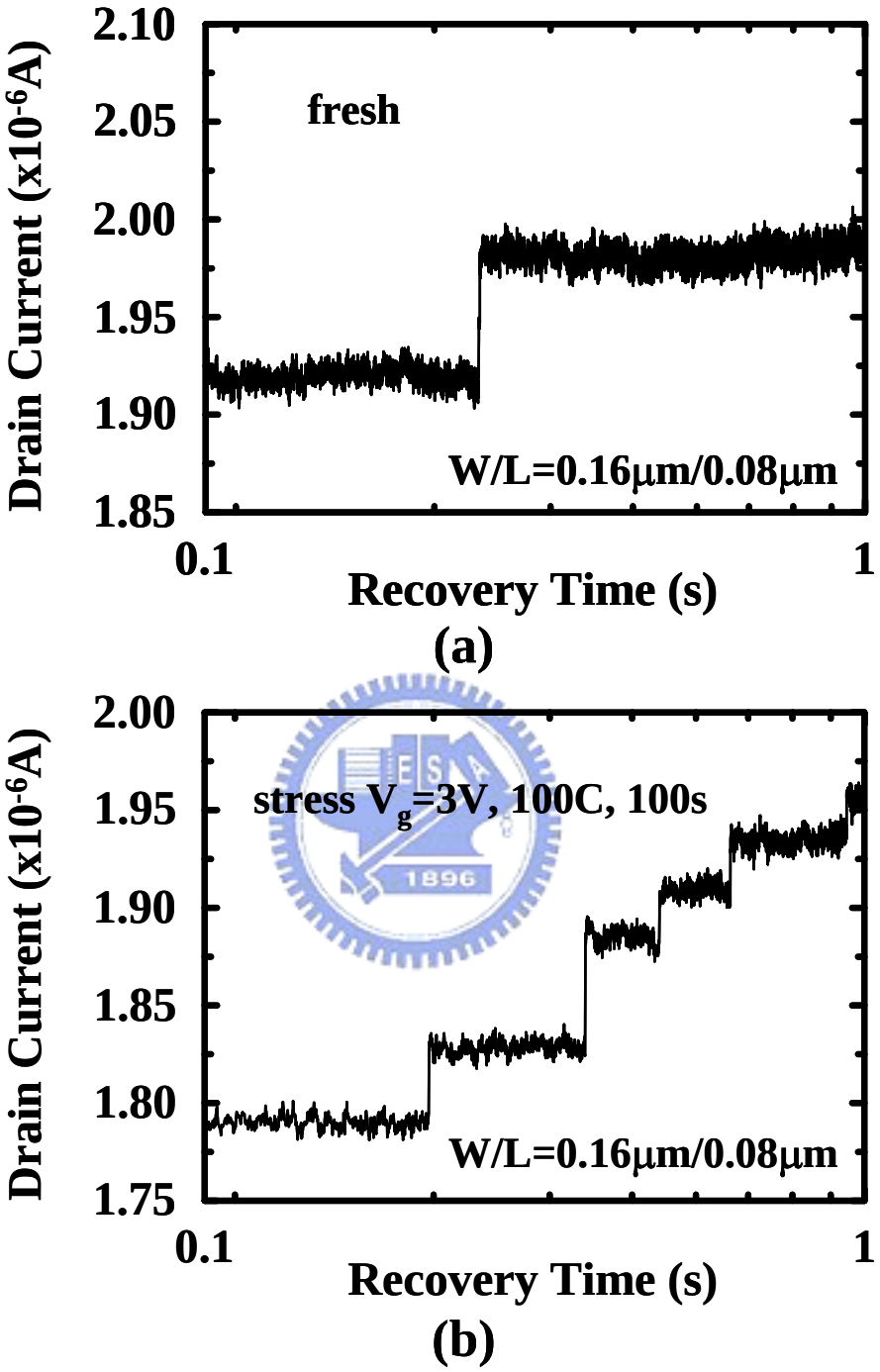


Fig. 4.12 Drain current recovery transient in a small-area device with W/L=0.16μm/0.08μm: (a) for stress time=0s only one current jump is observed, and (b) after stress at V_g=3V, T=100C for 100s, up to five current jumps are presented. The increase in the number of current jumps indicates additional traps are created during stress.

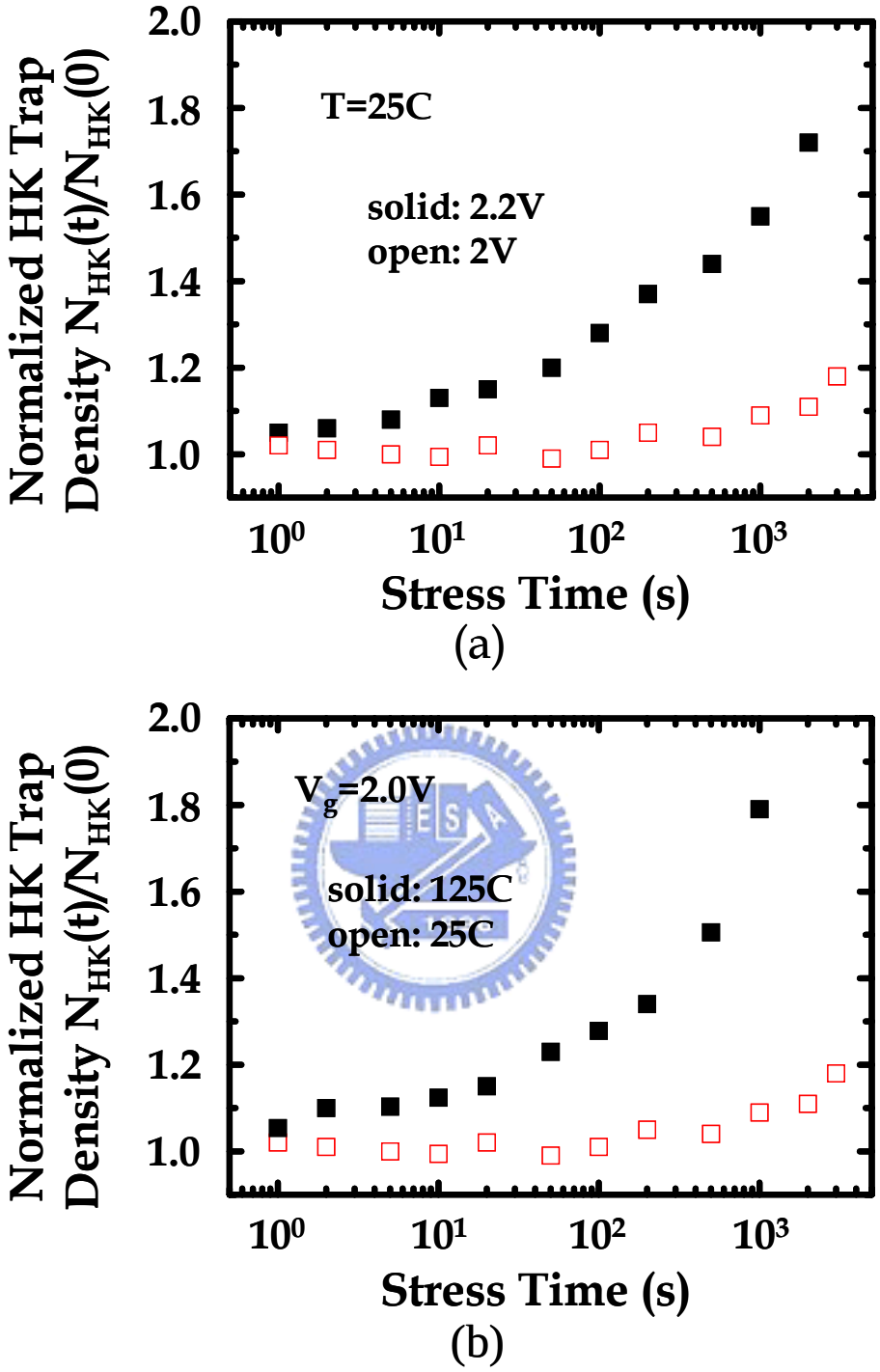


Fig. 4.13 High-k trap density versus stress time from the CP technique. The trap density is normalized to the initial high-k trap density. (a) Stress temperature=25C, (b) stress gate voltage=2V.

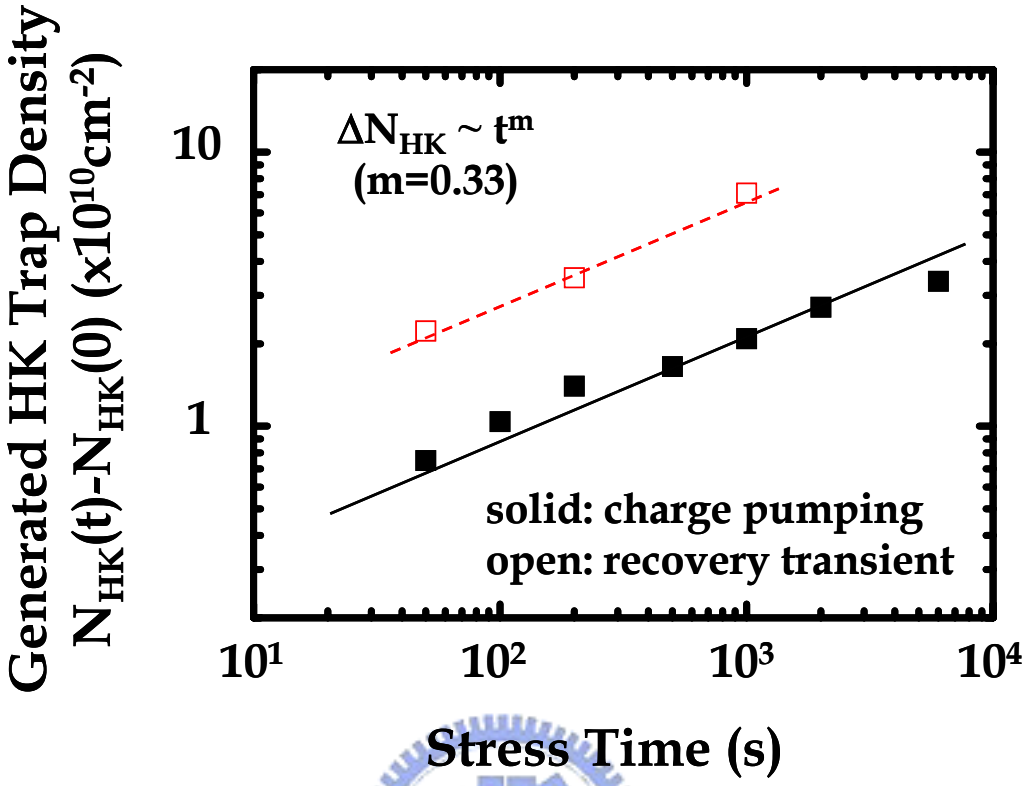


Fig. 4.14 Generated high-k trap density versus stress time from the recovery transient and the CP techniques. The stress condition is V_g=2.2V, T=25C.

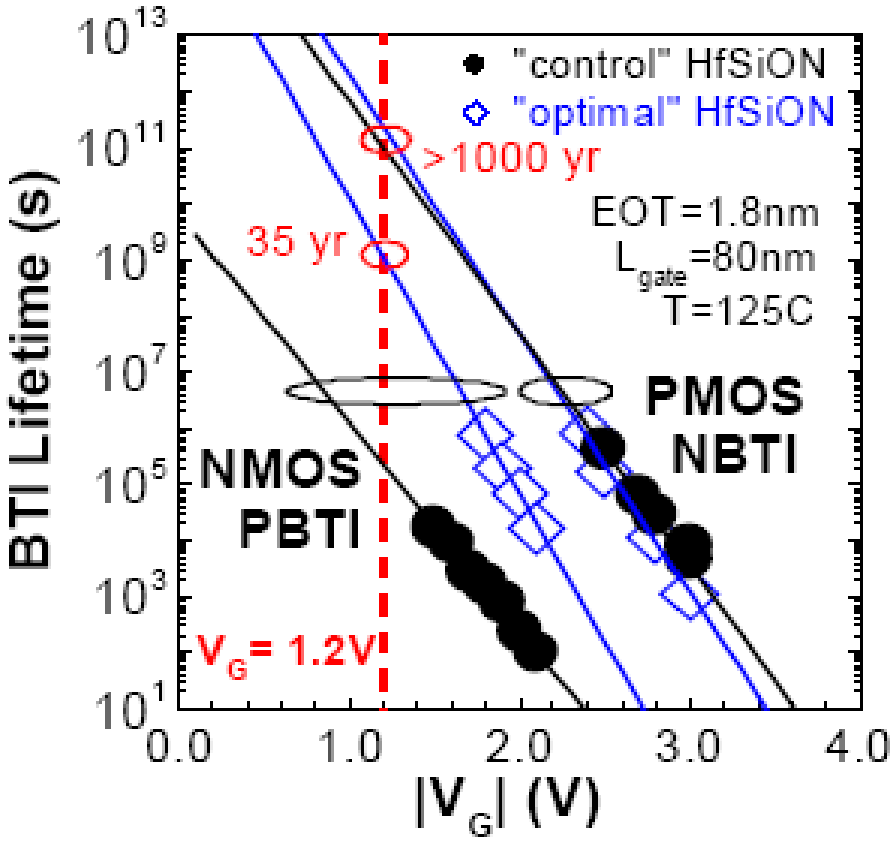


Fig. 4.15 Comparison of BTI-limited lifetime between the "optimal" and the "control" devices [4.2]. The BTI measurements are performed by using a conventional characterization method.

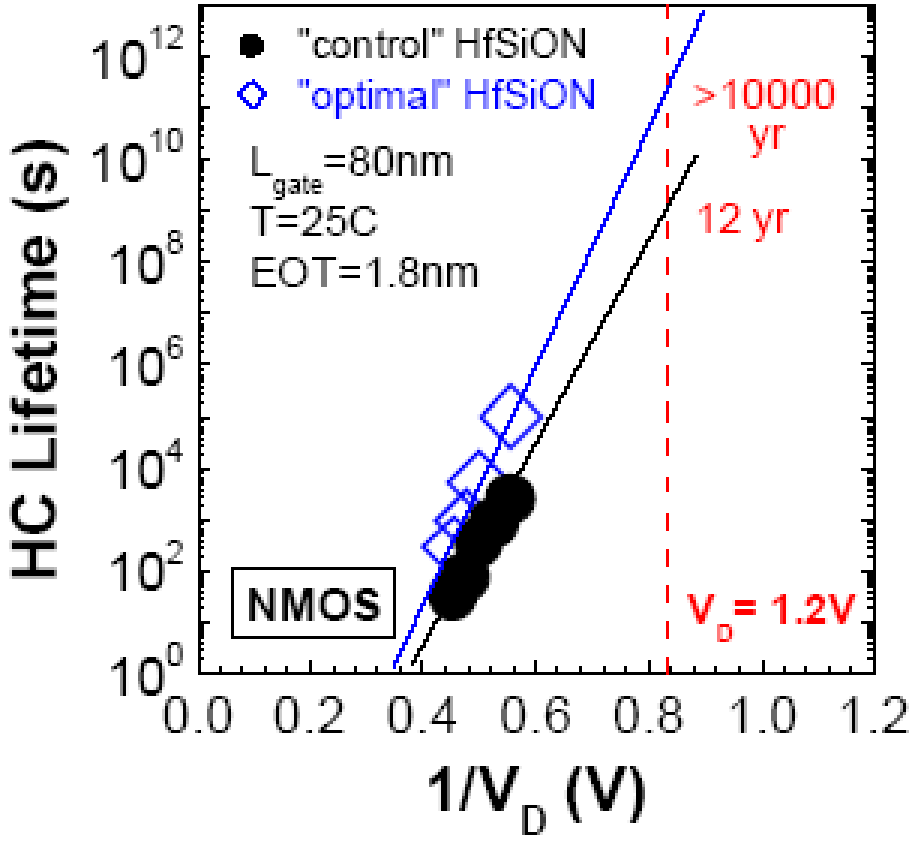


Fig. 4.16 Comparison of hot carrier injection limited lifetime between the "optimal" and the "control" devices [4.2]. The measurements are performed by using a conventional characterization method.

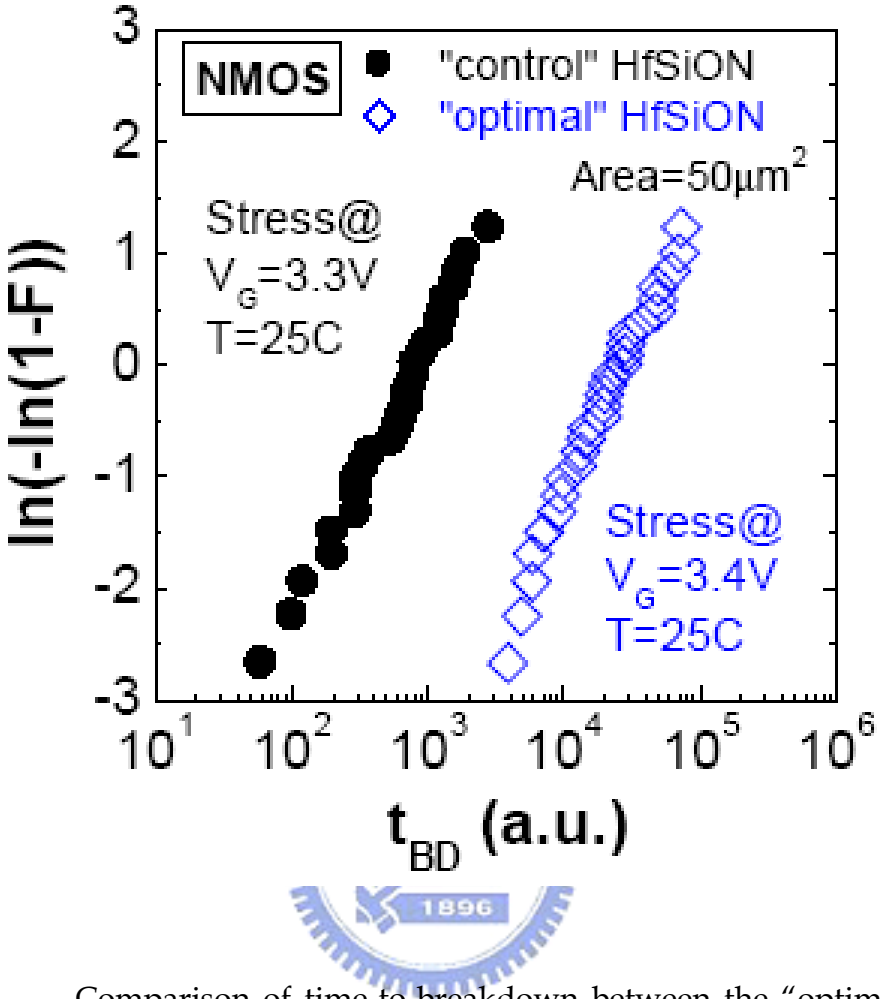


Fig. 4.17 Comparison of time-to-breakdown between the "optimal" and the "control" devices [4.2].

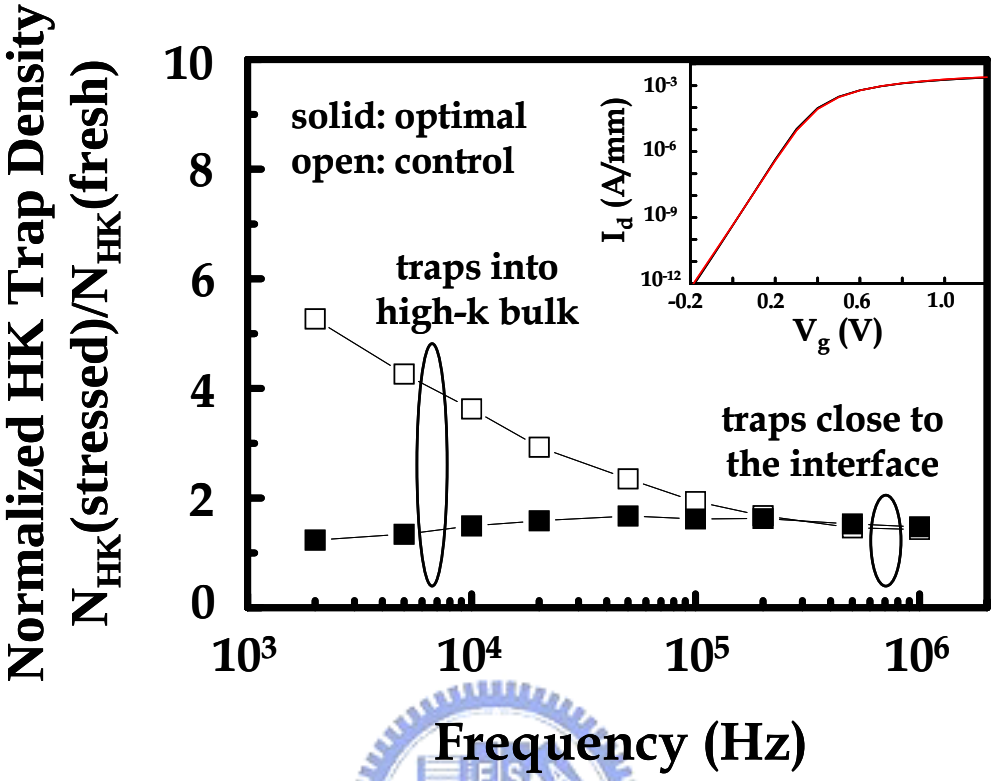


Fig. 4.18 PBTI stress robustness of two devices: "optimal" vs. "control." While both have an identical transfer characteristic shown in the inset, the device with optimal nitrogen incorporation presents a better immunity against PBTI stress in terms of high-k trap generation.

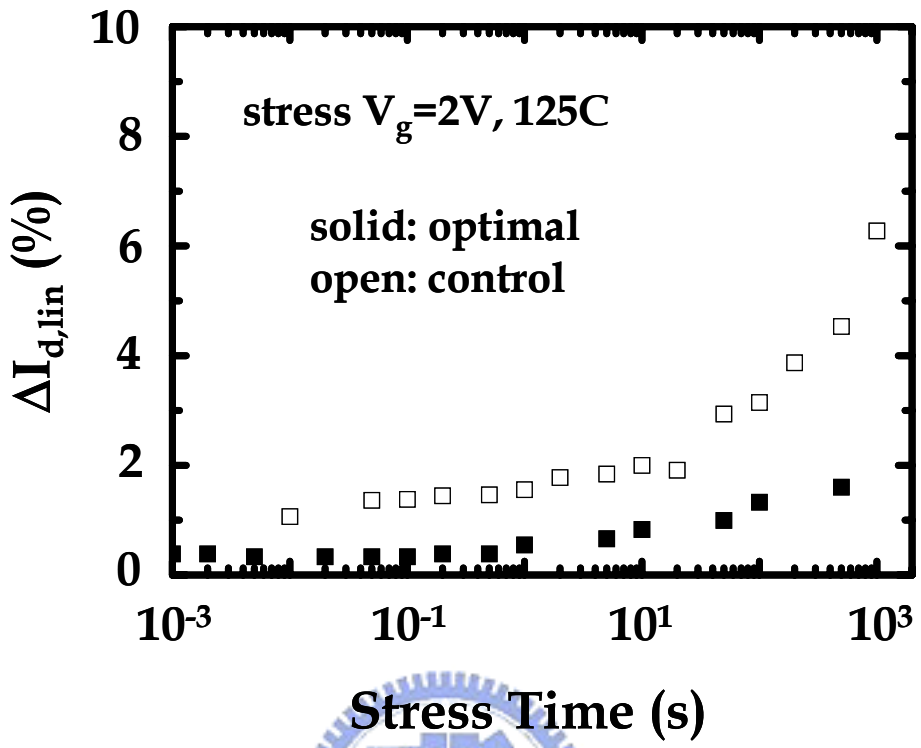


Fig. 4.19 Comparison of I_d degradation between "optimal" and "control." The optimal samples show a lower initial trap density and better stress immunity (i.e., a larger τ_c).

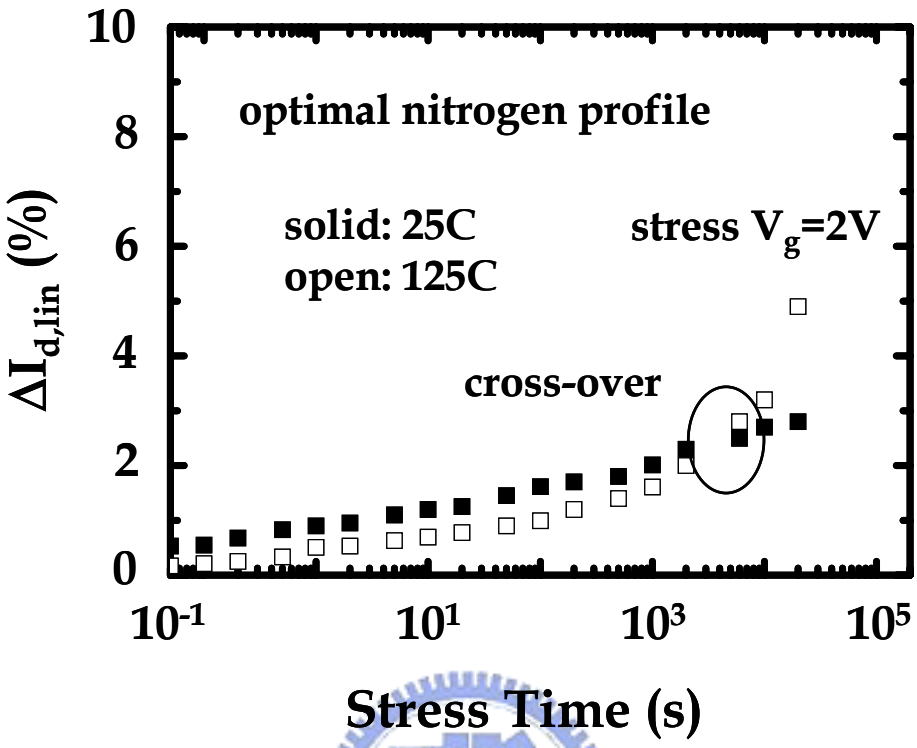


Fig. 4.20 Stress temperature effect on I_d degradation rate in the "optimal" sample. The cross-over is still observed, indicating the existence of two-stage degradation.