13. **Lecture 13: Scaling Theory of HCI Degradation**

13.1 **Hot Carrier Injection Degradation**

Unlike NBTI, HCI degradation is a phenomenon that is observed in NMOS transistors. In NMOS transistors, during its operation, the charge carriers, electrons, gain enough energy (in other words, they become hot!) to initiate an impact ionization process within the substrate. The electrons and holes generated as a result of this process may get injected into the gate oxide thereby creating defects in the interface. These defects degrade the characteristics of the MOSFET by shifting the threshold voltage ($V_{th}$), as a result affecting the flow of current in the device. One may ask, why the HCI degradation is not usually observed in PMOS. It is because the charge carriers in PMOS, holes, have a lower mobility than electrons, and are easier to scatter than electrons. As a result, the average energy of the charge carriers in PMOS is low and thus making them incapable of initiating the impact ionization process.
13.2 Universality of HCI Degradation

At first glance, unlike NBTI degradation the degradation caused by hot carriers does not seem to be universal for all devices. For example, when four devices are tested using four different (accelerated testing) voltages, their change in current (Charge pump current) vs. time plots reveal four different plots which are not parallel to each other and neither do they share the same exponent (power law, \( \sim t^a \)). What is interesting is that when each of these curves is laterally shifted by some factor, they all turn out to be parts of the same plot measured at different times!
When a single sample was stressed for a considerable amount of time and its degradation observed, as can be seen from the plot above and the shifted plots from figure 1, the measured degradation curves (change in the charge pump current due to defect formation) does indeed resemble the universal curve. Therefore, the universality is ensured because any MOSFET whose HCI degradation is observed will produce a curve similar to the one above.

![Degradation curves](image)

Fig. 13.3: Degradation curve of a single device stressed at a certain voltage compared with the universal plot

### Comparison of HCI degradation with NBTI

<table>
<thead>
<tr>
<th>NBTI</th>
<th>HCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predominantly a PMOS related reliability issue</td>
<td>Predominantly an NMOS related reliability issue</td>
</tr>
<tr>
<td>Degradation occurs when the gate voltage ( V_G ) is negatively biased</td>
<td>Degradation occurs at various gate voltage ( V_G ), but the maximum amount of degradation occurs when ( V_G \approx V_D/2 )</td>
</tr>
<tr>
<td>Degradation caused due to non-energized (cold) holes</td>
<td>Degradation caused due to highly energized carriers (hot carriers) that can cause impact ionization</td>
</tr>
</tbody>
</table>
Traps generated are uniformly distributed | Traps are generated near the drain of the device and hence the damage caused is localized
---|---
Only breaking of Si-H bonds are involved | Breaking of Si-H and Si-O bonds are involved

### 13.3 Empirical Observations of HCI degradation

Hot Carrier Induced degradation of MOSFETs, has been studied under various conditions. Below are some plotted experimental, that speaks about the dependence of HCI degradation on the operation time, voltage and temperature of the device.

![Graphs showing time, voltage and temperature scaling](image)

(a)Time Exponent  (b)Voltage Scaling  (c)Temperature Scaling

Fig. 13.4: Time, Voltage and Temperature dependence of HCI degradation

As can be observed from the Fig. 13.4(a), the degradation does indeed deteriorate with time. An interesting observation in this plot is the significant improvement (couple of orders of magnitude) in the lifespan of the device when the dangling Silicon bonds were passivated with Deuterium (H$_2^2$), rather than Hydrogen (H$_2$). The HCI degradation also shows a power law ($\sim t^n$) time dependence, with the time exponent ranging from 0.3 to 0.7.

From Fig. 13.4(b), it can be observed that the degradation worsens with higher operating voltages (V$_{DD}$). This is understandable, since at higher voltages (with a given gate voltage) the charge carriers get more energized causing an increase in impact.
ionization efficiency. As a result at higher supply voltages, more hot carriers are generated to increase the effect of HCI degradation.

The most interesting observation is made in the plot of Fig. 13.4(c), where one can observe that the degradation reduces, as the temperature is increased! One of the reasons behind this could be due to reduced phonon scattering of channel electrons at lower temperatures. This increases the average energy of the channel electrons and therefore the number of hot carriers capable of causing impact ionization also increases. Therefore, a device working at a higher temperature will most likely have more electron scattering, thereby reducing the average electron energy and causing the number of degradation causing hot carriers to drop.

**Mechanism of HCI Degradation:**
To better understand the physical mechanism of the HCI degradation process, observe the following cases:

![Fig. 13.5: (right) CMOS inverter, (top) transition from high to low of VA with the transitions of low to high of VB, (bottom) drain current of the NMOS during the transition](image)

In a CMOS inverter, when the applied voltage (\(V_G = V_A\)) goes from high to low (1→0), the drain voltage of the NMOS (\(V_D = V_B\)) goes from low to high. As will be explained later, the NMOS experience maximum amount of hot carrier stress during this transition, especially when \(V_G \sim V_D/2\). Once the input of the CMOS inverter goes to 0V,
the NMOS is turned off (no current flow), whereas the PMOS turns on and may experience NBTI degradation.

The operation of an NMOS at various gate voltages is shown below:

**Case 1** ($V_G = 0V$): The input voltage ($V_G$) is 0V, and therefore the output voltage of the inverter ($V_D$ of the NMOS) is $V_{DD}$. As a result, as can be observed from the band diagram in the right, the energy band is highly bent to the right. There is a high electric field which can accelerate electrons towards the drain, but due to the transistor being turned off, the electron density in the channel is miniscule. Whatever, electron does enter the high-field region gains enough energy to cause impact ionization. The holes generated from these impact ionizations have high enough energy to be injected into the oxide due to favorable electric field between the gate and the drain.

**Case 2** ($V_G \to V_D/2$):

![Diagram showing NMOS energy band at $V_D = V_{DD}$ for $V_G = V_D/2$](image)

Fig. 13.6: *NMOS energy band at $V_D = V_{DD}$ for $V_G = V_D/2$*

As the input voltage increases, once the gate bias exceeds the threshold voltage ($V_{th}$) channel electrons begin to appear. Once the gate voltage reaches around $V_G \sim V_D/2$, due to the reduction of barrier height ($\Phi_B$), the number of electrons in the channel increases significantly. Moreover, due to the gate induced vertical field the horizontal electric field near the drain is reduced. This makes the hot electrons in this region to interact with phonons and subsequently lose energy; as a result the holes generated through impact ionization by these electrons are of lower energy. Therefore, these holes have a reduced probability of getting injected into the oxide layer.
Case 3 \((V_G \rightarrow V_D)\):

The electric field close to the drain region reduces, as the gate voltage is increased to \(V_G = V_D\). The electric field in this region is given by:

\[
E_e = \frac{V_D - V_{D,SAT}}{l}
\]

where, \((V_{D,SAT})\) is the saturation voltage and \((l)\) is the pinch-off length.

As the gate voltage \((V_G)\) is increased, the saturation voltage \((V_{D,SAT})\) also increases, which decreases the electric field which accelerated the electrons. Therefore, even though the number of electrons increases with \((V_G)\), the number of hot carriers (energized electrons) reduces, thereby diminishing the impact ionization efficiency. Hence, as the gate voltage further \((V_G)\) increases, the electron density in the channel increases but the number of HCl degradation causing hot carriers decreases.
Fig. 13.8: (Top) electron ($I_e$) and hole ($I_h$) component of the gate current in an NMOS), (Middle) net gate current ($I_{G,\text{Net}}$), (Bottom) Causes of degradation in different regions (different gate voltages ($V_G$)).

Therefore to summarize the mechanism of HCI degradation at different gate voltage operations, observe the curves in the figure above. Initially, when the gate voltage is close to zero, due to the presence of a steep band bend at the drain side of the NMOS, an occasional electron in the channel gains enough energy to impact ionize and create ‘hot holes’ (that constitute the gate current). These hot holes are directed towards the gate oxide and break Si-O bonds (as observed in Region I of the figure above). The hot electrons, on the other hand, will be repelled by the electric field and very little fraction of them can enter the gate, as a result their contribution to the gate current at this stage is negligible. As the gate voltage increases, the ($I_e$) electron component of the gate current begins to rise, and since the hole and electron components of the gate current have opposite polarity, the net gate current forms a bell shaped curve in this region (due to the two components cancelling each other).

When $V_G \sim V_D/2$ (Region II), the number of hot electrons increases significantly, that results in an increase in density of broken Si-H bonds, thereby generating defects that ultimately degrade the device. In regions III and IV, as can be observed for high $V_G$ ($\geq V_D$), the hot carrier density decreases and therefore so does the degradation. If the $V_G$ is
high enough, the electrons injected into the gate can be trapped in the oxide and can cause shift in threshold voltage ($V_{Th}$).

### 13.4 Difference between hydrogen and deuterium passivation

As previously observed from one of the empirical observations, it was noticed that passivation of the Si dangling bonds with Deuterium ($H_2^+$) rather than Hydrogen ($H_2$) made significant difference in the lifetime of the device. To understand why it was more difficult to break Si-D bond rather than Si-H bonds, it was initially presumed that the heavier weight of deuterium in comparison to hydrogen played a role. It was thought that a heavier deuterium would be more difficult to displace far from its equilibrium location than a hydrogen atom, hence the bond dissociation would be more difficult for deuterium when compared to hydrogen. It was later realized that due to the higher mass of deuterium (compared to hydrogen), the vibrational spectra is for Si-H and Si-D are different. The vibrational frequency of Si-D bond, as can be observed from the figure below, is far closer to the frequency of the transverse optical (TO) mode of bulk Si than that of the Si-H bond. Therefore, the vibrational mode of the Si-D is strongly coupled to Si(TO), than that of the Si-H mode and thus the vibrational energy of the Si-D bond can flow/decay into the TO mode of bulk Si at a faster rate than that of Si-H bond. Hence, cascading (which eventually leads to bond dissociation) in Si-H is much easier than that in Si-D since immediate relaxation takes place in the latter due to good coupling.
Fig. 13.9: (a) Generation of $N_{IT}$ when passivated with hydrogen and deuterium, (b) Desorption Yield vs. Sample voltage for Si-H and Si-D bonds (Hess’ Experiment), (c) Frequencies associated with various modes

13.5 Multi-Terminal Measurements:

During the on state of the device, the channel inversion current is responsible for generation of hot carriers. In order to determine the dominant leakage current responsible for the generation of hot carriers in the off state, the drain current ($I_D$) was measured under various configurations. The device was biased in several configurations as shown in figure (b), so as to observe the drain current by shutting off the various components of the leakage current. It was observed that when both the gate and source were floated, the only current that showed up was the bulk leakage current. Therefore, out of the four configurations shown above, only the surface field initiated band-to-band tunneling current is the main contributor of $I_D$, and hence the dominant leakage current responsible for the generation of hot carriers in the off state.
Fig. 13.10: (a) $I_D$ measurement for various configurations to identify the dominant current component. (b) Current components for individual configurations: Active components are indicated by the arrows and inactive components are crossed out. The same value of off-state $I_D$ (e.g., $V_G = 0V$, $V_D > 5V$) for all configurations except 2T indicate a surface field initiated BTBT as the dominant current component, responsible for HCI damage of NMOS transistors. [Figure from “Universality of Off-State Degradation in Drain Extended NMOS Transistors”, Varghese, Kufluoglu, Alam]

### 13.6 Time Exponent of Si-H Bond Dissociation

The interface-trap generation rate for HCI can be written as a balance between dissociation and repassivation rates of Si-H bonds.

$$\frac{d N_{IT}}{dt} = k_F(N_0 - N_{IT}) - k_R N_H(0) N_{IT}$$  \hspace{1cm} (13.2)

Generally $N_{IT}$ is small compared to $N_O$ and derivative of $N_{IT}$ with respect to time is small compared to the fluxes on the right hand side, as a result the above expression can be written as:
The diffusion of hydrogen, HCl degradation, diffuses in a conical fashion since it takes place only near the drain-end. Volume under the curve:

\[ \frac{k_F N_0}{k_R} = N_H(0) N_{IT} \]  \hspace{1cm} 13.3

Fig. 13.11: (a) Diffusion of hydrogen in a conical fashion, (b) 3D diffusion profile for hot carrier degradation (c) 2D diffusion profile for hot carrier degradation, at \( t_2 = t \), diffusion peak also increases.

\[ N_{IT} = \frac{1}{3} \times \pi \times (\sqrt{D_H t})^2 \times N_H^{(0)} \]  \hspace{1cm} 13.4

Therefore,

\[ N_{IT} = \sqrt{\frac{1}{3} \times \pi \times \frac{k_F \times N_0}{k_r} \times D_H \times t} \]  \hspace{1cm} 13.5

As can be seen from the equation above, the diffusion of hydrogen atom (H) in the case of hot carrier degradation has time dependence with an exponent of (1/2).
13.7 **Time Exponent of Si-O Bond Dissociation**

As mentioned earlier, Si-O bonds that are broken primarily by ‘hot holes’ cannot be repassivated like the Si-H bonds once they break off. It is primarily because the Si-O bonds form at a very high temperature (>1000ºC) and thus once the bond disintegrates it is impossible to repassivate them in a normal MOSFET. The oxide layer contains amorphous (SiO$_2$) and they have various ring sizes within the oxide layer. Therefore, breaking off these bonds is associated with energies that are different for different ring sizes (the lower the energy required to break off a ring, the higher the number of defects generated).

Since there is no repassivation to worry about ($k_r = 0$), the number of Si-O bonds broken can be written as:

$$\frac{dN_{IT}(t)}{dt} = k_f (N_o - N_{IT}(t))$$  \hspace{1cm} 13.6

The solution of the differential equation is: $N_{ITi} = N_o \left[ 1 - e^{-k_f(t)} \right]$, where each ring size (i) will have different values of ($k_{fi}$) and therefore the number of defects generated will be ($N_{ITi}$).

$$\frac{dN_{IT}}{dt} = \int_{E_0 - n\sigma}^{E_0 + n\sigma} k_f(E) \left[ g(E) - f(E, t) \right] dE$$  \hspace{1cm} 13.7

Also, $\int g(E_A) dE_A \equiv N_o$. Therefore,

$$\frac{dN_{IT}(t)}{dt} = \int_{E_0 - n\sigma}^{E_0 + n\sigma} k_f(E_A) g(E_A) dE_A$$

$$\approx \int_0^{\infty} k_f(E_A) g(E_A) dE_A$$  \hspace{1cm} 13.8

where,
\[ g(E_A, E) = \frac{1}{\sigma} \frac{e^{\frac{(E_A - E)}{\sigma}}}{\left[ e^{\frac{(E_A - E)}{\sigma}} + 1 \right]^2} \] and \( k_f = k_o \times e^{\frac{E_A}{k_B T}} \) \hspace{1.5in} 13.9

Therefore,

\[ \frac{N_{IT}}{N_o} = 1 - \frac{1}{1 + \left(\frac{t}{T}\right)^n}, \text{where} \ n = \left(\frac{k_B T_e}{\sigma}\right)^p \] \hspace{1.5in} 13.10

Fig. 13.12: (right) If there was only one type of Si-O ring type, then the (HCI) degradation due to breaking of Si-O plot would like the figure on the (right). Figure 13: (right) Different types of Si-O ring sizes require different energies to break the bonds. (left) Different energies required to break Si-O bonds in different ring sizes lead to different degradation vs. time plots, but their summation yields the universal (red)HCI degradation curve.
13.8 Conclusion

In this lecture, the degradation of a MOSFET due to hot carrier injection is discussed. The physical mechanism of the degradation process has been discussed in detail. The degradation is caused mainly due to breaking of Si-O and Si-H bonds due to impact ionization of highly energized charges that generates interface traps that results in \( (V_{Th}) \) shift and ultimately deteriorates the performance of the device. The universality of the degradation process is also discussed along with the time exponents of Si-H and Si-O bond dissociation.

References