Gate Oxide Breakdown

Navid Azizi and Peter Yiannacouras
Outline

- Motivation
- Background
- Root Causes for Gate Oxide Breakdown
- Symptoms of Gate Oxide Breakdown
- Failure Models
- Prediction of Gate Oxide Breakdown
- Protection Against Gate Oxide Breakdown
- Conclusion
Motivation

- As technology is scaling, $t_{ox}$ is getting thinner

Why?

- To reduce power, $V_{DD}$ is lowered
  - To maintain performance
  - To control short channel effects
    - Gate Oxide must be made thinner

- With scaling, Gate Oxide Reliability becomes an issue
  - Electric Fields within the Gate Oxide grow larger with scaling
  - More and more transistors on chip
Background
Transistor Structure

Gate Oxide Breakdown
Gate Oxide Traps

- Defects in the Gate Oxide are called Traps
  - They can trap charges

- Traps are usually neutral except for
  - Near the anode they quickly become negatively charged
  - Near the cathode they quickly become positively charged
Root Causes
What is Gate Oxide Breakdown?

- Breakdown is defined as the time when there is a conduction path from the anode to the cathode through the gate oxide

- Traps allow for creation of conduction path

Outline of this section
- First we will see how traps lead to conduction paths
- Then we will investigate different physical methods for the creation of traps
  - The mathematics for these different physical models will be dealt later
Traps within Gate Oxide

- Traps start to form in the Gate Oxide
  - originally
    - Non-overlapping
    - Do not conduct
Soft Breakdown

- As more and more traps are created
  - Traps start to overlap
  - Conduction Path is created

- Once this conduction path is created we have **Soft Breakdown (SBD)**
Thermal Damage

- Conduction leads to heat
- Heat leads to thermal damage
- Thermal Damage leads to Traps
- More Traps leads to more conduction

Gate: PolySilicon

SiO$_2$

Substrate
Hard Breakdown

- Silicon in the breakdown spots melts
- Oxygen is released
- Silicon Filament is formed from Gate to Substrate (Hard Breakdown)
Photographs of Gate Oxide Breakdown

- Breakdown region pictured through emission microscopy
  - Photon emission at breakdown regions

- Dark region indicates area where Silicon has melted


Photographs of Gate Oxide Breakdown

- TEM Image of Breakdown Spot
- Substrate below Gate Oxide Breakdown

Trap Generation

- Know how traps can cause Gate Oxide Breakdown
- How are traps created?
- Different Models (i.e. we’re not exactly sure how)
  - Thermochemical Model
  - Anode Hole Injection
  - Hydrogen Release
  - Channel Hot Carriers
  - Irradiation

- Discuss the Physical Reasons
  - Math that leads to reliability projections for above models will be presented later
Thermochemical

- Model shows good agreement with data at low Electric Fields

- Structure of SiO$_2$

- Bond Angle between O-Si-O is always 109°

- Bond angle between Si-O-Si ranges from 120° to 180°
  - Bond is severely weakened above 150°
  - Can lead to bond breakage
Thermochemical - cont

- After bond breakage
  - Oxygen Vacancy

- Important Facts about this new structure
  - Si-Si is a very weak bond
  - Si-O bond is highly polar
Thermochemical - cont

- Go over polarization of polar molecules within an electric field
  - Polar molecules have a default polarization
  - In the presence of an electric field polarization can change
Thermochemical - cont

- When the Electric Field is applied to the oxide
  - The highly polar Si-O bonds within the oxide become polarized
  - The lattice becomes distorted
  - Each molecule of SiO₂ not only feels $E_{\text{ox}}$ but $E_{\text{loc}}$
  - Si-Si bonds become strained and break
Thermochemical – cont

- After the Si-Si bond breaks
  - The remaining electrons cause a hole trap
Anode Hole Injection

- Model shows good agreement with data at high Electric Fields

- High Electric Fields
  - Large tunneling current (electrons) through the oxide
  - Electrons have high Kinetic Energy
  - Electron hits the Gate Anode and transfers energy to Hole
  - Hole tunnels back into the Gate Oxide
  - Hole creates trap
Anode Hole Injection

- How do holes create Traps?
  - Holes break Si-O bonds
  - Two bond breakage near a Si atom can cause a permanent trap

Hydrogen Release Model

- Very similar to Anode Hole Injection Model
  - The AHI rate is too small to produce the defects that lead to breakdown
  - Use Hydrogen instead of Holes to produce traps

- Just as in AHI high energy electrons tunnel through oxide
  - Break Si-H bond at interface of gate oxide
  - $H^+$ ion (proton) is released into the oxide
  - Proton reacts with oxygen vacancies to produce traps
    - $(\text{Si-Si})+H^+ \rightarrow \text{Si-H}^+ -\text{Si}$
Channel Hot Carriers

- Thermochemical, AHI and HR models can all explain gate oxide breakdown when there is no potential difference between drain and source
  - There is data, however, that shows that gate oxide breakdown is more likely when there is a potential difference between drain and source

- Hot Carriers
  - Electrons and Holes who, in the presence of high lateral fields, gain sufficient energy that they are no longer in equilibrium with the lattice

- The hot carriers create an electron-hole pair by impact ionization in the channel
  - Hole enters the substrate
  - Electron enters the gate oxide and may cause traps
Irradiation

- Irradiation with ions can lead to oxide defects
- Irradiation has no immediate impact by itself, the transistor works as it should
- But transistors that have been irradiated, and then stressed break down more quickly
- Exact nature of defects caused due to irradiation in gate oxide is unknown
Symptoms
Symptoms of Breakdown

- **Transistor Characteristics**
  - Hard Breakdown
  - Soft Breakdown

- **Circuit Characteristics**
  - Inverter
  - Digital Logic
  - SRAMs
  - RF Circuitry
Hard Breakdown

- Current path exists from the Gate to the Channel
  - Large increase in gate current
    - ~2 orders of magnitude larger than normal when the transistor is on
    - ~6 orders of magnitude larger than normal when the transistor is off

- Current path is characterized by a breakdown resistance $R_G = V_G/I_G$
  - $R_G$ depends on breakdown locations
    - Increases linearly over drain and source regions due to the length of the drain and source extensions

---

Hard Breakdown Continued

- Breakdown over channel can be modeled with this circuit
  - $V_G > 0$
    - Current is injected from the gate to the channel where it goes to the drain and source
  - $V_G < 0$
    - Electrons are injected from the gate through the breakdown path
    - Diffuse through the substrate
    - Collect at source and drain

Soft Breakdown

- Not much change in transistor characteristic
  - Increased off state leakage current

- Transistor On-state
  - Increased gate leakage
  - With technologies with thin $t_{ox}$’s additional gate current may be large compared to intrinsic gate tunneling leakage

- Transistor Off-State
  - If breakdown occurs near drain, increase in GIDL of 5 orders of magnitude
  - Due to negative charge trappings in the oxide over the overlap region

- Transistors with low W/L
  - Breakdown region may form considerable portion of gate
  - Transconductance will drops of 50%
  - Saturation current of 30%
Inverter Characteristics

- Inverter stressed with positive voltages
  - NMOS is damaged
  - Ground current is increased
- Inverter stressed with negative voltages
  - PMOS is damaged
  - VDD current is increased
- Positive or Negative Stresses
  - Input current follows output current

Inverter Characteristics - cont

- DC Transfer Characteristic
  - Positive Stress
    - Can’t produce a good ‘0’
  - Negative Stress
    - Can’t produce a good ‘1’

Digital Logic

- Breakdown doesn’t cause that large a problem
  - Node B has extra loading and can’t be pulled completely high
  - Node C can’t be pulled completely low
  - But, next logic stages will clean the signal up
Digital Logic - cont

- Ring Oscillator Example
  - Functions, albeit at lower frequency, after many breakdowns
  - Increase in leakage current

SRAM

- Breakdown can occur in 3 different places
  - Drain
  - P-source
  - N-source
SRAM - cont

- Static Noise Margin (stability) of SRAMs decreases with breakdown
  - N-source and p-source breakdowns induce an asymmetry in the butterfly curve reducing the SNM
    - P-source breakdown is not so bad, because the NMOS is strong enough to combat the effects
  - N-source breakdown results in decreased SNM because the PMOS is weak and cannot deliver enough current to combat the extra leakage source
  - Drain breakdowns reduce the output swing of the SRAM, reducing it’s SNM

**RF Circuitry**

- Analog circuit designed with many specifications
  - A single breakdown can cause the circuit to stop functioning at its operating point

- Transistors are usually operated in saturation
  - Increased hot-carriers

- Oxide breakdown in a LNA led to
  - 5dB (3x) decrease in gain
  - Noise Figure increased from 2dB to 3dB
  - Frequency of minimum reflection shifted by 600MHz and at operating point has changed from -27dB to -9dB (increase of 62x)

---

Failure Models
Goals of our Failure Model

- Mapping from device parameters
  - Area, thickness, activation energies, etc.

- and usage conditions
  - Field, current, temperature, etc.

- to breakdown occurrence
  - Time ($t_{bd}$) or charge ($Q_{bd}$)
What do we know?

- Device characteristics, usage conditions
- Trap generation
- Breakdown

Breakdown Projection Models

Percolation Models
**Percolation Models**

- Given that trap generation occurs with some probability, what is the probability that a breakdown occurs?

**Tile-based**
- Developed by Sune in 1990
- Models gate oxide as a plane made up of small tiles
- Traps occur randomly in any tile
- After certain number of traps occurred in a tile, tile breaks down

\[
\ln(-\ln(1 - F)) = \ln\left(\frac{A}{a}\right) + \ln\left(ab_0p - \ln\left(\sum_{n=0}^{n_{bd}-1} \frac{(at_{ox}p)^n}{n!}\right)\right)
\]

- Is a Weibull distribution as expected
- Lacks predictive power in failing to relate \(n_{bd}\) to \(t_{ox}\)
Percolation Models

Sphere-based

- Degraeve, 1995
- Used 3-D model where traps were represented as spheres
- Only parameter is sphere radius
- Monte Carlo Simulations yielded distributions that were Weibull and similar to what was seen in practice.
- Related Weibull slope $\beta$ to $t_{ox}$
- Accounted for non-uniform thickness

Success!

- Is the currently accepted percolation model
- Comparing to experimental data, sphere diameter $= 0.9$ nm
- Suggests that thickness less than this will die quickly (on the first trap)
Percolation Models

**Cube-based**
- Stathis, 1999
- Used 3-D model made up of cubes
- Only parameter is cube size
- Monte Carlo Simulations yielded distributions that were weibull and similar to what was seen in practice.
- Related Weibull slope $\beta$ to $t_{\text{ox}}$

- Comparing to experimental data, cube size = 2.7 nm
- Size too big since oxides thinner than that were seen to be working reliably
Percolation Models

Analytical Cube-based
- Sune, 2001
- Similar as previous
- Expressed analytically as follows:

- Say $\lambda$ is the probability of a cube becoming a trap
- The reliability of the gate oxide is:
  \[ R_{bd} = [1-F_{col}(\lambda)]^N = [1 - \lambda^n]^N \]
  and the Weibull is
  \[ W_{bd} = \ln[-\ln(1-F_{bd})] = \ln[-N\ln(1 - \lambda^n)] \]
  Or, since $\lambda << 1$,

  \[ W_{bd} = \ln(N) + n\cdot\ln(\lambda) \]

- Clearly a Weibull, only parameter is cube size

But what is $\lambda$?
- This is where percolation models stop
- But for comparison:

  \[ \lambda = \lambda_0 Q^\alpha \]

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$a_0$</th>
<th>$a_{here}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere</td>
<td>0.56</td>
<td>0.9nm</td>
<td>1.17nm</td>
</tr>
<tr>
<td>Cube</td>
<td>1</td>
<td>2.7nm</td>
<td>2.34nm</td>
</tr>
</tbody>
</table>
Percolation Models and Soft/Hard Breakdown

- Percolation models make no distinction between hard and soft breakdown.

- This supports notion of soft and hard breakdown stemming from same cause and the only difference being after-effects (thermal runaway).

- This hypothesis was verified by Sune by showing distributions of first soft breakdown coinciding with distributions of first hard breakdown.
Breakdown Projection Models

- Breakdown projection models are closest thing we have to failure models
- Goal is to predict time to failure
- Very controversial – physicists working on percolation models criticize all research in this area.
- Often incomplete/unusable
- Two different families of breakdown projection models: E and 1/E
E and $1/E$ models

- Empirical discrepancy between breakdown at low and high fields

**E Model**

\[
\ln(t_{bd}) \propto \frac{\Delta H_0}{k_B T} - \gamma E_{ox}
\]

**1/E Model**

\[
\ln(t_{bd}) \propto \frac{\Delta H_0}{k_B T} - G(1/E_{ox})
\]
Breakdown Projection Model Candidates

- There are several models
  - Traditionally sided with either E or 1/E model, not both
  - Many not developed enough to be used yet
  - Some were absorbed by better models

- The two main candidates:
  1. AHI model (1/E model)
  2. Thermochemical model (E model)

- Both are currently attempting to unify the E and 1/E model
AHI Model

- Proclaimed cause of breakdown: Electrons
  - Tunneling electrons dissipate energy creating holes

- Model is as follows:

  \[ t_{bd} = \frac{Q_{bd}}{J_n} \]

  \[ Q_{bd} = \frac{Q_p}{a_p} \exp\left(\frac{B}{E} \Phi_p^{3/2}\right) \]

- Note the 1/E dependence
- Once calibrated agrees very well with data from high fields
- Can explain switch to E physics by accounting for minority ionization (very nasty equation)

- J_n – gate current density (quantum)
- Q_p – critical hole fluence at breakdown = 0.1 C/cm^2 (approx)
- B, \( \Phi_p \) are also known
- \( a_p \) – is probability a tunneling electron causes a trap. This value is unknown and cannot be calculated. Must use curve fitting to calibrate equation!!!
Thermochemical Model

- Proclaimed cause of breakdown: Electric Field
  - No place for electrons or holes

- Model is:

\[
t_{bd} = A_0 \exp \left( \frac{(\Delta H)_0 - 7.2e \cdot A \cdot E_{ox}}{k_B T} \right)
\]

- \(\Delta H_0\) – enthalpy of activation for trap generation (known)
- \(k_B\) – Boltzmann’s constant
- \(T\) – temperature
- \(A_0\) and \(A\) are known parameters

- Note the E dependence
- Purely quantitative
- Weaker agreement with data, does okay with low fields
Thermochemical Model - Enhanced

- Can account for E and 1/E effects by considering simultaneous reactions

\[ k_{2a} \quad k_{2b} \text{ – reaction rate increased} \]

\[ k_1 \]

\[ k_{eff} = k_1 + \frac{k_{2a}k_{2b}(\exp(-k_{2a}t) - \exp(-k_{2b}t))}{k_{2b}\exp(-k_{2a}t) - k_{2a}\exp(-k_{2b}t)} \]

\[ t_{bd} = \frac{\ln\left(\frac{1}{f_{crit}}\right)}{k_{eff}} \]

- \( k_{2a} \) insignificant
- \( k_{2a} \) dominates
- E dependence
- 1/E dependence
Prediction
Prediction

- Choice of model still controversial, still being researched

- Can not be done at the design phase
  - TC model can, but agreement is for small range of E and even then questionable
  - AHI model requires experiment and curve fitting

- Further complicated
  - Breakdowns may not cause failures
  - Coupling between parameters
    - Model & field
    - Physical constants & device geometry

- Prediction done through trial & error using accelerated testing
  1. Apply CVS/CCS for elevated Temperature and Voltage
  2. Extrapolate for Temperature, Voltage, and device Area
  3. Using Weibull, extrapolate for all failure rates
Prediction

- Initial steps to automating
- Plugin to BERT (Berkeley Reliability Tool)
- Simulates breakdown producing failure rates once given:
  - Usage environment (current)
  - Failure time ($t_{bd}$)
  - An experimentally determined mapping from gate thickness to the density of defects which span the thickness
- Not applicable to design phase prediction, hope is feedback will induce “design for reliability”

Protection
Protection against Gate Oxide Breakdown

- We have seen that breakdown depends on the
  - Electric Field (Thermochemical and AHI models)
  - Hot Carriers

- What can we do to reduce the probability of Breakdown
  - Guarantee that the oxide doesn’t experience Electric Fields larger than it was designed for (Voltage across gate should not be larger than VDD)
  - Minimize the current through a transistor when it is in saturation
Bitline Reduction Scheme

- Leakage in SRAMs is becoming important for both power and performance concerns
  - Underdrive the pass-transistors by 100mV when the cell is not active to lower bitline leakage
    - But now the voltage across the gate is $V_{DD} + 100\text{mV}$
    - Reduce cell voltage to $V_{DD} - 100\text{mV}$
      - But now we lose some SRAM stability
        - Increase the cell area

- Tradeoff area for reliability

---

**PRCH**

**WL**

**BL**

100mV
RF Circuitry Protection

- We’ve seen before that RF circuitry is very sensitive to gate oxide breakdown

- \( g_m \) stage
  - Hot Carriers are exponentially related to \( V_{ds} - V_{dsat} \)
  - \( V_{ds} \) is 550mv, \( V_{dsat} \) is 200mV

- Current Switching Stage
  - In Saturation
  - Carry lots of current
RF Circuitry Protection

- $g_m$ stage
  - Add an extra transistor to lower $V_{ds}$ across transistors

- Current Switching Stage
  - Add extra PMOS current source to remove common-mode current from current switching transistors
Conclusion

- Gate-oxide breakdown caused by trap generation

- Trap Generation Models
  - AHI
  - Thermochemical
  - No unified model

- Predicting gate-oxide breakdown is difficult

- To protect against gate-oxide breakdown
  - Voltage across gate-oxide should not be larger than VDD
  - Reduce hot-carriers