Device Reliability Physics: Past, Present, and Future

Jim Stathis
IBM
Yorktown Heights, NY
Acknowledgements

- Barry Linder
- Miaomiao Wang
- Ernest Wu
- Emmanuel Yashchin
- Eduard Cartier
- Fei Xue (summer intern, U. Texas Austin)

This work was performed by the Research Alliance Teams at various IBM Research and Development Facilities
Introduction

- **What is Reliability?**
  - The *probability* of operating a product for a given *time* under specified *conditions* without *failure*

- **Why Reliability matters?**
  - To evaluate if the *product performance* meets *expectation*
  - *Tradeoffs* between reliability, performance and cost

**Lifetime projection**

- Temperature Scaling (T)
- Area Scaling (A)
- Percentile Scaling (P)
- Voltage Scaling (V)

*Usage conditions*

E. Yashchin, IRPS (2012)
Implications of reliability

- From the *atomic scale*
- To the *system level*
• Big Data
  - Most advanced, dense technology
  - 3D chip stack, cooling, optical bus….
  - System level reliability
  - Voltage and power limits

• Disposable electronics
  - Medical devices
  - Shelf life

• The Internet of Things
  - 10s of billions of connected devices
  - Sensors, phones, tablets, cars, houses, utilities…
  - Diverse operating environments and lifetimes

Source: http://www.iontorrent.com/
device technology trends

Performance and complexity

- Stressors
- Oxide scaling limit
- Metal gate high-k
- Finfet
- 3D integration
- Nanowire
- Graphene

Time
Moore Marches On to the Dennard Drumbeat

- 1970
  - Transistors per chip: 2,000

- 1980
  - Transistors per chip: 60,000

- 2007
  - Transistors per chip: 800,000,000

- 2012
  - Transistors per chip: >2,000,000,000

- Beyond 2012
  - 20,000,000,000 transistors in one chip-sized 3D package

- Tox-scaling for short channel control

- Voltage, dimensions and doping are scaled by same factor
  - $\uparrow$ Density: $\sim \alpha^2$
  - $\uparrow$ Speed: $\sim \alpha$
  - Power Density: $\sim$ Const.
Oxide thickness trend

Data from IEDM and VLSI Technology conferences

- Oxide/oxynitride
- High-κ (EOT)

ITRS 2011
Device Scaling for DENSITY

- Aggressive gate dielectric ($T_{inv}$) scaling for improved short channel control
- W&L and Pitch scaling
  - Increased random doping fluctuations
  - Spacer thickness becoming comparable to older gate dielectric thickness (~10nm)

![Diagram of device structure](image)

### Node vs Device Pitch (nm)

<table>
<thead>
<tr>
<th>Node</th>
<th>Device Pitch (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>170-180</td>
</tr>
<tr>
<td>32</td>
<td>120-130</td>
</tr>
<tr>
<td>22</td>
<td>80-100</td>
</tr>
<tr>
<td>15</td>
<td>65-80</td>
</tr>
<tr>
<td>11</td>
<td>50-65</td>
</tr>
</tbody>
</table>
evolution toward fully depleted devices
Implications for reliability

- Atomic scale fin geometry
- Threshold voltage tuning by metal gate work function
  - Oxide field modulation
  - Physics of oxide breakdown

Effect of Depletion Charge on Oxide Field

- Many researchers use the following expression to approximate the vertical field in the interfacial region in MOS devices in inversion:

\[ E_{ox} = \frac{(V_g - V_t)}{T_{inv}} \]

- This expression is not exact
  (It is accurate to ~15% for \( (V_g - V_t) > 1V \))

- Correct expression is \( E_{ox} = \frac{(V_g - V_t + V_{dep})}{T_{inv}} \)
  where \( V_{dep} = \frac{(\text{depletion layer charge})}{C} \)

  \[ V_{dep} \approx \frac{q \cdot N_a \cdot D_{f_{in}}}{C_{ox}} \]

  \( V_{dep} \approx 0.16V \) for \( 1e18 \) doping for \( 1nm \) \( T_{inv} \)
  and \( V_{dep} \approx q \cdot N_a \cdot D_{f_{in}} / C_{ox} \) \(~0\) for undoped fin

- (Hence, \( E_{ox} \approx \frac{V_g}{T_{inv}} \) could be a better expression, and indeed this works well for standard N+ Poly CMOS)
• Classical expression for threshold voltage is \( V_t = \Phi_{ms} + 2 \cdot \psi_f + V_{dep} \)

Where \( V_t \) is defined as band bending = \( 2 \psi_f \)

• However, operationally we define \( V_t \) by constant inversion charge (i.e. fixed \( I_d/W \)):

\[
V_t = \Phi_{ms} + (\psi_f + \psi_i) + V_{dep} \\
= \Phi_{mi} + \psi_i + V_{dep}
\]

Where \( \psi_i \) is defined as the band bending for fixed \( I_d/W \)

\textit{Note} only the last term contains doping dependence

• Thus \( E_{ox} = (V_g - V_t + V_{dep})/T_{inv} = (V_g - (\Phi_{mi} + \psi_i))/T_{inv} \)

\textit{is independent of doping, and depends only on metal work function}
Experimental evidence:
Oxide field is independent of well doping

Gate current in strong inversion is independent of doping

Capacitance in strong inversion is independent of doping
Experimental evidence:
Oxide field dependence on metal work function

Gate current in strong inversion is dependent on $V_g - V_t$ (from Metal work function)

Capacitance in strong inversion is dependent on $V_g - V_t$ (from Metal work function)
Field normalization: Take-away message

- **To compare gate stack intrinsic reliability:**
  - **DO NOT** adjust for Vt when comparing devices with different Vt from substrate side (well doping, halos, etc)
  - **DO** adjust for Vt when comparing devices with different gate work function
n-FinETs (undoped) vs. bulk n-FETs ($N_a \sim 1 \times 10^{17} \text{cm}^{-3}$) – PBTI

Equivalent lifetimes at Equivalent field

$E_{ox} = (V_g - V_t + V_{dep}) / T_{inv}$

Lifetime extracted @ $E_{ox} = 7 \text{MV/cm}$

Superior lifetime for FinFET at matched $V_g$ and $V_t$
Other evidence

- At constant overdrive (Vg-Vt), increasing doping appears to make NBTI worse
- 22nm FinFET vs 32nm Planar (nFET improved by ~100mV)

Chadwin Young, WODiM (2012)

Intel, VLSI-T (2012)
Oxide Breakdown: Physical models for SiON/Poly-Si CMOS

- **Thermochemical (field-driven) ("E")**
  - Field driven defect generation
  - Bond-breaking energy lowered by dipole term

- **Hydrogen release ("Vg" or power law)**
  - Si-H bond breaking followed by secondary reactions

- **Anode hole injection ("1/E")**
  - Hot holes generated in anode electrode
  - Constant critical hole fluence \( Q_p \sim 0.1 \text{ C/cm}^2 \)

---


Do any of these models apply for High-k / Metal-Gate?
Percolation model

- Widely accepted, common to all physical models
- Defects generated randomly in oxide until a conducting ("percolating") path is formed
- Explains thickness-dependent Weibull slope in SiON/Poly-Si
- Has not been validated for HKMG (limited thickness range available!)


- Random defect generation
- Connection from one electrode toward the other
Progressive Breakdown: Growth of breakdown spot

- Generation of additional defects (traps) near the localized BD spot
- Increasing diameter of BD spot → increasing conductance

B.P. Linder et al., IRPS 2003, p. 402.
Why Progressive Breakdown is Important: Failure Distribution

\[
f(t) = \int_{0}^{\infty} f_{BD}(t - \Delta t_{PBD}) f_{PBD}(\Delta t_{PBD}) d\Delta t_{PBD}
\]

- Convolution of breakdown time distribution and post-breakdown growth time distribution

- Curvature on Weibull plot

- Huge gain in reliability margin

Progressive BD component in metal gate devices

- <1µA progressive BD before catastrophic BD
- Current increase immediately from 10uA to 100uA at breakdown
- Small progressive BD component: Gradual increase in current and increase in noise observed before breakdown

![Graph showing sense current vs. total time](image)

- **Sense Current @1V (A)**
  - Total time (s)
  - **V_s=2.5V**
  - **V_s=2.3V**
  - **V_s=2.2V**
  - **V_s=2.4V**

![Graph showing gate current vs. time](image)

- **Gate Current [A]**
  - **Catastrophic BD**
  - **Charge trapping**
  - **Slow growth phase**
  - **No change in Vt trend vs. time**

- N-MOSFET
- 2x10^-8cm^2
- V_G= +2.6V

S. Lombardo, et al., Sematech, ISAGST’06 (Austin).
Breakdown statistics in SiON/Poly-Si devices

- Classical Weibull distribution of intrinsic 1st breakdown in Poly-SiON

- Progressive BD distribution (I_fail criterion) in Poly-SiON

- Monte-carlo simulations


Breakdown statistics in High-k/Metal Gate devices (or, Goodbye Weibull ?)

- Curvature in Weibull plot for HKMG “1st BD”
  - Increasing the failure criterion, increases VBD for low percentiles.
    - signature of two-stage breakdown mechanism.

\[
\ln (-\ln (1-F)) = \ln (A/1\mu m^2)
\]
Various Explanations for non-Weibull distribution in HKMG

- **Dual layer percolation model**
  - Different defect generation rates for interface and high-k layer
  - Uncorrelated breakdown paths generated randomly on the high-k and interfacial layer area

- **Correlated breakdown**
  - A breakdown path initiated in one layer induces local increase of defect generation rate in the other layer.
  - The mathematics of this mechanism is very similar to progressive breakdown.

- **Progressive breakdown**
  e.g., S. Sahhaf et al., Trans. Elect. Dev. 56, p. 1424 (2009).
General Aspects of the Statistical Nature of Breakdown

- **“Weakest link”:**
  - The weakest link in a chain controls the failure of the whole chain
  - If any one spot on a dielectric breaks, the entire device is broken
  - If any transistor fails, the whole circuit or chip fails

- The Weibull Distribution is an ‘extreme value’ distribution in $\ln(x)$ and is appropriate for a “weakest-link” type of problem, however, it is not the only valid distribution.

- General formulation for minimum failure time:

  $$F_{\text{min}}(t) = 1 - [1 - F(t)]^N$$

- Assume $N$ independent competing “cells”, each described by an (arbitrary but identical) distribution of failure times.

- Cell with the shortest lifetime causes failure of the sample
  - Note: Here we are not considering successive (multiple) breakdown, which also introduces curvature in Weibull scale – see e.g. E. Wu, IRPS 2012

- The Weibull distribution is a special case: if $F$ is Weibull, $F_{\text{min}}$ is also Weibull. (This is the usual area scaling relation.)
One case:
Min-Log distribution for progressive and metal-gate breakdown

- Assume $N$ independent competing “cells”, each described by a log-normal distribution of failure times. \( F_{min}(t) = 1 - [1 - F(t)]^N \)

- Excellent & statistically rigorous fits to high-k / metal gate breakdown data
  - Cell size (or $N$) is a parameter

5-parameter model: S. Tous et al., IEEE EDL, 29, p.949 (2008); IRPS 2010; J. Suñé et al., IEDM 2010
min-log model: E. Yashchin et al., IRPS 2012
Unique feature of HKMG breakdown: AC effect in NFET

- **HKMG**: interrupted quasi-DC stress not equivalent to pure DC stress
  - At high percentiles, interrupted stress has shorter $T_{FAIL}$
  - Distributions may merge or cross at low percentiles
    - Complicated relation depending on all stress parameters and percentiles

- **SiON-Poly**: unipolar AC stress nearly equivalent to DC
  - Slight increase in TBD with frequency for bipolar AC
  - Basis for common usage of interrupted DC stress for SILC

E. Wu, IEDM 2011 presentation
Voltage Dependence of AC breakdown in high-k nFET

- AC stress has lowest T_FAIL at low voltage

Relation to charge trapping:
- Universal voltage acceleration is obtained when charge trapping is taken into account

Possible relation to defect recovery in RRAM?
- Relation to product usage conditions?
Final Thoughts

- SiO\textsubscript{2} reliability was studied intensively nearly 50 years.
  - Continuous scaling in thickness from ~1 micron to ~1 nanometer motivated long term fundamental research and development.
  - In spite of this, charge trapping, interface state generation, and breakdown are still somewhat mysterious.

- New devices and materials give us new physics to think about, and challenge our established models
  - Fully depleted devices, high-k, metal gate…III-V…not even thinking about graphene yet!
  - To what extent can our understanding of SiO\textsubscript{2}/poly reliability be transferred to these new devices and materials?
  - Because of the accelerated pace and increasing complexity of technology development, physical models are far from complete for HKMG gate stacks,

- The critical importance of reliability requires continued fundamental research and understanding of wearout and failure mechanisms to support reliability qualifications for present and future technologies