Comparison of Strategies for Redundancy to improve Reliability concerning Gate Oxide Breakdown

Hagen Sämrow, Claas Cornelius, Frank Sill, Andreas Tockhorn, Dirk Timmermann

17.02.2009, Bremen
Outline

- Basics and Motivation
  - Approaches for reliability enhancements
  - Gate oxide breakdown
- Redundancy strategies
  - Theoretical fundamentals
  - Results
- Conclusion / Outlook
Motivation – Known approaches

Reliability

Transient failures

Soft error resilience
- Hardening techniques
- Reusing debug resources for redundant flipflops [Mitra]

Permanent failures

Yield enhancements
- Layout modifications
- Redundancy

Initial failures

Failures occurring at runtime

Little effort put into lifetime reliability enhancements

Initial failures

Permanent failures

Transient failures
Motivation – Known approaches

Lifetime reliability enhancements

High level:
- Dynamic system management to adapt operation conditions in response to an observed hardware usage [Srinivasan et al.]

Low level:
- Random insertion of redundant transistors to improve yield [Sirisantana et al.]
- Improvement of this approach by a controlled insertion at those instances which are most vulnerable to gate oxide breakdown [Sill et al.]
Basics – Gate oxide breakdown

Gate oxide breakdown – GOB:

- Point of time a conducting path between gate and substrate is generated

- Mainly dependent on:
  - Gate oxide thickness
  - Electrical field at the gate

- Causes:
  - Sudden extrinsic overvoltage: ESD – Electro-Static Discharge
  - Slow intrinsic destruction over time: TDDB – Time-Dependent Dielectric Breakdown
Basics – TDDB

- Physical mechanism: trap creation
Basics – TDDB

Initial traps

Poly Silicon

SiO$_2$

Substrate
Basics – TDDB

Initial traps

During operation: generation of overlapping traps

Poly Silicon

SiO₂

Substrate
Basics – TDDB

Initial traps

During operation: generation of overlapping traps

Soft breakdown: Creation of a conducting patch
Basics – TDDB

- **Initial traps**
- **During operation:** generation of overlapping traps
- **Soft breakdown:** Creation of a conducting patch
- **Heat dissipation**
- **Increasing current flow:**
  - Heat dissipation
  - Thermal damage
Basics – TDDDB

Initial traps

During operation: generation of overlapping traps

Soft breakdown:
Creation of a conducting patch

Increasing current flow
→ Heat dissipation
→ Thermal damage
Basics – TDDDB

- Initial traps
- Finally: Hard breakdown

Diagram showing a cross-section of a device with Poly Silicon, SiO₂, and Substrate layers, with a path marked by R → 0.
Basics – TDDDB

Finally: **Hard breakdown**

Model by Segura et al.
Basics – Scaling issues

- Scaling increases the gate oxide breakdown problems:
  - Increasing number of transistors within a die
Basics – Scaling issues

Scaling increases the gate oxide breakdown problems:
- Increasing number of transistors within a die
- Decreasing gate oxide thickness
- Increase of the electrical field due to non-ideal supply voltage scaling

\[ l_1 > l_2 \]
\[ t_{ox1} > t_{ox2} \]
\[ E_1 < E_2 \]
Theoretical fundamentals

- Failure rate $\lambda_{SYS}$ represents the rate at which an individual system suffers from individual faults.
- Reliability $R_{sys}(t)$ is the probability of the system to perform as desired at time $t$:
  \[ R_{SYS}(t) = e^{-\lambda_{SYS}t} \]
- Series system with $n$ equal components (component failure rate $\lambda$) fails if any component fails:
  \[ R_S(t) = [R(t)]^n = \left[e^{-\lambda t}\right]^n \]
- Parallel system works until all of its components fail:
  \[ R_P(t) = 1 - \left[1 - R(t)\right]^n = 1 - \left[1 - e^{-\lambda t}\right]^n \]
Mean Time To Failure ($MTTF_{SYS}$) of a system is the average time a system operates until it fails.

\[
MTTF_{SYS} = \int_{0}^{\infty} R_{SYS}(t) \, dt
\]
Redundancy strategies

\( n = \text{no. of transistors}; \ g = \text{no. of gates}; \ m = \emptyset \text{ no. of transistors per gate} \)

- **Basic multiplier**
  \[
  R_{BASIC}(t) = e^{-n\lambda t}
  \]

- **Block duplication**
  \[
  R_{BD}(t) = 1 - \left(1 - e^{-n\lambda t}\right)^2
  \]

- **Gate duplication**
  \[
  R_{GD}(t) = \left[1 - \left(1 - e^{-m\lambda t}\right)^2\right]^g
  \]

- **Transistor duplication**
  \[
  R_{TD}(t) = \left[1 - \left(1 - e^{-\lambda t}\right)^2\right]^n
  \]
Simulation setup

- Wallace multiplier
- Transistor level simulations with HSpice
- Industrial 65 nm gate library
- First order model of Segura et al.
Results – No defects
Results – Reliability with defects

Simulation results and theoretical curves

<table>
<thead>
<tr>
<th>Dupl.</th>
<th>MTTF&lt;sub&gt;GOB&lt;/sub&gt;</th>
<th>+ / -</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>0.537</td>
<td></td>
</tr>
<tr>
<td>Block</td>
<td>0.446</td>
<td>-17%</td>
</tr>
<tr>
<td>Gate</td>
<td>4.237</td>
<td>+ 789%</td>
</tr>
<tr>
<td>Trans.</td>
<td>13.552</td>
<td>+ 2524%</td>
</tr>
</tbody>
</table>
Results – Graceful degradation

- Increase of the delay with rising defects

![Graph showing the relationship between the number of defects and delay with trends for gate and transistor duplication.](image)
Results – Graceful degradation

Increase of the overall power with rising defects due to increased static power consumption.
Conclusion

- Need of design improvements for lifetime reliability
- Lowest abstraction level (transistor level) promises the most improvements concerning gate oxide breakdown
- Simpler integration of (as well good) gate level duplication into existing design flows and CAD tools
- Graceful degradation behavior in the presence of defects
Outlook

- Usage of more elaborate and complex breakdown models
- Additional implementation of devices with different gate oxide thickness
- Partial duplication of most vulnerable gates or transistors
- Investigation of the impact of soft breakdowns