Is CMOS more reliable with scaling?

TM Mak
Intel Corporation
Moore’s Law Continues

Heading toward 1 billion transistors in 2007

- 1970: 8008
- 1980: 8080, 8086
- 1990: Pentium® Processor
- 2000: Pentium® II Processor
- 2010: Pentium® IV Processor, Itanium™ Processor (McKinley)

Transistor growth over time:
- 1970: 1,000 transistors
- 2000: 1 billion transistors

Yearly milestones:
- Dec 2000
- June 2001
- Nov 2001
- May 2002
V\text{CC} and V\text{T} Scaling

- V\text{CC} is decreasing more rapidly than V\text{T}
- Transistor drive current is function of \((V\text{CC}-V\text{T})^n\)
Fewer and fewer atoms between the gate and channel

Placing a few $\text{SiO}_N$ species uniformly in billions of devices ??

65nm node $L_{\text{gate}}=30\text{nm}$

Gate oxide less than 3 atomic layers thick

PolySi

Silicon

CMOS reliability with scaling
Gate Oxide Wearout Hot e-

- Gate oxide fails (leakage current increases) in characteristic time dependent on electric field and temperature.
- Gate oxide reliability impose BI / Test voltage stress and Tj limits.
- Trapped charge also shift Vt costing performance degradation over time.
Shrinking Bathtub?

**Infant Mortality** (declining failure rate)
Due to Latent Reliability Defects
Goals: 500 DPM within 0-30 days & 200 FIT within 0-1 year

**Cumulative Fallout Vs. Time**
(follows a lognormal distribution)

**Impact of Burn In:**
Control Infant Mortality

**Wearout** (increasing failure rate)
Due to oxide wearout, EM, hot-e, etc
Goal: <0.1% failing for intrinsic reliability mechanisms

Scope of Burn In

Failure Rate

~1 year

<7 YR Early Wearout

7 YR Wearout Target

.10um .13um .18um

Time
Leakages dominate BI power

<table>
<thead>
<tr>
<th>Active Power</th>
<th>Leakage Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%-75%</td>
<td>20%-25%</td>
</tr>
<tr>
<td>Frequency Reduction by $10^7$</td>
<td>*BI Acceleration Factor</td>
</tr>
<tr>
<td>$\text{BI AF}_{\text{gat}}=2.5-3.5$</td>
<td>$\text{BI AF}_{\text{tr}}=4$</td>
</tr>
<tr>
<td>$\text{BI AF}_{\text{decap}}=6$</td>
<td>These percentages vary with the product!</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transistor Gate Leakage</th>
<th>Transistor Leakage</th>
<th>Decap Leakage</th>
</tr>
</thead>
</table>

BI Active Power

Burn-In Leakage Power @ T=100°C/$V_{CC}=1.61$V
Burn-In Thermal Trends

Model Assumptions

- BI Voltage = 1.4X Vcc
- BI Temp = 90 or 100 C
- Leakage = 3X (130nm)
- % Low Vt = 20
- Lo Vt Lkg = 10X Hi Vt Lkg

Will fit within the current BI envelope
Will fit within the NGBI envelope assuming depopulation and minor improvements
Will not fit into the NGBI envelope unless there are major improvements

Cost of BI increases or more infant mortality fails
Thermal Runaway

- Thermal runaway is a destructive positive feedback condition that can occur when inadequate thermal control is combined with a silicon process technology where leakage increases exponentially with temperature.

Test sockets can be destroyed by thermal runaway

Ref: M Miller, NGBI, 2001
• Dopant Fluctuation causes Vt variation
  – Substantial variation even for close proximity
  – Affect device that require symmetry
• SRAM array (substantial due to # of cells), differential sense amp, current mirrors

Ion implantation for Vt control

Dopant under gate reduces to thousands/hundreds of dopant atoms

TM Mak
CMOS reliability with scaling
Leakage as a defect mode: data retention

- Minute leakage at storage node can cause node to change state when it should not
  - Fault model is a high R bridge across source to drain
- Will not fail if cell is refreshed frequent enough
  - Usual test method is “wait”
Latches and domino gates are equivalent storage nodes

- Failure mode may not be detected at speed
  - \( R \) is not small enough to have appreciable performance difference
  - *Running at low frequency or left alone will cause state loss*

- Coupled with the system and chip design techniques to save power, such as stop clock, frequency changes, Vcc changes, device may have a hard time remembering its machine states

\( V_t \) mismatches exacerbate leakage effects

Leakage increases with wearout!!
Shrinking Process decrease charge per node

Soft error is a function of stored charge at sensitive nodes

\[ Q = CV \]

i.e., \( C_{\text{node}} \) and \( V_{\text{cc}} \)
Logic circuit subjected to SEU

- Logic is not immune to SER
  - All feedback nodes are susceptible
- Errors in compute elements may become SDC (silent data corruption) or at best system crash (equally undesirable)
  - Low end system may be OK with a reboot; not acceptable for a server
  - SDC is the most vulnerable; won’t know unless computation is repeated at another time
Crosstalk (reliability?)

Interconnect Geometry

Coupling vs. Substrate capacitance

Noise $\propto C_c \frac{dV}{dt}$

~30 GHz
Frequency Doubles in Two Years

Noise $\propto C_c \frac{dV}{dt}$
Crosstalk can also be a transient event

- Multiple coupling possible in meshes of wires
- Individual analysis may give a healthy report
  - But their combined forces (e.g. buses) may create enough of a slowdown or significant glitch
  - Not correlated signals may be very hard to analyze (timing windows, functional sensitizable)
- May appear as a random system failure
  - Data dependent and/or machine state dependent

*When will your neighbors all gang on you?*
Summary

• Scaling will bring less reliable electronics unless we come up with new solutions
  – Scaling bring more leakage (Igate and Ioff)
  – Both manufacturing issues (Burn-in and lack thereof) and long term quality/reliability issues (field infant mortality, hot electron degradation and wearout)
  – Leakage problems further exacerbated with Vt mismatches
  – Scaling also increase logic susceptibility to Soft Error
  – Signal integrity may appear as a reliability issue

• Traditional fault tolerance solution don’t work
  – Excessive power consumption
  – 2X+ cost

• Fault tolerance largely an architecture solution; does it mean less testing with adoption of Fault tolerance?