Electrical Variability due to Layout Dependent Effects: Analysis, Quantification, and Mitigation on 40 and 28nm SOC Designs

Mark Zwolinski
mz@ecs.soton.ac.uk
1. Introduction
2. Stress Effects
3. Variability Analysis Flow
4. Results and Discussions – 40nm CMOS Technology
   a) Transistor variations: Vth, Idsat, Ioff
   b) Cell timing and leakage variations
5. 28nm Results
6. Mitigation Strategies
7. Conclusions
1. Introduction

- Layout dependent variations (context dependent):
  - A. Variation in poly pitch.
  - B. Well-proximity effects.
  - C. Intentional and unintentional Stress: LOD, STI, DSL and SiGe.
  - D. Pattern dependent dishing and oxide erosion.
  - E. Rapid thermal anneal (RTA) process.

- We focus on stress effects including Diffusion Spacing Effects (OSE) and Well Proximity Effects (WPE).

- Cadence tool LEA is used to analyze stress effects as results of layout context.

- Device and cell variability due to stress are analysed.

- Mitigation strategies for lower systematic variability are discussed.
2. Stress Effects

- **Unintentional stress: LOD and STI:**
  - **LOD** effect is due to mechanical compressive stress induced at boundary of OD.
  - Proportional to the distance to OD boundary.
  - Layout dependent but not context dependent.

- **STI** becomes compressive as the wafer cools down.
  - The wider STI the higher the stress.
  - Width of STI (Active to Active Spacing) has a strong role in determining the stress.
  - Context dependent.
2. Stress Effects (Cont.)

- Intentional stress: DSL and e-SiGe:
  - DSL applies a SiN film to create tensile stress on n- and compressive stress on p-MOS.
  - The boundaries of compressive and tensile are synthesized from the well layer.
  - Context dependent.

- Well Proximity Effects: WPE →
  - MOS close to well edge exhibits a difference in Vth andIds from that of the device located remotely from well edge.

- Poly Space Effects (PSE)
  - not context dependent if cells have dummy polys (1st polys have same dimensions as poly gate), and effects of 2nd polys are less.
3. Variability analysis flow

- **Library Variability analysis**
  - Understand and quantify context variability
  - Perform quantified area/timing variability architectural and layout tradeoffs
  - Prioritize various layout optimizations or mitigation strategies
  - Optimize selection of context for characterization

- **Device Variability analysis**
  - Vth characterisation.
  - Idsat characterization.
  - Ioff characterization.

- **Path Variability analysis**
  - Analysis of critical paths, Clock Trees, etc
  - More accurate timing analysis and reduce margins
### 3. Path Variability

- **LEA path variability flow:**
  - Create data from Encounter
  - Extracts critical cells with context
  - Launch LVS to extract stress effects due to contexts on critical cells
  - Compute delay difference and back-annotate timing

- **LEA path variability flow is used:**
  - To run analysis on critical paths, Clock Trees, etc
  - To provide more accurate timing analysis and reduce margins
  - In Standalone mode or from Encounter
4. Results: Vth variability, 40 nm

Small logic cell: Inverter consists of 1 n-MOS and 1 p-MOS
1000 random contexts,
Top and Bottom: Regular Layout (Filler Cells),
Left and Right: Random cells from the library.

<table>
<thead>
<tr>
<th>Vth</th>
<th>Spread (mV)</th>
<th>Relative Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMOS</td>
<td>13.4</td>
<td>2.84</td>
</tr>
<tr>
<td>PMOS</td>
<td>14.8</td>
<td>2.88</td>
</tr>
</tbody>
</table>

40nm, n-MOS, Vds=50mV, T=−40°C

40nm, p-MOS, Vds=−50mV, T=−40°C
4. Results: $I_{dsatn}$, $I_{offn}$ variability, 40 nm

<table>
<thead>
<tr>
<th>NMOS</th>
<th>Spread</th>
<th>Relative Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idsat</td>
<td>0.0268 (mA/µm)</td>
<td>5.7</td>
</tr>
<tr>
<td>Ioff</td>
<td>1.35 (pA/µm)</td>
<td>13.9</td>
</tr>
</tbody>
</table>
4. Results: $I_{\text{dsatp}}, I_{\text{offp}}$ variability, 40 nm

<table>
<thead>
<tr>
<th>PMOS</th>
<th>Spread</th>
<th>Relative Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{\text{dsat}}$</td>
<td>0.0168 (mA/µm)</td>
<td>8.3</td>
</tr>
<tr>
<td>$I_{\text{off}}$</td>
<td>1.06 (pA/µm)</td>
<td>20.0</td>
</tr>
</tbody>
</table>
4. Results: Cell variability, 40 nm

<table>
<thead>
<tr>
<th>CELLS</th>
<th>Drive Strength</th>
<th>Max delay spread (%)</th>
<th>CELLS</th>
<th>Drive Strength</th>
<th>Max Output slew spread (%)</th>
<th>CELLS</th>
<th>Drive Strength</th>
<th>Max Leakage spread (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverter</td>
<td>2</td>
<td>15.39</td>
<td>Buffer</td>
<td>1</td>
<td>13.53</td>
<td>Inverter</td>
<td>1</td>
<td>30.48</td>
</tr>
<tr>
<td>Inverter</td>
<td>1</td>
<td>11.74</td>
<td>Inverter</td>
<td>2</td>
<td>13.25</td>
<td>Inverter</td>
<td>2</td>
<td>28.25</td>
</tr>
<tr>
<td>Inverter</td>
<td>0.5</td>
<td>10.90</td>
<td>Inverter</td>
<td>1</td>
<td>12.79</td>
<td>Inverter</td>
<td>0.5</td>
<td>26.83</td>
</tr>
<tr>
<td>Inverter</td>
<td>4</td>
<td>10.49</td>
<td>Buffer</td>
<td>0.5</td>
<td>12.57</td>
<td>Buffer</td>
<td>1</td>
<td>24.19</td>
</tr>
<tr>
<td>Buffer</td>
<td>1</td>
<td>9.86</td>
<td>Buffer</td>
<td>2</td>
<td>10.56</td>
<td>Inverter</td>
<td>4</td>
<td>22.16</td>
</tr>
<tr>
<td>Buffer</td>
<td>8</td>
<td>4.75</td>
<td>Inverter</td>
<td>32</td>
<td>6.70</td>
<td>Buffer</td>
<td>6</td>
<td>12.97</td>
</tr>
<tr>
<td>Buffer</td>
<td>20</td>
<td>4.40</td>
<td>Inverter</td>
<td>20</td>
<td>6.62</td>
<td>Buffer</td>
<td>12</td>
<td>12.96</td>
</tr>
<tr>
<td>Buffer</td>
<td>16</td>
<td>3.65</td>
<td>Buffer</td>
<td>8</td>
<td>6.20</td>
<td>Buffer</td>
<td>8</td>
<td>11.60</td>
</tr>
<tr>
<td>Buffer</td>
<td>24</td>
<td>3.18</td>
<td>Inverter</td>
<td>16</td>
<td>5.88</td>
<td>Buffer</td>
<td>32</td>
<td>11.28</td>
</tr>
<tr>
<td>Buffer</td>
<td>32</td>
<td>2.97</td>
<td>Inverter</td>
<td>24</td>
<td>5.54</td>
<td>Buffer</td>
<td>24</td>
<td>10.41</td>
</tr>
</tbody>
</table>

Worst and best 5 clock tree cells variability from 40 nm cell library
4. Discussion: Cell variability, 40 nm

Smaller cells demonstrate higher variability and bigger cells have lower variability as shown in the above table.

Cell variability is affected mostly by:

1. Devices at the border of the cell because they have higher variability.

2. The proportion (Percentage) of devices along the cell border affects the whole variability.

3. The left and right context cells (WPE and OSE at lateral effects).

4. And top and bottom contexts cells have negligible effects because of the Dummy diffusions (smaller OSE Vertical effects).
5. 28nm devices

- **28nm MOSFETs variability due to stress.**
  - $V_{th}$ and $I_{dsat}$ Variability of n-MOS increase with technology scaling.
  - $I_{off}$ variability is much smaller.
  - Variability of p-MOS is about half of 40nm devices, which may be because of DSL and SiGe technologies.
  - Up to 4% of timing and 0.2% of leakage variability of cells’ are found.
  - The 28nm technology is better than 40nm according to context dependent variability.
## 5. Results of 28nm devices – Cell variability

<table>
<thead>
<tr>
<th>CELLS</th>
<th>Drive Strength</th>
<th>Max delay spread (%)</th>
<th>CELLS</th>
<th>Drive Strength</th>
<th>Max Output slew spread (%)</th>
<th>CELLS</th>
<th>Drive Strength</th>
<th>Max Leakage spread (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverter</td>
<td>20</td>
<td>1.6103</td>
<td>Inverter</td>
<td>24</td>
<td>3.7275</td>
<td>Inverter</td>
<td>1</td>
<td>0.1631</td>
</tr>
<tr>
<td>Inverter</td>
<td>24</td>
<td>1.4437</td>
<td>Buffer</td>
<td>1</td>
<td>3.5992</td>
<td>Inverter</td>
<td>0</td>
<td>0.1631</td>
</tr>
<tr>
<td>Inverter</td>
<td>3</td>
<td>1.2694</td>
<td>Buffer</td>
<td>2</td>
<td>3.4314</td>
<td>Inverter</td>
<td>2</td>
<td>0.1321</td>
</tr>
<tr>
<td>Buffer</td>
<td>4</td>
<td>1.1821</td>
<td>Buffer</td>
<td>8</td>
<td>3.2862</td>
<td>Inverter</td>
<td>3</td>
<td>0.1296</td>
</tr>
<tr>
<td>Inverter</td>
<td>2</td>
<td>1.1562</td>
<td>Buffer</td>
<td>4</td>
<td>3.1377</td>
<td>Buffer</td>
<td>2</td>
<td>0.1037</td>
</tr>
<tr>
<td>Inverter</td>
<td>8</td>
<td>0.728</td>
<td>Inverter</td>
<td>20</td>
<td>1.2376</td>
<td>Buffer</td>
<td>16</td>
<td>0.0128</td>
</tr>
<tr>
<td>Inverter</td>
<td>12</td>
<td>0.69</td>
<td>Inverter</td>
<td>16</td>
<td>1.2306</td>
<td>Buffer</td>
<td>20</td>
<td>0.0103</td>
</tr>
<tr>
<td>Buffer</td>
<td>32</td>
<td>0.653</td>
<td>Buffer</td>
<td>6</td>
<td>1.2293</td>
<td>Inverter</td>
<td>32</td>
<td>0.0095</td>
</tr>
<tr>
<td>Buffer</td>
<td>24</td>
<td>0.528</td>
<td>Inverter</td>
<td>12</td>
<td>1.1215</td>
<td>Buffer</td>
<td>24</td>
<td>0.0086</td>
</tr>
<tr>
<td>Inverter</td>
<td>16</td>
<td>0.091</td>
<td>Inverter</td>
<td>32</td>
<td>1.0738</td>
<td>Buffer</td>
<td>32</td>
<td>0.0063</td>
</tr>
</tbody>
</table>

Worst and best 5 clock tree cells’ timing and leakage variations from 28 nm cell library

Cell variability at 28nm technology is smaller than that of 40nm’s.
6. Possible Mitigation Strategies

1. Effects of diffusion spacing (OSE)

✓ We found that the context dependent systematic variability does not become worse by removing the dummy diffusion.

**Strategy 1:** Remove top and bottom dummy diffusions in future generations.

<table>
<thead>
<tr>
<th>$I_{DSAT}$</th>
<th>Dummy OD</th>
<th>Min (mA/µm)</th>
<th>Max (mA/µm)</th>
<th>Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMOS</td>
<td>With</td>
<td>0.48520</td>
<td>0.51371</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td>W/O</td>
<td>0.47707</td>
<td>0.50599</td>
<td>6.06</td>
</tr>
<tr>
<td>PMOS</td>
<td>With</td>
<td>0.20345</td>
<td>0.22089</td>
<td>8.57</td>
</tr>
<tr>
<td></td>
<td>W/O</td>
<td>0.20663</td>
<td>0.22479</td>
<td>8.79</td>
</tr>
</tbody>
</table>

✓ Bigger vertical diffusion spacing is good for p-MOS current, bigger spacing is suggested to trade-off n and p-MOS current.

**Strategy 2:** Making the top context cells a little farther from the victim.

✓ Device performance benefits from smaller horizontal diffusion spacing:

**Strategy 3:** Making the left and right context cells closer to the victim.
6. Possible Mitigation Strategies (Cont.)

2. Effects of Well Proximity (WPE)

✓ The smaller distance from gate to well edges, the higher WPE.

Strategy 4: Try to place bigger cells at the left and right of the victim.

Schematic diagram of physical distances between gate to well edges

3. Effects of Poly Spacing (PSE)

✓ Stress Variability is not obviously affected by removing dummy polys.

<table>
<thead>
<tr>
<th>$I_{DSAT}$</th>
<th>Dummy Poly</th>
<th>Min (mA/µm)</th>
<th>Max (mA/µm)</th>
<th>Variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMOS</td>
<td>With</td>
<td>0.48520</td>
<td>0.51371</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td>W/O</td>
<td>0.487635</td>
<td>0.516094</td>
<td>5.84</td>
</tr>
<tr>
<td>PMOS</td>
<td>With</td>
<td>0.20345</td>
<td>0.22089</td>
<td>8.57</td>
</tr>
<tr>
<td></td>
<td>W/O</td>
<td>0.20417</td>
<td>0.22154</td>
<td>8.51</td>
</tr>
</tbody>
</table>
7. Conclusions

1. Context dependent stress effects (Oxide Spacing, Well Proximity Effects) are significant for nano devices and cell variability.

2. LEA is used for device, cell, circuit variability analysis due to stress.

3. 40nm MOSFET variability due to stress are reported.

4. Cell variability of 40nm due to stress are affecting SOC designs.

5. Mitigation strategies to reduce context dependent variability.

6. 28nm MOSFETs variability due to stress.

7. Maximum context dependent variability at 40 and 28nm technologies.

<table>
<thead>
<tr>
<th>Process</th>
<th>Vth</th>
<th>Idsat</th>
<th>Ioff</th>
<th>Delay/Slew</th>
<th>Leakage</th>
</tr>
</thead>
<tbody>
<tr>
<td>40nm</td>
<td>2.88%</td>
<td>8.3%</td>
<td>20.0%</td>
<td>15.39%</td>
<td>30.48%</td>
</tr>
<tr>
<td>28nm</td>
<td>3.91%</td>
<td>11.5%</td>
<td>2.18%</td>
<td>3.73%</td>
<td>0.163%</td>
</tr>
</tbody>
</table>
Acknowledgements

This work was supported by a Knowledge Transfer Secondment, funded by the Engineering and Physical Sciences Research Council

Thanks to:

Yangang Wang, Andrew Appleby, Mark Scoones, Sonia Caldwell, Touqeer Azam, Philippe Hurat and Chris Pitchford