Electrical modeling of imperfect lithographic patterning

Puneet Gupta*
Tuck-Boon Chan, Rani S. Ghaida
Dept. of EE, University of California Los Angeles
(puneet@ee.ucla.edu)

Work partly supported by NSF, UC Discovery IMPACT and SRC.

NanoCAD Lab

http://nanocad.ee.ucla.edu/
Outline

• Introduction
• Modeling Poly and Active Imperfections
• Modeling Line-Ends
• Design-Flow Adoption Challenges
• Electrical Impact of Double Patterning Lithography (DPL) Imperfections
• Conclusions
Scaling and Lithography Problems

Figure courtesy Synopsys Inc.
Existing compact device models (e.g., BSIM) do not handle non-rectangular geometries.
Where Are Electrical Models of Patterning Imperfections Needed?

• Cells characterization
• Electrically-driven OPC
  – Converting shape into current
• Contour-based design analysis
  – Estimate power and performance.
• Design rule optimization
• Transistor shape optimization
  – Optimizes non-rectangular transistor for delay-leakage tradeoffs.
Why Wires Are Not Important

- Width variation averages over long wires.
- Resistance and capacitance change in opposite directions as line width changes.

FreePDK 45nm process
Simulation at Chip-Level

- Delay and switching power <3%.
- Impact of wire variation is exaggerated as averaging effect is ignored.

<table>
<thead>
<tr>
<th>Interconnect layers (variation)</th>
<th>Δ delay (%)</th>
<th>Δ Switching power (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2 (+10%)</td>
<td>0.89</td>
<td>1.46</td>
</tr>
<tr>
<td>M2 (-10%)</td>
<td>-0.75</td>
<td>-0.69</td>
</tr>
<tr>
<td>M3 (+10%)</td>
<td>1.90</td>
<td>2.83</td>
</tr>
<tr>
<td>M3 (-10%)</td>
<td>-1.62</td>
<td>-1.85</td>
</tr>
<tr>
<td>M4 (+10%)</td>
<td>0.77</td>
<td>1.64</td>
</tr>
<tr>
<td>M4 (-10%)</td>
<td>-0.65</td>
<td>-0.84</td>
</tr>
<tr>
<td>M5 (+10%)</td>
<td>0.08</td>
<td>0.50</td>
</tr>
<tr>
<td>M5 (-10%)</td>
<td>-0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>M6 (+10%)</td>
<td>0.22</td>
<td>0.65</td>
</tr>
<tr>
<td>M6 (-10%)</td>
<td>-0.19</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Total gates=43K Total area=0.2mm² FreePDK 45nm process
Non-Rectangular Transistor Modeling

• Existing compact device models (e.g., BSIM) do not handle non-rectangular geometries

• Device models for shape imperfections:
  – Polysilicon gate shape contours [Gupta SPIE’06]
  – Diffusion rounding [Gupta ASPDAC’08, Chan VLSID’10]
  – Line-end shortening: gate not completely formed [Gupta DAC’07]
  – Line-end rounding: “tapering”, “necking” or “bulging” [Gupta PMJ’08]
Polysilicon Rounding Model

- Line-edge roughness and poly rounding lead to NRG transistor

- Equivalent gate length (EGL) can be used to represent the current behavior of the transistor to communicate to SPICE
Narrow Width Effect (NWE)

- Dopant densities, well-proximity effects, line-end capacitive coupling, etc. change with distance from STI edge
  - Non-uniform Vth along channel width
  - Ion/loff vs. W plot is not perfectly linear
- The extent and kind of behavior are very process-dependent

<table>
<thead>
<tr>
<th>Variation sources</th>
<th>Vth edge/Vth middle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fringe capacitance</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Well proximity</td>
<td>&gt;= 1</td>
</tr>
<tr>
<td>STI Stress</td>
<td>&lt;= 1</td>
</tr>
</tbody>
</table>
Modeling Location Dependent $V_{th}$

- Threshold voltage modeled as a function of location along channel width
  
  $$V_{th}(x) = \begin{cases} 
  V_{th}(\text{middle}) - K_1(x - w)^2 + K_2(x - w) & 0 \leq x \leq w \\
  V_{th}(\text{middle}) & w \leq x \leq W - w \\
  V_{th}(\text{middle}) - K_1(W - x - w)^2 + K_2(W - x - w) & W - w \leq x \leq W 
  \end{cases}$$

- $K_1$ and $K_2$ can be fitted purely in SPICE regime
  - NWE effect in BSIM $\rightarrow I_{off}$ vs. Width plot
  - $V_{th}$ vs. location can be fitted such that $I_{off}$ of transistor slices match $I_{off}$ vs. Width plot

- Parameters of $V_{th}$ model are estimated using $I_{off}$ data, which is much more sensitive to $V_{th}$
Device Level Modeling Results

TCAD

Uniform $V_{th}$

Location dependent $V_{th}$
Compact Model for Circuit Simulation

• EGLs depend on transistor working states
  – EGLs are extracted at \(|V_{gs}| = 0\) and \(|V_{gs}| = V_{dd}\) for leakage and timing analysis, respectively

• Alternatives:
  – Model a transistor by multiple smaller transistors connected in parallel [Sreedhar ICCD’08]
    ➢ Accurate but number of transistors increases
  – Fit \(L_{\text{eff}}\) and \(V_{th}\) for \(I_{on}\) and \(I_{off}\)
    ➢ Only a set of parameters for a transistor
Other Circuit Models

• Express gate length as a function of $V_{gs}$ in device’s model (e.g., BSIM)
  – Given $L_{eff}$ at $V_{gs} = 0$ and $V_{gs} = V_{dd}$,
  – Intermediate gate length can be estimated using close form equation [Singhal DAC’07]

• Model the impact of gate length variation using voltage dependent current source [Shi ICCAD’06]
  – $I-V$ curve is calculated based on transistor’s shape.
  – $\Delta I$ due to non-rectangular gate is extracted and modeled as a current source connected in parallel to the transistor
The Flip Side

- Use the models to draw non-rectangular transistors intentionally to reduce power
- Proposed alternative: shape the transistor channel to create a dominant device
  - Lower leakage, faster delay, smaller capacitance
- 90nm simulation results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C5315</td>
<td>1.96</td>
<td>1.95</td>
<td>31.93</td>
<td>30.46</td>
<td>4.6</td>
</tr>
<tr>
<td>C6288</td>
<td>5.62</td>
<td>5.61</td>
<td>39.66</td>
<td>38.38</td>
<td>3.2</td>
</tr>
<tr>
<td>C7552</td>
<td>3.19</td>
<td>3.19</td>
<td>36.78</td>
<td>35.08</td>
<td>4.6</td>
</tr>
<tr>
<td>i2</td>
<td>0.86</td>
<td>0.86</td>
<td>13.55</td>
<td>12.80</td>
<td>5.5</td>
</tr>
<tr>
<td>i3</td>
<td>0.45</td>
<td>0.45</td>
<td>6.07</td>
<td>5.74</td>
<td>5.4</td>
</tr>
</tbody>
</table>
Its not only “L”: Diffusion Rounding

- Diffusion rounding occurs due to printing imperfection.
  - Diffusion routing
  - Pwr/Gnd connections
- Modeled as trapezoid gate to investigate electrical performance.

Victor Moroz, Munkang C. & Xi-Wei Lin SPIE 2009
Developing a Physical Diffusion+Poly Rounding Model

- To capture two dimensional E field, slice channel according to its distribution
  - For each slice, $L_{\text{eff-i}} = L_i$
- Effective width is derived using gradual channel approximation:
  \[
  W_{\text{eff-i}} = \frac{(W_{s-i} - W_{d-i})}{\ln(W_{s-i} / W_{d-i})}
  \]
- $V_{th}$ varies due to NWE and asymmetry between source and drain
  \[
  \Delta V_{th-\text{effective}} = \Delta V_{th-\text{Narrow width}} + \Delta V_{th-\text{CS}}
  \]
- Using charge sharing model:
  \[
  \Delta V_{th-\text{CS}} = \frac{qN_a W_c}{2LC_{\text{ox}}} \left[ \frac{2(L_d W_d + L_s W_s)}{W_d + W_s} - (L_d + L_s) \right]
  \]
Total Currents

- Each slice is rectangular with equivalent L, W and $V_{th}$:

$$I_{total} = \sum_{i=1}^{n} f(L_i, W_i, V_{th_i})$$

- Second order effects (DIBL, short channel effects, etc) are implicitly considered in BSIM.
- Evaluate $I_{total}$ at $V_{gs} = 0V$, $V_{ds} = V_{dd}$ (off)
  $$V_{gs} = V_{dd}, V_{ds} = V_{dd} \text{ (on)}$$
- With $I_{total}$, equivalent device for circuit simulation can be obtained using EGL or other methods.

Can be obtained using conventional compact model e.g., (BSIM).
TCAD vs Model (Diffusion Rounding only)

- Asymmetrical $I_{on}/I_{off}$ when rounding happens at Drain/Source terminals
  - $\Delta V_{th}$ varies according to drain/source ratio
# Poly+Diffusion Rounding

<table>
<thead>
<tr>
<th></th>
<th>L1 (nm)</th>
<th>L2 (nm)</th>
<th>W_d (nm)</th>
<th>W_1 (nm)</th>
<th>W_2 (nm)</th>
<th>Error (%)</th>
<th>TCAD cal.</th>
<th>SPICE cal.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I_on</td>
<td>I_off</td>
<td>I_on</td>
</tr>
<tr>
<td>Diffusion rounding</td>
<td>45</td>
<td>45</td>
<td>155</td>
<td>26</td>
<td>0</td>
<td>-2.1</td>
<td>-0.8</td>
<td>-2.0</td>
</tr>
<tr>
<td>(Source side larger)</td>
<td>45</td>
<td>45</td>
<td>155</td>
<td>45</td>
<td>0</td>
<td>-2.0</td>
<td>0.7</td>
<td>-1.9</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>45</td>
<td>155</td>
<td>78</td>
<td>0</td>
<td>-2.8</td>
<td>0.4</td>
<td>-2.7</td>
</tr>
<tr>
<td>Poly rounding only</td>
<td>55</td>
<td>45</td>
<td>155</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>-0.7</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>45</td>
<td>155</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>-0.2</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>45</td>
<td>155</td>
<td>45</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>-1.4</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>45</td>
<td>155</td>
<td>0</td>
<td>45</td>
<td>NA</td>
<td>NA</td>
<td>-2.8</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>45</td>
<td>155</td>
<td>45</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>-2.4</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>45</td>
<td>155</td>
<td>0</td>
<td>45</td>
<td>NA</td>
<td>NA</td>
<td>-0.7</td>
</tr>
</tbody>
</table>

**Average error:**

- (Diffusion layer rounding only)
  - TCAD calibrated model = 1.6%
  - SPICE calibrated model = 1.7%

- (Poly+ Diffusion layers rounding)
  - SPICE calibrated model = 2.7%
Application on Logic Cells

- At 100nm defocus
  \( \Delta \) Delay = 5%
  \( \Delta \) Leakage = 9%
- Design rule can be optimized.
Line-End Imperfection

- line-end shape changes fringing capacitance and narrow width effect
- Fringing capacitance can be modeled by

\[ C_{f\_total} = \sum_i C_{f\_si} + C_{f\_gate\_edge} \]  

[Gupta PMJ’08]
Electrical Impact of Line-End Problems

- **LEE vs. Capacitance**
  Line-end extension increases $C_g$ because there exists fringe capacitance between line-end extension and channel.

- **Capacitance vs. $V_{th}$**
  $C_g$ affects $V_{th}$, narrow width effect
  - $C_g$ increases $\rightarrow$ $V_{th}$ decreases
  - $C_g$ decreases $\rightarrow$ $V_{th}$ increases

- **$V_{th}$ vs. Current**
  $I_{on}$ and $I_{off}$ are functions of $V_{th}$
  - $V_{th}$ increases $\rightarrow$ $I_{on}$, $I_{off}$ decrease
  - $V_{th}$ decreases $\rightarrow$ $I_{on}$, $I_{off}$ increase
Misalignment Model

- There exists misalignment error between gate and diffusion processes.
- Overlapping region (=actual channel) can vary according to misalignment error:
  - Increase linewidth variation.
- Misalignment has a probability, $P(m)$.

$$I_{exp} = \sum_{m=1}^{5} P(m) \cdot I(m)$$
Optimizing Line-End of SRAM

SRAM Bitcell Layout vs. Line-End Design Rule

(Line-End Length, Sharpness) vs. (Leakage, Area)

Large $n$ is better for leakage variation
but it increases OPC and Mask costs.

According to the taper shape, LEE design rule can be optimized to reduce bitcell size.
Line-End Shortening (LES)

- Polysilicon does not cover active region completely
  - Sources: Misalignment and line-end pullback

- Transistor suffering LES:
  - Functionally correct
  - High Leakage power
  - May have hold time violation
Design Flow Integration

• Full-custom/Analog designs
  – SPICE or SPICE-like analyses flows
  – $W_{eq}, L_{eq}$ per transistor is sufficient

• Cell-based digital designs
  – Static analysis flows based on standard cell abstraction
    • One cell is 2-100 transistors
    • Timing/power views stored in pre-characterized “.lib” files
  – Analysis done at PVT “corners”
  – State of art 45nm logic designs have 10M+ cells and 50M+ transistors → Hierarchy preservation essential
Adoption Challenge #1: Simulation Runtime

- “Expected” runtime ~ 1M instances/2 hrs
  - ~1nm accuracy needed for timing analysis
  - Multiple focus, exposure and overlay conditions?

- Tricks to play
  - Simulate only the gate area on Poly and Diff
  - Parallelization
  - Leverage pre-simulated cells
  - Mix of rule-based and model-based approaches
  - Filter simulation areas
    - Timing criticality: simulate only near critical instances
    - Geometric criticality: pattern-based or graph-based filtering

- Added complication: need for incrementality
  - Timing/power optimization \(\rightarrow\) incrementally resimulate after change
  - Trick: use methods which do not require (significant) layout change.
    - E.g., multi-Vt
Adoption Challenge #2: Uniquification

- Lithography simulation + NRG model → potentially all instances of a cell master may be different
  - E.g., 10 Leq steps, 10 transistors in a cell → $10^{10}$ unique cell instances possible
  - Typical cell library size = 1000 cells
  - Typical design size = 10M instances
  - Uniquification and flattening → 10000X increase in library size → intractable STA, etc runtimes; data management nightmare

- Solutions/research needs:
  - Smart pruning of cell variants
    - Snap to pre-chosen set of variants; or
    - Generate minimal set of additional variants
    - Design-context (power/timing) aware
  - Incremental characterization/estimation of variants
    - Transistor-level analysis methods to leverage pre-existing “.libs”

- Similar problems for any systematic variation analysis
  - RTA, strain, etch…
Adoption Challenge #3: SPICE vs. Litho Corners

• Typical BSIM corner methodology
  – Based on a reference pattern context
    • FF, SS & TT correspond to the device placed in the reference context
    • Within this context, parameters (tox, Vt0, etc.) are fitted from silicon over multiple L and W bins
  – Litho-dependency in the pattern contexts outside the reference pattern is not accounted for
    • Prohibitive to cover all contexts
    • Some limited context-dependent “re-centering” of the model

• Typical litho process window
  – Across focus, exposure with multiple patterns

• No explicit connection between L/W variation in litho vs. SS-FF L/W variation in SPICE \( \rightarrow \) No way to connect litho simulation across PW to circuit power/performance analysis
Starting Point: Compact Model for Channel’s Shape

- NRG transistor are modeled as transistor slices connected in parallel
- Detailed description of transistor slices is costly
  - \((\text{transistor \#}) \times (\text{slices \#}) \times (\text{geometrical info})\)
- Example Compact Shape Model:
  - Ignore narrow width effect → slices are independent → can be rearranged

Approximate channel slices by a trapezium
  - \(L\) and \(W\) replaced by \(L_{\text{min}}, L_{\text{max}}, W\) → 1 extra layout-dependent parameter extracted by device extraction
Patterning Methods – Now and Future

• Next generation lithography is not ready at 22nm
  – EUV, nanoimprint and electron beam direct write
• RETs alone are unlikely to be enough
• Alternative solutions:
  – DPL → pitch relaxation using 2 separate exposure/etch steps
  – Interference assisted lithography → form 1D grating and remove unwanted features with a trim-exposure
  – Source-mask optimization → enhance printability using pixellated source and limited set of layout patterns
• Challenges of these solutions:
  – impose restrictions on layout
  – carry serious implications on design
Double Patterning Lithography

- ≈ 2X pitch relaxation
- But many challenges and implications for design
Within Layer Overlay

- Within-layer overlay translates into linewidth/spacing variation
  - depending on process flavor
- For devices (poly)
  - Gate spacing affects liner stress
  - Gate-to-contact spacing affects source/drain resistance, gate-to-contact cap, and liner stress
- For wires [Ghaida SPIE’09]
  - Delay variation can reach up to 17% for a line segment but...
  - Max. variation = 3.4% for a path
    - Indirect benefit due to congestion
    - Averaging
  - Up to 50mV increment in peak crosstalk glitch
Bimodality Problem

- Different exposure/etch steps → two CD populations
- Overlay is another contributor to bimodality
- Large CD/delay variability (e.g., 34% $3\sigma$ increase - by ASML study)

\[
3\sigma^2_{pooled} = \frac{3\sigma^2_{p1}}{2} + \frac{3\sigma^2_{p2}}{2} + \left(\frac{3}{2} |\mu_{p1} - \mu_{p2}|\right)^2
\]

- Loss of spatial correlation
- Timing problems: clock skew and worse timing slack (e.g., 53ps and 46ps assuming 6nm CD difference [Jeong ASPDAC’09])
Other Layout Dependent Sources of Variability

- Layout-dependent stress variation (e.g., 15% $\Delta I_{on}$)
- Well proximity effect on $V_{th}$ (e.g., up to 10% delay increase)
- Etch introduces CD variability with strong dependence on pattern-density within a few microns range
- RTA used in the fabrication of ultra-shallow junctions
  - Long-range effect (few millimeters)
  - Affects $I_{on}/I_{off}$ ratio and $V_{th}$.
- CMP imperfections of dishing and erosion
  - Causes interconnect $RC$ variability
  - Depends on line-width/spacing and pattern-density within a long-range (up to 100 micron)
Summary

• Lithographic variation is a major source of gate’s length and width variations.
  – Wires not all that important
  – Non-rectangular transistor modeling can reduce pessimism in design rules as well as enable accurate power/performance analyses.
  – Adoption of electrical model strongly depends on
    • RET and patterning technologies.
    • Layout restrictions for manufacturability.
    • Contribution of lithography to total electrical variability

• Other sources of layout-dependent variability
  – Layout-dependent stress variation (e.g., 15% $\Delta I_{on}$)
  – Well proximity effect on $V_{th}$ (e.g., up to 10% delay increase)
  – Etch bias
  – RTA induced $V_{th}$
  – CMP imperfections of dishing and erosion