Nanolithography and Design-Technology Co-optimization Beyond 22nm

David Z. Pan
Dept. of Electrical and Computer Engineering
The University of Texas at Austin
http://www.cerc.utexas.edu/utda
50+ Years Ago, ...

Still

There's Plenty of Room at the Bottom
- An Invitation to Enter a New Field of Physics

Richard P. Feynman, 1959

The Moore, The Better!
Outline

♦ Introduction

♦ Nanolithography for 22nm and Beyond
  ♦ Double Pattern Lithography
  ♦ Emerging Lithography

♦ Some Other Design-Technology Co-optimization Issues
  ♦ NBTI/PBTI
  ♦ 3D Integration: TSV, Stress, Reliability

♦ Conclusions
Nanometer Issues

Litho

Random defects

Etch

CMP

 CHIP performance

Occurrence

Defect Size

Disappearance

Rounding

Pinch, Pullback

Random defects

(F, O)

Mask

Bowing

Necking

Direct

Ion

IdsatP

IdsatN
“Next” Generation Lithography

193i w/ DPL

EUV

Nanoimprint
Don’t Forget Other Objectives

Interconnect determines the overall performance
Power/leakage/thermal issues
Other “technology” related issues: NBTI, HCI, FINFET

(source: ITRS)  (source: Intel)

…
More Moore and More than Moore

♦ More Moore: continue pushing the envelope, 22nm, 15nm (14nm), 11nm, 8nm (ITRS)
  › Computational Scaling (pushing 193nm)
  › Double Patterning
  › Emerging Nanolithography

♦ More than Moore: New design-technology co-optimization issues
  › Vertically – 3D IC integration
  › New device/material: FINFET, optical interconnect, …
  › Nano-X
  › ……
Outline

♦ Introduction

♦ Nanolithography for 22nm and Beyond:
  ♦ Double Pattern Lithography
  ♦ Emerging Lithography

♦ Some Other Design-Technology Co-optimization Issues
  ♦ NBTI

♦ 3D Integration: TSV, Stress, Reliability

♦ Conclusions
Double Patterning Lithography

- For 22nm and 16nm, the industry most likely will adopt double patterning lithography (DPL)
- A key problem is overlay control
  - Double exposures, masks, …
- Intelligent CAD solution to compensate unwanted overlay effects or even take advantage of them!
- [Yang et al, ASPDAC’ 2010]
  - A new layout decomposition framework
  - Graph-theoretic, multi-objective
Issues with DPL

Minimum Stitch Insertion

Overlay Compensation

1) Yield loss with overlay
2) Area increase due to overlap margin

[Lucas SPIE’08]

Without Overlay Compensation

With Overlay Compensation

\[ \begin{align*}
C_1 - \Delta C_1 & \quad C_2 - \Delta C_2 \\
1^\text{st} \text{ patterning} & \quad 1^\text{st} \text{ patterning}
\end{align*} \]

\[ \begin{align*}
C_1 - \Delta C_1 & \quad C_2 + \Delta C_2 \\
1^\text{st} \text{ patterning} & \quad 2^\text{nd} \text{ patterning}
\end{align*} \]
## Comparisons with Previous Works

<table>
<thead>
<tr>
<th></th>
<th>Balanced Density</th>
<th>Overlay Compensation</th>
<th>Stitch Minimization</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Yao+, ICCAD08]</td>
<td>No</td>
<td>No</td>
<td>Yes (ILP)</td>
<td>NP-Complete</td>
</tr>
<tr>
<td>[Yuan+, ISPD09]</td>
<td>No</td>
<td>No</td>
<td>Yes (ILP)</td>
<td>NP-Complete</td>
</tr>
<tr>
<td>[Xu+, ICCAD09]</td>
<td>No</td>
<td>No</td>
<td>Yes (ILP)</td>
<td>NP-Complete</td>
</tr>
<tr>
<td><strong>Our Framework [ASPDAC10]</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (Bi-Partitioning)</td>
<td>Polynomial Time $O(N\log N)$</td>
</tr>
</tbody>
</table>


Benefits of Balanced Density

S38584: 13% and 87%

S38584: 50% and 50%

C432: 27% and 73% (7 stitches)

C432: 50% and 50% (17 stitches)
Overlay Compensation Result

- Without TDD: 9% variation (Weight=0.0)
- One stitch: 5.274% variation (Weight=0.2)
- Three stitches: 1.098% variation (Weight=0.5)
- Nine stitches: 0.018% variation (Weight=1.0)
Spacer-type DPL

- SADP (self-aligned double patterning)
- Core mask and trim mask
- Less overlay cf. LELE
Challenges in SADP

- A single width of sidewall spacer
- Does not allow ‘stitch’ points
- SADP currently in production only for 1D patterns
  - NAND Flash memory applications
- SADP for 2D random logic patterns is challenging
- [Ban et al., DAC’11] proposes systematic techniques to perform layout decomposition for general 2D patterns
How to Solve Coloring Conflicts?

- The space/width of the merged region should be equal or larger than the minimum space/width of the trim mask.
- Trim mask overlay at the merged region
22nm Metal1 Standard Cell

(1) Target layout
(2) Mandrel & spacer
(3) Trim mask
(4) Final patterns
Electronic Beam Lithography

- Maskless technology, which shoots desired patterns directly into a silicon wafer
- Low throughput is its major hurdle
  - Variable Shaped Beam (VSB)

Total number of 11 shots are needed
Character Projection (CP) Technology

- Print some complex shapes in one electronic beam shot, rather than writing multiple rectangles.

3 shots only
Overlapped Characters

- The number of characters is limited due to the area constraints of the stencil.

- By overlapping adjacent characters/sharing blank spaces, more characters can be put on the stencil.
Character candidates to be considered

Stencil

Order Matters

Out of Stencil
Stencil Planning and Optimization

**#shots (projection time)**

- **NON-OVERLAP**: Blue bars
- **GREEDY**: Purple bars
- **PROPOSED**: Green bars

**#characters on stencil**

- **NON-OVERLAP**: Light blue bars
- **GREEDY**: Purple bars
- **PROPOSED**: Green bars

**#CPU (logscale)**

- **NON-OVERLAP**: Light blue bars
- **GREEDY**: Purple bars
- **PROPOSED**: Green bars

51%, 14% reduction on shot number over previous ILP-based approach without overlapping characters and greedy algorithm.
Outline

♦ Introduction

♦ Nanolithography for 22nm and Beyond:
  ♦ Double Pattern Lithography
  ♦ Emerging Lithography

♦ Some Other Design-Technology Co-optimization Issues
  ♦ NBTI and Clock Network Design
  ♦ 3D Integration: TSV, Stress, Reliability

♦ Conclusion
What is NBTI?

♦ NBTI is a key failure mechanism for PMOS
♦ Cause PMOS Vth to drift when driven by GND
  › E.g., $|\Delta V_{TH}| = +60\text{mV}$ after 10 years
  › 30% increase in inverter delay
♦ NBTI-Induced Skew Management in Gated Clock Trees [Chakraborty+, DATE 2009, ISPD 2010]
  › Main problem: clock gating cause imbalance between different clock buffers/receivers
  › Key idea: try to balance NBTI degradation
  › Both circuit design (run time) and CAD techniques (design time)
♦ Similar principle holds for PBTI
Clock Gating Induced $\Delta V_{TH}$ Imbalance

- Using NAND gate reduces SP0 at output
- Using NOR gate increases SP0 at output
- In both cases, $\Delta V_{TH}$ mismatch will exist!
If \( \{ GATE = FALSE \} \) \nElse If \( \{ SELECT = 0 \} \) \nElse

\[
\begin{align*}
CLK\_OUT &= CLK \\
CLK\_OUT &= 0 \\
CLK\_OUT &= 1
\end{align*}
\]
Determine clock gating NAND/NOR during design
  › Not runtime (less penalty and no SELECT signals)

Main idea:
  › Optimally pick NAND and NOR gates for clock gating

Symbolic SP0 Propagation

SP0 Aware Delay Characterization

Symbolic Arrival Time Computation

Skew Minimization Formulation (ILP)
Delay is Function of CLK Gating Assignment

\[
\begin{align*}
    \text{DINV}(0.5) + \\
    X2 \cdot \text{DNAND}(0.5) + X2' \cdot \text{DNOR}(0.5) + \\
    (X4 \cdot \text{DNAND}(0.72 - X2 \cdot 0.5) + X4' \cdot \text{DNOR}(0.75 - X2 \cdot 0.5))
\end{align*}
\]
## Results

<table>
<thead>
<tr>
<th>CKT</th>
<th>Solver Time (s)</th>
<th>OUR Skew (ps)</th>
<th>All NAND (ps)</th>
<th>All NOR (ps)</th>
<th>10 Rand. (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.14</td>
<td>2.80</td>
<td>4.41</td>
<td>9.02</td>
<td>7.24</td>
</tr>
<tr>
<td>B</td>
<td>0.06</td>
<td>2.18</td>
<td>3.23</td>
<td>5.84</td>
<td>4.96</td>
</tr>
<tr>
<td>C</td>
<td>1.41</td>
<td>4.13</td>
<td>6.4</td>
<td>9.28</td>
<td>7.05</td>
</tr>
<tr>
<td>D</td>
<td>0.81</td>
<td>3.03</td>
<td>5.04</td>
<td>9.74</td>
<td>6.21</td>
</tr>
<tr>
<td>E</td>
<td>0.12</td>
<td>2.76</td>
<td>5.46</td>
<td>10.21</td>
<td>7.04</td>
</tr>
<tr>
<td>F</td>
<td>0.09</td>
<td>3.94</td>
<td>6.21</td>
<td>12.23</td>
<td>11.82</td>
</tr>
<tr>
<td>G</td>
<td>0.47</td>
<td>3.88</td>
<td>6.75</td>
<td>13.07</td>
<td>10.58</td>
</tr>
<tr>
<td>H</td>
<td>0.09</td>
<td>2.59</td>
<td>3.91</td>
<td>8.44</td>
<td>5.38</td>
</tr>
<tr>
<td>Avg:</td>
<td></td>
<td>1</td>
<td>1.56X</td>
<td>2.19X</td>
<td>1.33X</td>
</tr>
</tbody>
</table>

- **Age the circuit to 10 years**
- **Our > Rand > NAND > NOR solution**
- **Significantly tightens the skew budget**
3D IC Integration

Better Performance
- Massive Bandwidth
- Reduced Interconnect Delays
- Power Reduction (Less IO driver)
- Higher Functionality/Space
- Heterogeneous Integration

Smaller Size
- 3D Maximizes Space Utilization

Lower Cost
- Lower Cost vs. Next-gen Device
- Reuse of Proven SIP

[Courtesy of Dr. H.-M., Tong, ASE]
3D IC Yield

<table>
<thead>
<tr>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
<th>$Y_5$</th>
<th>$Y_6$</th>
<th>$Y_7$</th>
<th>$Y_8$</th>
<th>$Y_9$</th>
<th>$Y_{10}$</th>
<th>Overall Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>95.0%</td>
</tr>
<tr>
<td>99.5%</td>
<td>99.5%</td>
<td>90.0%</td>
<td>90.0%</td>
<td>90.0%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>99.5%</td>
<td>70.0%</td>
</tr>
</tbody>
</table>

$Y_1 = \text{Joint Yield}$

$Y_2 = \text{Repassivation/RDL Yield}$

$Y_3 = \text{Interface Yield}$

$Y_4 = \text{TSV Yield}$

$Y_5 = \text{Interface Yield}$

$Y_6 = \text{Repassivation/RDL Yield}$

$Y_7 = \text{Joint Yield}$

$Y_8 = \text{Joint Yield}$

$Y_9 = \text{Substrate Yield}$

$Y_{10} = \text{Joint Yield}$

[Courtesy of Dr. H.-M., Tong, ASE]
Thermal Stress Impact Near TSV

<table>
<thead>
<tr>
<th>Material</th>
<th>CTE in $10^{-6}$/K at 20°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>3</td>
</tr>
<tr>
<td>W</td>
<td>4.5</td>
</tr>
<tr>
<td>Cu</td>
<td>17</td>
</tr>
</tbody>
</table>

CTE: Coefficient of thermal expansion

[Dao+, ICICDT’ 2009]

TSV: 250 °C ~ 400 °C process (Higher than operating temperature)
Since Cu has larger CTE than Si, tensile stress is in Si near TSV.

< Tensile stress >

Cu TSV  Silicon

[Selvanayagam+, ECTC’ 08, TAP’ 09]

< Fast NFET, slow PFET with tensile stress >

[H.S. Yang, IEDM’ 2004]
Stress Aware Design Flow [Yang+, DAC’10]

- Pre-placed TSV location
- Stress estimation induced by TSVs
  - Mobility change ($\Delta \mu / \mu$) calculation
  - Cell characterization with mobility (Cell name change in Verilog)
    - Stress aware Verilog netlist
    - Verilog, SPEF merging for 3D STA
      - 3D Timing Analysis with PrimeTime
        - Critical gate selection
          - TSV stress aware layout optimization

- Liberty file having cell timing with different mobility
- Verilog netlist
- Optimized layout with TSV stress
Stress Effect on Mobility & Current

**CMOS (Stress: 200MPa, R=r)**

- **NMOS:** $0.5 \Delta \mu$ ($\Delta I_{ds}: +1.5\%$)
- **PMOS:** $0.6 \Delta \mu$ ($\Delta I_{ds}: +1.8\%$)

- **NMOS:** $0.75 \Delta \mu$ ($\Delta I_{ds}: +2.25\%$)
- **PMOS:** $-0.1 \Delta \mu$ ($\Delta I_{ds}: -0.3\%$)

**Cell characterizations based on distance and orientation are needed.**
Cell instantiation depending on location

**INVX1_P4_P6**

$(\Delta \mu/\mu)_e = +4\%$

$(\Delta \mu/\mu)_h = +6\%$

**INVX1_P8_N14**

$(\Delta \mu/\mu)_e = +8\%$

$(\Delta \mu/\mu)_h = -14\%$

**INVX1_P8_N8**

$(\Delta \mu/\mu)_e = +8\%$

$(\Delta \mu/\mu)_h = -8\%$

**INVX1_P2_0**

$(\Delta \mu/\mu)_e = +2\%$

$(\Delta \mu/\mu)_h = 0\%$

Identify hole and electron mobility variation according to TSV induced stress

- Rename cells based on the mobility

Cell naming: INVX1_P8_N8

P8: +8% electron mobility variation

N8: -8% hole mobility variation
Inverter Delay Dependence on Stress

- Electron Mobility Variation:
  \[ \Delta \mu_e : 0\% \sim 24\% \text{ in our test case} \]
  \[ \Rightarrow \Delta D_{\text{falling}} : \text{up to } 7.5\% \]

- Hole Mobility Variation:
  \[ \Delta \mu_h : -22\% \sim 10\% \text{ in our test case} \]
  \[ \Rightarrow \Delta D_{\text{rising}} : \text{more than } 20\% \]
## TSV Specification

<table>
<thead>
<tr>
<th>Width</th>
<th>Landing pad</th>
<th>KOZ</th>
<th>Height</th>
<th>Dielectric</th>
<th>Resistance</th>
<th>Capacitance</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.14um</td>
<td>4.54um</td>
<td>0.4um</td>
<td>20um</td>
<td>0.2um</td>
<td>0.1</td>
<td>70fF</td>
</tr>
</tbody>
</table>

## Stress effect on critical paths

<table>
<thead>
<tr>
<th>Circuit</th>
<th>#Cells</th>
<th>Without TSV stress</th>
<th>With TSV stress</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Longest Delay(ns)</td>
<td>TNS(ns)</td>
<td>Longest Delay(ns)</td>
</tr>
<tr>
<td>IDCT</td>
<td>14,864</td>
<td>12.07</td>
<td>-21293</td>
<td>11.91</td>
</tr>
<tr>
<td>8051</td>
<td>15,712</td>
<td>4.78</td>
<td>-7868</td>
<td>4.94</td>
</tr>
<tr>
<td>8086</td>
<td>19,895</td>
<td>9.56</td>
<td>-8557</td>
<td>9.56</td>
</tr>
<tr>
<td>MAC2</td>
<td>29,706</td>
<td>7.72</td>
<td>-17561</td>
<td>7.72</td>
</tr>
<tr>
<td>ETHERNET</td>
<td>77,234</td>
<td>18.3</td>
<td>-476</td>
<td>18.95</td>
</tr>
<tr>
<td>RISC</td>
<td>88,401</td>
<td>8.28</td>
<td>-1249</td>
<td>8.34</td>
</tr>
<tr>
<td>B18</td>
<td>103,711</td>
<td>11.28</td>
<td>-2082</td>
<td>11.25</td>
</tr>
<tr>
<td>DES_PERT</td>
<td>109,181</td>
<td>8.61</td>
<td>-2801</td>
<td>8.64</td>
</tr>
<tr>
<td>VGA_LCD</td>
<td>126,379</td>
<td>8.01</td>
<td>-543</td>
<td>8.14</td>
</tr>
<tr>
<td>B19</td>
<td>168,943</td>
<td>13.01</td>
<td>-5539</td>
<td>12.98</td>
</tr>
<tr>
<td>average</td>
<td>75,403</td>
<td>10</td>
<td>-6,797</td>
<td>10</td>
</tr>
</tbody>
</table>
## Result: Timing Optimization

### Critical path manual optimization (Circuit: 8051)

<table>
<thead>
<tr>
<th>Logic Depth</th>
<th>Gate</th>
<th>Original</th>
<th>Optimized</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ΔHole(%)</td>
<td>ΔElectron (%)</td>
<td>Gate</td>
</tr>
<tr>
<td>1</td>
<td>NOR3X1</td>
<td>-2</td>
<td>14</td>
<td>DFFPOSX1</td>
</tr>
<tr>
<td>2</td>
<td>AND2X1</td>
<td>-12</td>
<td>12</td>
<td>NOR3X1</td>
</tr>
<tr>
<td>3</td>
<td>INGX1</td>
<td>-6</td>
<td>12</td>
<td>INGX1</td>
</tr>
<tr>
<td>4</td>
<td>INGX1</td>
<td>-12</td>
<td>12</td>
<td>INGX1</td>
</tr>
<tr>
<td>5</td>
<td>AND2X1</td>
<td>-16</td>
<td>16</td>
<td>AND2X1</td>
</tr>
<tr>
<td>6</td>
<td>BUFX2</td>
<td>6</td>
<td>12</td>
<td>BUFX2</td>
</tr>
<tr>
<td>7</td>
<td>AOI22X1</td>
<td>4</td>
<td>10</td>
<td>AOI22X1</td>
</tr>
<tr>
<td>8</td>
<td>INGX1</td>
<td>0</td>
<td>10</td>
<td>INGX1</td>
</tr>
<tr>
<td>9</td>
<td>OR2X1</td>
<td>-4</td>
<td>10</td>
<td>OR2X1</td>
</tr>
<tr>
<td>10</td>
<td>OR2X2</td>
<td>-16</td>
<td>18</td>
<td>OR2X2</td>
</tr>
<tr>
<td>11</td>
<td>NOR3X1</td>
<td>0</td>
<td>14</td>
<td>NOR3X1</td>
</tr>
<tr>
<td>12</td>
<td>NAND3X1</td>
<td>-4</td>
<td>14</td>
<td>NAND3X1</td>
</tr>
<tr>
<td>13</td>
<td>BUFX2</td>
<td>-4</td>
<td>14</td>
<td>BUFX2</td>
</tr>
<tr>
<td>14</td>
<td>OR2X2</td>
<td>-</td>
<td>8</td>
<td>OR2X2</td>
</tr>
<tr>
<td>15</td>
<td>AOI22X1</td>
<td>-16</td>
<td>16</td>
<td>AOI22X1</td>
</tr>
<tr>
<td>16</td>
<td>OAI21X1</td>
<td>-4</td>
<td>14</td>
<td>OAI21X1</td>
</tr>
<tr>
<td>17</td>
<td>NOR3X1</td>
<td>2</td>
<td>14</td>
<td>NOR3X1</td>
</tr>
<tr>
<td>18</td>
<td>AOI21X1</td>
<td>-18</td>
<td>18</td>
<td>AOI21X1</td>
</tr>
<tr>
<td>19</td>
<td>INGX1</td>
<td>-16</td>
<td>16</td>
<td>INGX1</td>
</tr>
<tr>
<td>20</td>
<td>OAI21X1</td>
<td>6</td>
<td>14</td>
<td>OAI21X1</td>
</tr>
</tbody>
</table>

| Path Delay | 4.937 | 4.618 | -6.5% |
Result: Cell Perturbation

Original cell placement

Rising critical optimization with hole contour

Falling critical optimization with electron contour

After cell perturbation
Consider TSV stress during placement [ICCAD’10]

Full-chip TSV stress modeling with multiple TSVs and physical layout optimization issues [Mitra+, ECTC’11]

TSV EMI analysis [Pak+, ECTC’11]

Due to vast difference in size differences
Conclusion

♦ Some new research problems in nanolithography and design-technology co-optimization
  › Pushing the lithography limits:
    » double patterning, triple/quadruple patterning
    » E-beam lithography (stencil planning, e-beam proximity effects)
    » EUV lithography (flare effects, etc.)
  › Resilient design with built-in compensation and error correction (NBTI/PBTI, overlay effects, etc.)
  › 3D-IC manufacturability and reliability issues
  › ……

♦ Holistic treatment in a vertically integrated manner
Synergistic Design-Technology Co-opt

- Need good **levers at different levels** for design-technology co-optimization (DTC)

“Give me a place to stand on, and I can move the earth.”
- Archimedes’ Lever

DTC lever for your sub-22nm billion transistor design!