Statistical Circuit Optimization Considering Device and Interconnect Process Variations

I-Jye Lin, Tsui-Yee Ling, and Yao-Wen Chang

The Electronic Design Automation Laboratory
Department of Electrical Engineering
National Taiwan University
March 17, 2007
Outline

Introduction
Deterministic Algorithm
Statistical Algorithm
Experimental Results
Conclusions
Outline

Introduction
Deterministic Algorithm
Statistical Algorithm
Experimental Results
Conclusions
Interconnect Process Variation

Interconnect delay and reliability highly affect VLSI performance.
The variability of interconnect parameters will raise up to 35%.

The worst-case corner models cannot capture the worst-case variations in interconnect delay.
— Liu et al., DAC 2000

The interconnect optimization guided by statistical analysis techniques has become an inevitable trend.
— Visweswariah, SLIP 2006
Previous Work in Statistical Optimization

Statistical gate sizing with timing constraints using Lagrangian Relaxation.


Statistical power minimization by delay budgeting using second order conic programming.

- Orshansky et al., DAC 2005.

Statistical gate sizing using geometric programming

- Patil et al., ISQED 2005.

No statistical optimization work consider both interconnect and device sizing.
### Comparison with Previous Work

<table>
<thead>
<tr>
<th></th>
<th>Orshansky’s work (DAC 2005)</th>
<th>Our work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sizing variable</strong></td>
<td>Gate only</td>
<td>Gate and wire</td>
</tr>
<tr>
<td><strong>Delay Model</strong></td>
<td>Linear model (linear term)</td>
<td>Elmore delay model (nonlinear term)</td>
</tr>
<tr>
<td><strong>Objective</strong></td>
<td>Power</td>
<td>Area</td>
</tr>
<tr>
<td><strong>Constraint</strong></td>
<td>Timing</td>
<td>Power, timing, thermal</td>
</tr>
</tbody>
</table>

Due to the nonlinear term introduce by the Elmore delay model, the optimization using both gate and wire sizing will be much harder to solve.
Delay Model

Our delay model and timing constraint:

- Elmore delay model

\[
D_i = R_g \left( \frac{C_w X_w L_w + C_g X_i}{X_j} \right) + R_w L_w \left( \frac{C_w X_w L_w / 2 + C_g X_i}{X_w} \right)
\]

- Timing constraint: \[ a_i \geq a_j + D_i \]
- \[ a_i = \text{arrival time of gate } i \]

Delay model and timing constraint used in previous work in DAC 2005:

\[ a_i \geq a_j + d_i^0 + d_i \]

- \[ d_i^0 = \text{delay due to the sizing for maximum slack} \]
- \[ d_i = \text{slack added to node } i \text{ due to the loading} \]
Statistical Circuit Optimization with SOCP

Second-order conic programming (SOCP)

Minimize $f^T x$
subject to $\|A_i x + b_i\|_2 \leq c_i^T x + d_i$
$Fx = g$

- Convex optimization
- Theoretical runtime $O(N^{1.3})$
- Orshansky (DAC 2005), Davoodi (DAC 2006)

Second-order conic constraint:
$\|A_i x + b_i\|_2 \leq c_i^T x + d_i$

Approximation method

Linear terms!
Nonlinear (quadratic) terms are not applicable!
Approximation Method

Fix the gate size in the timing constraint.
- Reduce the timing constraint from quadratic order to linear order.

Approximate the gate sizes by a two-stage flow.
- Iteratively reduce the approximation errors.
- The flow is similar to Sequential Linear Program (SLP).

Solve the SOCP problem under current constraints (gate size fixed)

Update the gate size of timing constraints and form a new SOCP problem

convergence or max iterations

Finish
Our Contributions

The first work of statistical optimization on circuit interconnect and devices
  – Previous work considers only circuit devices (gates).
  – Statistical optimization for considering both interconnect and devices is much harder.

The first work that statistically optimizes the area with thermal- and timing-constrained parametric yields
  – Most existing statistical optimization considers only timing.

The first work capable of analytically transforming the statistical RC model into an SOCP
  – Previous work uses linear delay model
Outline

Introduction

Deterministic Algorithm

Statistical Algorithm

Experimental Results

Conclusions
# of paths may grow exponentially to the circuit size. To reduce problem size, we distribute the timing information to each node.

\[ D_i \leq a_i \quad i = 1, \ldots, s \quad / \ast \text{primary inputs} \ast / \]

\[ a_j + D_i \leq a_i \quad i = s + 1, \ldots, n+s \text{ and } \forall j \in \text{input}(i) \]

\[ a_j \leq D^B \quad j \in \text{input}(m) \quad / \ast \text{primary outputs} \ast / \]
Electron Migration (EM) lifetime reliability of metal interconnects is governed by the well-known Black’s equation:

\[
TTF = A^* \cdot j^{-n} \cdot \exp\left(\frac{Q}{k_B T_m}\right),
\]

The design is reliable when

\[
\frac{j_{avg}^2}{\exp\left(\frac{Q}{k_B T_m}\right)} \leq \frac{j_0^2}{\exp\left(\frac{Q}{k_B T_{ref}}\right)}.
\]

\[
T_m \leq \frac{Q \cdot T_{ref}}{Q - 2K_B T_{ref}(\ln j_0 - \ln j_{avg})} = T^{B'}.
\]

\[
T_m = \Delta T_{self-heating} + T_{environment}
\]
Average Temperature of the Chip

The average temperature of the chip, $T_{\text{avg}}$, can be estimated by:

$$T_{\text{avg}} = T_{\text{air}} + R_n \left( \frac{P_{\text{tot}}}{A} \right)$$

- $P_{\text{tot}}$: total power consumption of the chip
- $T_{\text{air}}$: ambient temperature
- $R_n$: thermal resistance of the substrate and the package
- $A$: chip area

— Banerjee et al., ISPD 2001.
Power Constraint

Need to constrain chip’s temperature under a reasonable bound during the optimization:

\[ \alpha_i c_i \leq P_i^{B'}, \quad P_i^{B'} = \frac{P_i^B}{V_{DD}^2 f} \]

- For simplicity, consider the dynamic power consumption only.
- \( P_i^B \): the power bound of the gate \( i \)
- \( c_i \): the downstream capacitance of the gate \( i \)
- \( \alpha_i \): switching activity of component \( i \)
Deterministic Formulation

\[
\begin{align*}
\text{Minimize} & \quad \sum_{i=s+1}^{n+s} l_i x_i \\
\text{subject to} & \quad T_{mi} \leq T_{i}^B, i \in W, \\
& \quad D_i \leq a_i, i = 1, \ldots, s \\
& \quad a_j + D_i \leq a_i, i = s + 1, \ldots, n + s \text{ and } \forall j \in \text{input}(i) \\
& \quad a_j \leq D^B, j \in \text{input}(m) \\
& \quad \alpha_i c_i \leq P^{B'}, i = s + 1, \ldots, n + s \\
& \quad L_i \leq x_i \leq U_i, \forall s + 1 \leq i \leq n + s.
\end{align*}
\]

\(f\): working frequency; \(\alpha_i\): switching activity of component \(l\); \(C_j\): load capacitance of component \(l\); \(\omega\): path in the path set \(\Omega\).
Outline

Introduction
Deterministic Algorithm
Statistical Algorithm
Experimental Results
Conclusions
Variation Models

Introduce two process parameters as the variation sources: Inter-layer dielectric (ILD) thickness (H), and metal thickness (T).

R and C can be approximated by the first-order Taylor expression:

\[ R = R_{\text{nom}} + a_1 \Delta T, \]
\[ C = C_{\text{nom}} + b_1 \Delta T + b_2 \Delta H \]

- \( a_1, b_1, b_2 \) are sensitivities calculated by the differential differentiation of:

\[ R = \frac{\rho}{WT}, \]

\[ \frac{C_{\text{gnd}}}{\epsilon} = \frac{W}{H} + 3.28 \left( \frac{T}{T + 2H} \right)^{0.023} + \left( \frac{S}{S + 2H} \right)^{1.16} \]


\( R_{\text{nom}}/C_{\text{nom}} \): nominal value of \( R/C \)

\( \Delta T/\Delta H \): random deviation of metal thickness/ILD thickness

\( C_{\text{gnd}}/\epsilon \): normalized capacitance

\( S \): space between parallel lines
RC Variability

Assume T and H are Gaussian, the variability magnitude of R and C can easily be calculated by:

\[ \sigma_R^2 = a_1^2 \sigma_T^2 \]
\[ \sigma_C^2 = b_1^2 \sigma_T^2 + b_2^2 \sigma_H^2 \]

Apply the *interconnect delay variation metric* to calculate the variability of the product of R and C.

- Well captured by a normal distribution with 1.2% average error of the mean delay and 3.8% average error of the standard deviation.
- Blaauw et al., DAC 2004.
**Statistical Formulation**

**Deterministic formulation**

\[
\begin{align*}
\text{Minimize} & \quad \sum_{i=s+1}^{n+s} l_i x_i \\
\text{subject to} & \quad T_{mi} \leq T_i^B \\
& \left\{ \begin{array}{l}
D_i \leq a_i \\
a_j + D_i \leq a_i \\
a_j \leq D^B \\
\alpha_i c_i \leq P^{B'} \\
L_i \leq x_i \leq U_i
\end{array} \right.
\end{align*}
\]

**Statistical formulation**

\[
\begin{align*}
\text{Minimize} & \quad \sum_{i=s+1}^{n+s} l_i x_i \\
\text{subject to} & \quad \text{Prob} \left( T_{mi} \leq T_i^B \right) \geq \delta \\
& \left\{ \begin{array}{l}
\text{Prob} \left( D_i \leq a_i \right) \geq \zeta \\
\text{Prob} \left( a_j + D_i \leq a_i \right) \geq \zeta \\
\text{Prob} \left( a_j \leq D^B \right) \geq \zeta \\
\text{Prob} \left( \alpha_i c_i \leq P^{B'} \right) \geq \eta \\
L_i \leq x_i \leq U_i
\end{array} \right.
\end{align*}
\]

– $\delta/\zeta/\eta$: Thermal/Timing/Power yield constraint
Transformation into SOCP

**Theorem:** Given independent Gaussian random vectors $a_i$ and bound vectors $b_i$, the parametric yield ($\eta$) problem is as follows:

\[
\begin{align*}
\text{Minimize} & \quad \sum x_i \\
\text{subject to} & \quad \text{Prob} (a_i^T x_i \leq b_i) \geq \eta,
\end{align*}
\]

the problem can be reformulated as an SOCP:

\[
\begin{align*}
\text{Minimize} & \quad \sum x_i \\
\text{subject to} & \quad (a_i^T x_i) + \Phi^{-1} (\eta) (x^T \Sigma x)^{1/2} \leq b_i.
\end{align*}
\]

- $\Phi^{-1}$: the cumulative density inverse function

---

The EDA Laboratory

**Transformation Flow**

\[ P \left( a_i^T x_i \leq b_i \right) \geq \eta \]

\[ u = a_i^T x_i \]

\[ \text{variance: } \sigma^2 \]

\[ \text{mean: } \mu \]

\[ P \left( \frac{u - \mu}{\sigma} \leq \frac{b_i - \mu}{\sigma} \right) \geq \eta, \]

zero mean unit variance

Gaussian variable

\[ \Phi \left( \frac{b_i - \mu}{\sigma} \right) \geq \eta \]

cumulative density function

\[ \frac{b_i - \mu}{\sigma} \geq \Phi^{-1} (\eta) \]

\[ \mu + \Phi^{-1} (\eta) \sigma \leq b_i \]

\[ (\bar{a}_i^T x_i) + \Phi^{-1} (\eta) \left( x^T \Sigma_i x \right)^{1/2} \leq b_i \]
Thermal & Power Constraints in SOCP Form

Thermal constraint:

\[ P \left( T_{mi} \leq T_i^B \right) \geq \delta \]

\[ \bar{T}_{mi} + \phi^{-1}(\delta) \sqrt{\sigma_1^2 x_{w1}^2 + \sigma_2^2 \frac{1}{x_{w2}^2} + \sigma_3^2 + \sigma_4^2} \leq T_i^B, \]

Power (Thermal distribution) constraint:

\[ P \left( \alpha_i c_i \leq P_i^{B'} \right) \geq \eta \]

\[ \alpha_i \bar{c}_i + \phi^{-1}(\eta) \sqrt{\sigma_5^2 \sum_{i=s+1}^{n+s} (x_i^2)} \leq P_i^{B'}, \]
Timing Constraint in SOCP Form

\[ D_i = R_g \left( C_w X_w L_w + C_g X_i \right) / X_j + R_w L_w \left( C_w X_w L_w / 2 + C_g X_i \right) / X_w \]

Only \( X_w \) is the sizing variable

\[ D_i = R_j C_w X_w L_w + R_j C_i + R_w C_w (L_w)^2 / 2 + R_m L_w C_i / X_w \]

Timing constraint:

\[ P(a_j + D_i \leq a_i) \geq \zeta \]

\[ a_j + \bar{D}_i + \phi^{-1} (\zeta) \sqrt{\frac{\sigma^2}{6} x_w^2 + \frac{\sigma^2}{7} + \frac{\sigma^2}{8} + \frac{\sigma^2}{x_w^2}} \leq a_i \]
Statistical Problem Formulation using SOCP

\[ \mathbf{P} : \text{Minimize} \]

\[ \sum_{i=s+1}^{n+s} l_i x_i \]

subject to

\[ T_{mi} + \phi^{-1}(\delta) \sqrt{ \sigma_1^2 x_{mi}^2 + \sigma_2^2 \frac{1}{x_{mi}^2} + \sigma_3^2 + \sigma_4^2 } \leq T_i^B, \]

Thermal constraint

\[ a_j + D_i + \phi^{-1}(\zeta) \sqrt{ \sigma_5^2 x_w^2 + \sigma_i^2 + \frac{\sigma_8^2}{x_w^2} } \leq a_i, \]

Timing constraint

\[ \alpha_i \bar{c}_i + \phi^{-1}(\eta) \sqrt{ \sigma_5^2 \sum_{i=s+1}^{n+s} (x_i^2) } \leq P_{B'}, \]

Power constraint

\[ L_i \leq x_i \leq U_i, \; \forall s+1 \leq i \leq n+s. \]
Begin

Formulate the problem with the RC variation

Assign the values of the current gate sizes to the gate size variables in the timing constraint

Formulate the problem into SOCP with gate size as fixed value in the timing constraint

Solve it with the interior point method

Convergence or Max iterations

no

yes

Iteratively reduce the approximation errors.

End
Outline

Introduction
Deterministic Algorithm
Statistical Algorithm
Experimental Results
Conclusions
Experimental Setup

<table>
<thead>
<tr>
<th>Circuit Name</th>
<th>Circuit Size</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#Gate</td>
<td>#Wire</td>
<td>#Total</td>
</tr>
<tr>
<td>c17</td>
<td>11</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>c432</td>
<td>122</td>
<td>230</td>
<td>352</td>
</tr>
<tr>
<td>c499</td>
<td>246</td>
<td>396</td>
<td>642</td>
</tr>
<tr>
<td>c880</td>
<td>256</td>
<td>230</td>
<td>486</td>
</tr>
<tr>
<td>c1355</td>
<td>297</td>
<td>555</td>
<td>852</td>
</tr>
<tr>
<td>c1908</td>
<td>201</td>
<td>336</td>
<td>537</td>
</tr>
<tr>
<td>c2670</td>
<td>499</td>
<td>754</td>
<td>1253</td>
</tr>
<tr>
<td>c3540</td>
<td>429</td>
<td>1021</td>
<td>1450</td>
</tr>
<tr>
<td>c5315</td>
<td>927</td>
<td>1792</td>
<td>2719</td>
</tr>
<tr>
<td>c6288</td>
<td>1298</td>
<td>3596</td>
<td>4894</td>
</tr>
</tbody>
</table>

Implemented in C++ & applied the MOSEK optimization tool to solve it. Tested on the commonly used ISCAS85 benchmark circuits in this area. Used Design Compiler & Astro with UMC 0.18 μm technology library to synthesize and place the circuits.
## Experimental Results

Achieve 51%, 39%, and 26% area reductions for 70%, 84.1%, and 99.9% yield constraints, respectively.

Avg. / Max. # of the running iterations: 5.6 / 10

Timing constraint error bound: 2%

<table>
<thead>
<tr>
<th>Circuit name</th>
<th>Deterministic</th>
<th>70% yield</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>area (µm²)</td>
<td></td>
</tr>
<tr>
<td>c17</td>
<td>7160</td>
<td>0.06</td>
</tr>
<tr>
<td>c432</td>
<td>47752</td>
<td>0.24</td>
</tr>
<tr>
<td>c499</td>
<td>127103</td>
<td>0.41</td>
</tr>
<tr>
<td>c880</td>
<td>152804</td>
<td>0.37</td>
</tr>
<tr>
<td>c1355</td>
<td>174896</td>
<td>0.58</td>
</tr>
<tr>
<td>c1908</td>
<td>96968</td>
<td>0.33</td>
</tr>
<tr>
<td>c2670</td>
<td>275967</td>
<td>0.74</td>
</tr>
<tr>
<td>c3540</td>
<td>362409</td>
<td>1.10</td>
</tr>
<tr>
<td>c5315</td>
<td>913522</td>
<td>1.88</td>
</tr>
<tr>
<td>c6288</td>
<td>1455730</td>
<td>5.23</td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental Results of 84.1% and 99.9% yield

The lower the yield constraints, the better the area optimization. All constraints (timing, power, thermal) are met.

<table>
<thead>
<tr>
<th>Circuit name</th>
<th>84.1% yield</th>
<th>99.9% yield</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>area (µm²)</td>
<td>Area improv.</td>
</tr>
<tr>
<td>c17</td>
<td>3394</td>
<td>52.60%</td>
</tr>
<tr>
<td>c432</td>
<td>27860</td>
<td>41.66%</td>
</tr>
<tr>
<td>c499</td>
<td>57758</td>
<td>54.56%</td>
</tr>
<tr>
<td>c880</td>
<td>66420</td>
<td>56.53%</td>
</tr>
<tr>
<td>c1355</td>
<td>147397</td>
<td>15.72%</td>
</tr>
<tr>
<td>c1908</td>
<td>65020</td>
<td>32.95%</td>
</tr>
<tr>
<td>c2670</td>
<td>161426</td>
<td>41.51%</td>
</tr>
<tr>
<td>c3540</td>
<td>169331</td>
<td>53.28%</td>
</tr>
<tr>
<td>c5315</td>
<td>735838</td>
<td>19.45%</td>
</tr>
<tr>
<td>c6288</td>
<td>1109090</td>
<td>23.81%</td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>39.21%</td>
</tr>
</tbody>
</table>
# Delay, Power and Temperature Performance

Though the delay and the maximum metal temperature are increased, they all meet the given bounds.

- Fully utilized the constraint bound to get the best optimization results.

<table>
<thead>
<tr>
<th>Circuit Name</th>
<th>Delay (ns)</th>
<th>Power (mW)</th>
<th>Max T\text{_increase} (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bound</td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>c17</td>
<td>36.82</td>
<td>22.19</td>
<td>32.21</td>
</tr>
<tr>
<td>c432</td>
<td>247.65</td>
<td>154.59</td>
<td>136.62</td>
</tr>
<tr>
<td>c499</td>
<td>186.13</td>
<td>153.79</td>
<td>135.32</td>
</tr>
<tr>
<td>c880</td>
<td>253.43</td>
<td>208.84</td>
<td>170.15</td>
</tr>
<tr>
<td>c1355</td>
<td>274.55</td>
<td>203.45</td>
<td>241.45</td>
</tr>
<tr>
<td>c1908</td>
<td>222.91</td>
<td>161.09</td>
<td>136.11</td>
</tr>
<tr>
<td>c2670</td>
<td>290.84</td>
<td>176.66</td>
<td>229.94</td>
</tr>
<tr>
<td>c3540</td>
<td>507.80</td>
<td>308.59</td>
<td>245.85</td>
</tr>
<tr>
<td>c5315</td>
<td>445.58</td>
<td>313.52</td>
<td>421.54</td>
</tr>
<tr>
<td>c6288</td>
<td>1333.91</td>
<td>913.33</td>
<td>1148.41</td>
</tr>
<tr>
<td>Comparison</td>
<td>1</td>
<td>1.13</td>
<td>1</td>
</tr>
</tbody>
</table>
Outline

Introduction
Deterministic Algorithm
Statistical Algorithm
Experimental Results
Conclusions
Conclusions

Presented the first statistical work for area minimization under thermal and timing constraints by gate and wire sizing.

- Obtained much better results than those of the deterministic method.

Formulated statistical RC model by SOCPs which can be solved efficiently and effectively.

- Used more accurate delay model (Elmore delay model)
- Solved the problem by a two-stage approximation flow
  - Nonlinear terms are not applicable to SOCP
Thank You
Backup Slides
Temperature Distribution

Applying the Finite Difference Method (FDM), we can divide the whole chip into \( m \) mesh nodes and calculate each node’s temperature by

\[
G_2 T_P = \begin{bmatrix}
g_{1,1} & g_{1,2} & \cdots & g_{1,m} \\
g_{2,1} & g_{2,2} & \cdots & g_{2,m} \\
\vdots & \vdots & \ddots & \vdots \\
g_{m,1} & g_{m,2} & \cdots & g_{m,m}
\end{bmatrix}
\begin{bmatrix}
T_1 \\
T_2 \\
\vdots \\
T_m
\end{bmatrix}
= \begin{bmatrix}
P_1 \\
P_2 \\
\vdots \\
P_m
\end{bmatrix} = P_P
\]

- \( P_i \): \( i \)th mesh node’s power dissipation
- \( T_i \): \( i \)th mesh node’s temperature
- \( g \): power density of the heat sources (W/m\(^3\))

Temperature Dependent Delay

An inseparable aspect of electrical power distribution and signal transmission through the interconnects. Resistance is dependent on Temperature.

\[ r(x) = \rho_0 \left(1 + \beta \cdot T(x)\right) \]

- \( \rho_0 \): the resistance per unit length at reference temperature
- \( \beta \): the temperature coefficient of resistance (1/°C)
Interconnect Temperature Calculation

The interconnect temperature is given by

\[ T_m = \Delta T_{self-heating} + T_{environment} \]

\[ = I_{rms}^2 \cdot R \cdot \theta_{int} + T_{environment} \]

\[ = \frac{\sigma V_{cross}^2 t_{ox} t_m}{K_{ox} l^2 R_m} \cdot \frac{x_i}{x_i + \phi t_{ox}} + T_{environment}. \]

- \( x_i \): wire width
- \( \theta_{int} \): the thermal impedance of the interconnect line to the chip
- \( \sigma \): duty cycle
- \( V_{cross} \): cross voltage of wire
- \( t_{ox} \): the total thickness of the underlying dielectric
- \( t_m \): the thickness of the wire
- \( K_{ox} \): the thermal conductivity
- \( l \): wire length
- \( R_m \): the temperature dependent unit resistance
- \( \psi \): the heat spreading parameter

**Not linear functions**

**Least Square Estimator**
Least squares solves the problem by finding the line for which the sum of the square deviations (or residuals) in the d direction (the noisy variable direction) are minimized.

- Apply Cramer Rule to find the $A_1$ and $A_0$, which minimizes the square deviations

\[ y = A_1 x + A_0 \]

Cramer Rule:

\[
A_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(f_{1i} - \bar{f}_1)}{\sum_{i=1}^{N} (x_i - \bar{x})^2}
\]

\[
A_0 = \bar{f}_1 - A_1 \bar{x},
\]

$\bar{x}$ is the average of $x_i$, $\bar{f}_1$ is the average of $f_1(x_i)$
Approximation for Thermal Constraint

Let $N = 5$ and pick five sizes of $x_i$, we can approximate the thermal constraint by Least Square Estimator (LSE).

— Banerjee et al., DAC 1999