On the futility of statistical power optimization

Jason Cong
Puneet Gupta
John Lee
UCLA

Sponsored by SRC Contract 2006-TJ-1460, NSF award 811832, and UC Discovery Grant ele07-10283
Power Variability

- Leakage Power: 331% (2015)
- Total Power: 88% (2015)
- Performance: 63% (2015)

From ITRS Roadmap 2007 (Design)
Statistical Power Optimization

- Costs of upgrading to Statistical Power Optimization
  - Tools
    - Programming
    - Validation
  - Modeling
    - Extract statistics (Monte-Carlo runs)

- Limitations of Statistical Power Optimization
  - Errors in modeling physical behavior
  - Errors in predicting input / output combinations
Statistical Power Optimization

- Power measures are similar

What is the value of Statistical Power Optimization?
Evaluating the benefits of Statistical Power Optimization

I. Sub-optimality Bounds
   – “What is the maximum improvement that can be gained by optimizing statistically?”
   – Results for benchmark designs in 45 nm library with gate width sizing

II. Extension to Practical Solvers: Solution Rankings
   – Solvers that are non-optimal
   – If a deterministic solution is in the top 10% of all deterministic sizings, will it in the top 10% of all statistical solutions?
   – Experimental validation for w, l, vt
Statistical Power Optimization

- Works with the statistical power random variable:

\[ P = P_1 + P_d \]

\[ = \sum_i \kappa_i w_i e^{\alpha l_i + \beta l_i} e^{-\gamma v_i} e^{\eta_i(\Delta L_{dt} + \Delta L_{i,wd})} \]  
\[ + \sum_i \mu_i w_i (l_i + \Delta L_{dt} + \Delta L_{i,wd}) \]  

- Optimize w, l, vt
  - Help manage the variability in leakage / dynamic power
  - Make designs aware of the effects of variation
Assumptions

- Variations are in gate length only
  - Nominal channel length: 45nm
  - Die-to-die standard deviation: 1nm
  - Within-die standard deviation: .5nm

- Leakage power is Log-Normal

- Deterministic power is linear in gate sizes
  - For l and vt, rewrite in terms of z:
    \[ p = k_i (w_i e^{\alpha l_i^2 + \beta i e^{-\gamma v_i}}) = k_i z_i \]
  - Statistical power can then be written as:
    \[ p_{\text{statistical}} = k_i z_i e^{\eta_i (\Delta L_{\text{dtd}} + \Delta L_{i, \text{wid}})} \]

- Commercial tools return the optimal deterministic sizing solution
Contrast with Statistical Delay Optimization

- Benefits of statistical delay optimization have been shown
  - (c.f. Guthaus et. al GLSVLSI 2005)

- Corner based methods are competitive with full statistical delay optimization
  - (Najm DAC 2005, Burns et. al. DAC 2007)

- Our work is separate from the statistical delay question
  - Deterministic delay is used in this work
  - Delay model is only used for an initial deterministic solution
I. Sub-optimality bounds

**Given:**
- Optimal deterministic sizing solution
  - Synthesized to Nangate Open Cell Library (45nm standard cell library)

**Find:**
- What is the maximum improvement that can be gained by optimizing statistically?

**Example:**
- Gate width sizing examples for benchmark circuits
Calculating bounds: Overview

- Timing feasible region is complex - difficult to find bounds
- Use a simpler set that contains the timing feasible region
- Optimize over the simpler set to get a lower bound
- Use lower bound to find maximum improvement from Statistical Optimization
Calculating bounds: Example

- Visual example: two gates

- Deterministic Power Measure
- Statistical Power Measure
- Timing feasible region
- Iso-power plane
- Sizes with greater deterministic power
- Sizes with greater deterministic power
- Timing feasible region is Bounded by half-space!!!
Calculating bounds: Step 1

Bounding the timing feasible region

a. Deterministic power is linear in gate sizes, e.g.:

\[ p_d(\mathbf{w}) = p(\mathbf{w}_d^*) + \nabla p(\mathbf{w}_d^*)^T(\mathbf{w} - \mathbf{w}_d^*), \quad (\mathbf{w} \in \mathbb{R}^n) \]

b. Deterministic power optimum \( \mathbf{w}_d^* \): smallest power sizing in the timing feasible region:

\[ p_d(\mathbf{w}) \geq p_d(\mathbf{w}_d^*) \rightarrow \nabla p(\mathbf{w}_d^*)^T(\mathbf{w} - \mathbf{w}_d^*) \geq 0 \]

Timing feasible region is contained in a simpler region:

\[ \{ \mathbf{w} \mid p_d(\mathbf{w}) \geq p_d(\mathbf{w}_d^*) \} \subseteq \{ \mathbf{w} \mid 0 \leq \nabla p(\mathbf{w}_d^*)^T(\mathbf{w} - \mathbf{w}_d^*) \} \]
Calculating bounds: Step 2

Optimize over the simpler region

- Using non-linear programming to solve:

\[
\begin{align*}
\text{minimize} & \quad p_{\text{statistical}}(w) \\
\text{subject to} & \quad 0 \leq \nabla p(w_d^*)^T (w - w_d^*) \\
& \quad w_{\text{min}} \leq w \leq w_{\text{max}}
\end{align*}
\]

- The solution $w^*$ is a lower bound on the true statistical optimum $w^*_{\text{statistical}}$
  - Timing feasible region is relaxed to a larger, continuous region

Statistical power

Simpler region (contains timing feasible region)
Calculating bounds: Step 3

Create bound

- $w'$ is a lower bound on the statistical optimum, $w^\text{statistical}$

\[ p_{\text{statistical}}(w') \leq p_{\text{statistical}}(w^\text{statistical}) \leq p_{\text{statistical}}(w^\text{deterministic}) \]

suboptimality gap

- Bound the suboptimality gap using the percentage:

\[ \delta_{\text{so}} = \frac{p_{\text{statistical}}(w^\text{deterministic}) - p_{\text{statistical}}(w')}{p_{\text{statistical}}(w^\text{deterministic})} \]
Bounds for Benchmarks: Example

Worst Case Sub-optimalities:

<table>
<thead>
<tr>
<th></th>
<th>Mean Power $\delta_{so}$</th>
<th>Mean + 3 Sigma $\delta_{so}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic Power</td>
<td>.24% (= 13.24 - 13.21) / 13.24</td>
<td>2.5% (= 17.35 - 16.91) / 17.35</td>
</tr>
</tbody>
</table>

Optimized Sizes for Deterministic Power:

- Deterministic Power: 13.08 µW
- Mean Power: 13.24 µW
- Mean + 3 Sigma Power: 17.35 µW

Lower Bound Calculation:

- Mean Power Lower Bound: 13.21 µW
- Mean + 3 Sigma Power Lower Bound: 16.91 µW
Sub-optimality results: Leakage power optimization

### ISCAS ’85 benchmarks and ALU circuit

#### Synthesized speeds

<table>
<thead>
<tr>
<th></th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
<th>v4</th>
<th>avg</th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
<th>v4</th>
<th>avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>c432</td>
<td>0.3%</td>
<td>0.4%</td>
<td>0.3%</td>
<td>5.3%</td>
<td>7.7%</td>
<td>6.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c499</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>1.7%</td>
<td>1.0%</td>
<td>1.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c880</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>2.1%</td>
<td>1.7%</td>
<td>1.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1355</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>2.0%</td>
<td>3.9%</td>
<td>4.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1908</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>2.8%</td>
<td>2.1%</td>
<td>2.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c2670</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>1.2%</td>
<td>1.7%</td>
<td>2.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c3540</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>2.7%</td>
<td>2.7%</td>
<td>2.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c5315</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.2%</td>
<td>2.0%</td>
<td>1.5%</td>
<td>1.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c6288</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>2.2%</td>
<td>1.2%</td>
<td>1.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c7552</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>1.9%</td>
<td>2.7%</td>
<td>2.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alu</td>
<td>0.2%</td>
<td>0.2%</td>
<td>1.9%</td>
<td>2.7%</td>
<td>2.5%</td>
<td>1.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### (Upper bounds on the improvement from using Statistical Power Optimization)
Sub-optimality results:
Total power optimization

ISCAS ’85 benchmarks and ALU circuit

<table>
<thead>
<tr>
<th></th>
<th>Mean Power</th>
<th>Mean + 3 Sigma Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>switching</td>
<td>minimum</td>
<td>maximum</td>
</tr>
<tr>
<td>probability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>~0%</td>
<td>0.003%</td>
</tr>
<tr>
<td>0.10%</td>
<td>0.001%</td>
<td>0.004%</td>
</tr>
</tbody>
</table>

- The impact of statistical power variations is diminished by the dynamic power
  - Dynamic power is larger than leakage power
  - Deterministic and statistical dynamic power are highly linearly correlated
  - Variations in dynamic power are smaller
II. Solution Rankings

Question

- Suppose the deterministic solution is within the top 5% of all deterministic sizings
- Will this also be in the top 5% of all statistical solutions?

Experimental validation

- Generated random $w$, $l$, $vt$ assignments
- For each assignment:
  - Compared the deterministic power with the statistical power
Solution Rankings

Deterministic Power vs. Statistical Power
(random size assignments)

Relation is nearly linear!
(Small amount of noise)

Benchmark c432
Quantifying the correlation

- The deterministic and statistical powers are nearly linear relations:
  \[ p_{\text{statistical}}(w, l_{\text{eff}}, V_t) = \left( \alpha p_{\text{deterministic}}(w, l_{\text{eff}}, V_t) + \beta \right) + \text{error}(w, l_{\text{eff}}, V_t) \]

- Error statistics:

<table>
<thead>
<tr>
<th>Leakage Power</th>
<th>Mean Power</th>
<th>Mean + 3 Sigma Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>variables</td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>w</td>
<td>.004%</td>
<td>.03%</td>
</tr>
<tr>
<td>w, vt</td>
<td>.009%</td>
<td>.08%</td>
</tr>
<tr>
<td>w, vt, l</td>
<td>.016%</td>
<td>.14%</td>
</tr>
</tbody>
</table>

Total Power (switching frequency = .001)

| w, vt, l      | .005% | .10% | .024% | .077% | 3.3% | .98% |
Summary

- Presented framework to:
  - Bound the maximum improvement that can be gained by optimizing statistically
  - Experimentally compare the statistical quality of a deterministic sizing

- Statistical power optimization gives modest gains
  - Leakage power: on average 2-3% improvement at best
  - Total power: < 1% improvement at best

- Quality deterministic power solutions are quality statistical power solutions
  - The values correlate nearly linearly with small error
  - Expect the sub-optimality to be small
Future Goals

- Model Vt variations
- Statistical delay measures
- Generalized distributions