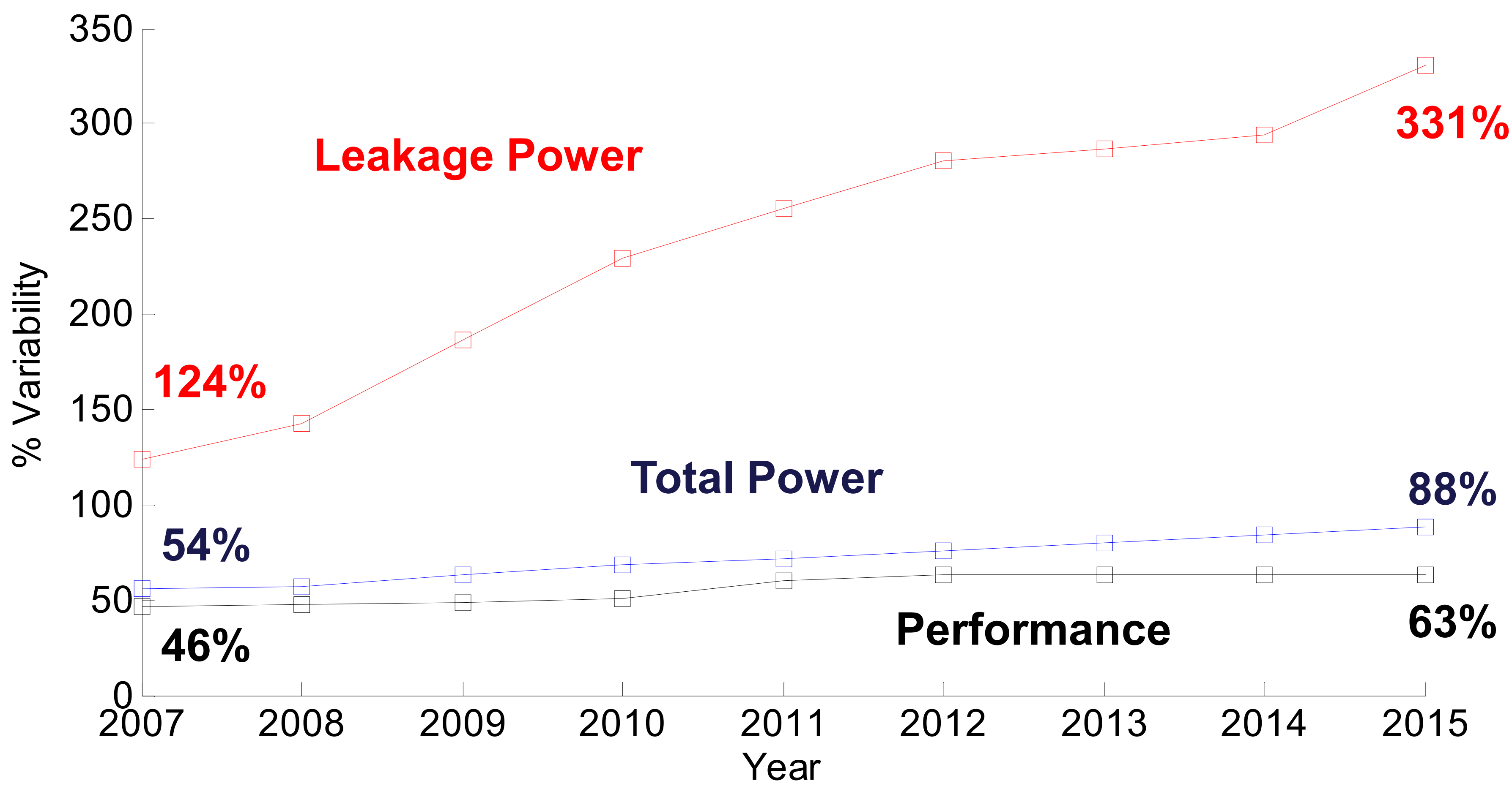
On the futility of statistical power optimization

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



From ITRS Roadmap 2007 (Design)

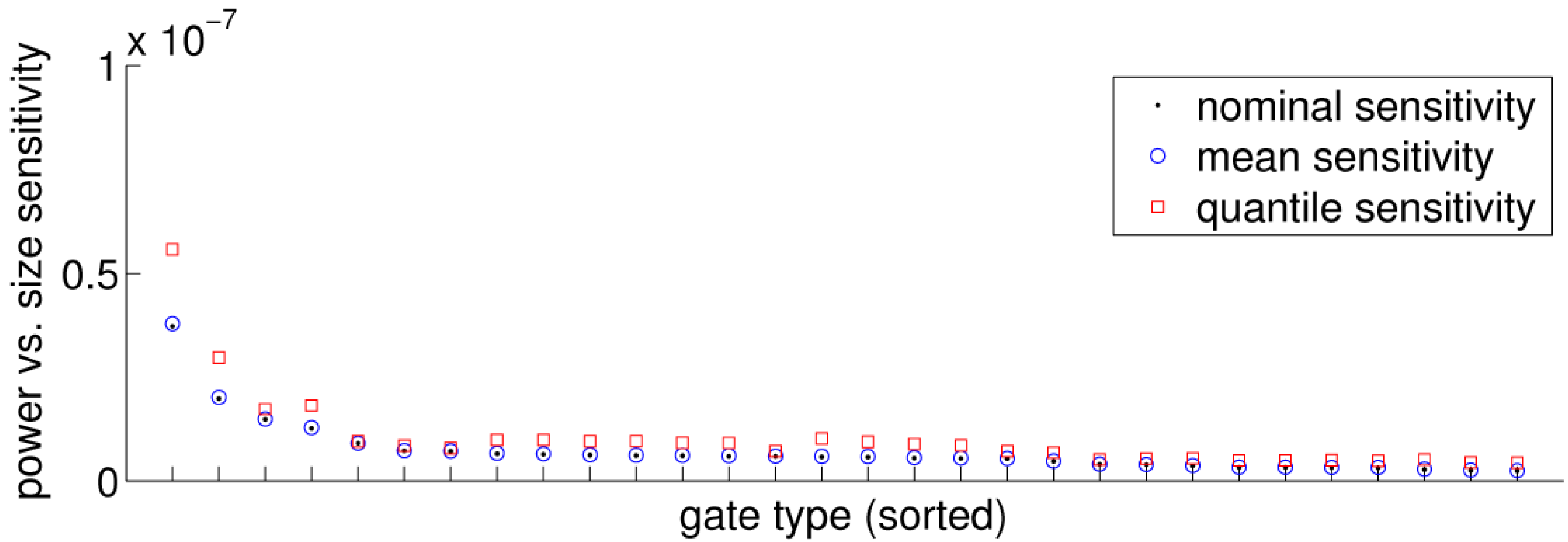
Statistical Power Optimization

- **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:**

$$\mathbf{P} = \mathbf{P}_{\mathrm{l}} + \mathbf{P}_{\mathrm{d}}$$

$$= \sum_i \kappa_i w_i e^{\alpha l_i + \beta l_i} e^{-\gamma v_{ti}} e^{\eta_i (\Delta \mathbf{L}_{\mathrm{dtd}} + \Delta \mathbf{L}_{i,\mathrm{wid}})} \quad \text{Statistical Leakage Power}$$

$$+ \sum_i \mu_i w_i (l_i + \Delta \mathbf{L}_{\mathrm{dtd}} + \Delta \mathbf{L}_{i,\mathrm{wid}}) \quad \text{Statistical Dynamic Power}$$

- **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:

$$\mathrm{p} = k_i(w_i e^{\alpha l_i^2 + \beta_i} e^{-\gamma v_{ti}}) = k_i z_i$$

  - Statistical power can then be written as:

$$\mathrm{p}_{\mathrm{statistical}} = k_i z_i e^{\eta_i (\Delta \mathbf{L}_{\mathrm{dtd}} + \Delta \mathbf{L}_{i,\mathrm{wid}})}$$

- **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)

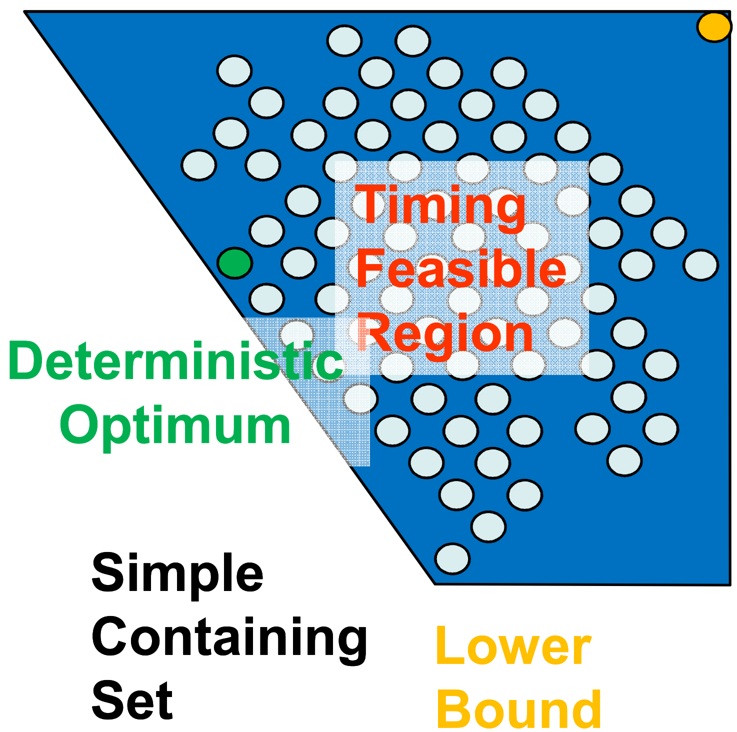
Find:

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

Example:

- **Gate width sizing examples for benchmark circuits**

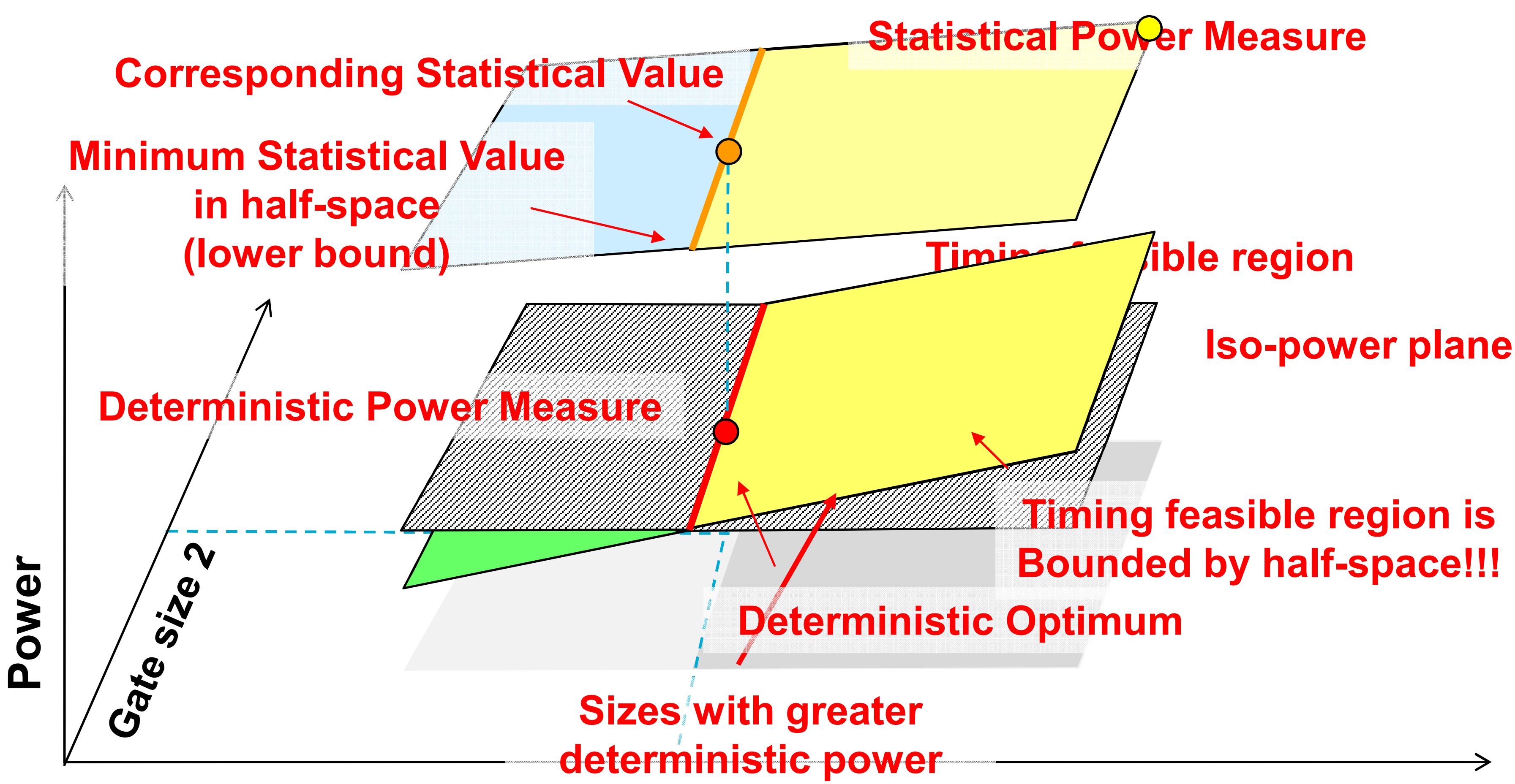
Calculating bounds: Overview



- 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



Calculating bounds: Step 1

Bounding the timing feasible region

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

$$p_{\mathrm{d}}(\mathbf{w}) = p(\mathbf{w}_{\mathrm{d}}^{*}) + \nabla p(\mathbf{w}_{\mathrm{d}}^{*})^{T}(\mathbf{w} - \mathbf{w}_{\mathrm{d}}^{*}), \quad (\mathbf{w} \in \mathbb{R}^{n})$$

b. Deterministic power optimum $\mathbf{w}_{\mathrm{d}}^{*}$: smallest power sizing in the timing feasible region:

$$p_{\mathrm{d}}(\mathbf{w}) \geq p_{\mathrm{d}}(\mathbf{w}_{\mathrm{d}}^{*}) \rightarrow \nabla p(\mathbf{w}_{\mathrm{d}}^{*})^{T}(\mathbf{w} - \mathbf{w}_{\mathrm{d}}^{*}) \geq 0$$

Timing feasible region is contained in a simpler region:

$$\{\mathbf{w} \mid p_{\mathrm{d}}(\mathbf{w}) \geq p_{\mathrm{d}}(\mathbf{w}_{\mathrm{d}}^{*})\} \subseteq \{\mathbf{w} \mid 0 \leq \nabla p(\mathbf{w}_{\mathrm{d}}^{*})^{T}(\mathbf{w} - \mathbf{w}_{\mathrm{d}}^{*})\}$$

Calculating bounds: Step 2

<u>Optimize over the simpler region</u>

- **Using non-linear programming to solve:**

$$\begin{aligned} \text{minimize} \quad & p_{\text{statistical}}(\mathbf{w}) \\ \text{subject to} \quad & 0 \le \nabla p(\mathbf{w}_{\text{d}}^{*})^{T}(\mathbf{w} - \mathbf{w}_{\text{d}}^{*}) \\ & \mathbf{w}_{\min} \le \mathbf{w} \le \mathbf{w}_{\max} \end{aligned}$$

Statistical power

Simpler region
(contains timing feasible region)

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

Calculating bounds: Step 3

Create bound

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

$$\underbrace{p_{\text{statistical}}(\mathbf{w}') \le p_{\text{statistical}}(\mathbf{w}^*_{\text{statistical}}) \quad \left(\le p_{\text{statistical}}(\mathbf{w}^*_{\text{deterministic}})\right)}_{\text{suboptimality gap}}$$

- Bound the suboptimality gap using the percentage:

$$\delta_{\text{so}} = \frac{p_{\text{statistical}}(\mathbf{w}^*_{\text{deterministic}}) - p_{\text{statistical}}(\mathbf{w}')}{p_{\text{statistical}}(\mathbf{w}^*_{\text{deterministic}})}$$

Bounds for Benchmarks: Example

Worst Case Sub-optimalities:

Mean Power δ_{so} : .24% (= 13.24-13.21) / 13.24

Mean + 3 Sigma δ_{so} : 2.5% (= 17.35-16.91) / 17.35

Deter… Po… Optin…

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

	Mean Power					Mean + 3 Sigma Power				
	v1 (fastest)	v2	v3	v4 (slowest)	avg	v1 (fastest)	v2	v3	v4 (slowest)	avg
c432		0.3%	0.4%		0.3%		5.3%	7.7%		6.0%
c499		0.2%	0.1%		0.1%		1.7%	1.0%		1.5%
c880	0.2%	0.2%	0.2%	0.2%	0.2%	1.7%	2.5%	3.0%	2.6%	2.5%
c1355	[illegible]	[illegible]	0.2%	0.1%	0.2%	2.1%	1.7%	1.2%	1.2%	1.5%
c1908			0.2%	0.2%	0.2%	2.0%	3.9%	4.1%	4.3%	3.6%
c2670	0.2%	0.2%	0.2%	0.2%	0.2%	2.8%	2.1%	2.0%	2.0%	2.2%
c3540	0.2%	0.2%	0.2%	0.2%	0.2%	1.2%	1.7%	2.6%	2.6%	2.0%
c5315	0.2%	0.2%	0.2%	0.2%	0.2%	2.7%	2.7%	2.5%	2.5%	2.6%
c6288	0.2%	0.2%	0.2%	0.2%	0.2%	2.0%	1.5%	1.8%	1.1%	1.6%
c7552	0.2%	0.2%	0.2%	0.2%	0.2%	2.2%	1.2%	1.7%	1.1%	1.5%
alu				0.2%	0.2%	1.9%	2.7%	2.5%	1.2%	2.1%

ISCAS '85 (c432–c7552)

Open Cores ALU (alu)

(Upper bounds on the improvement from using Statistical Power Optimization)

Sub-optimality results: Total power optimization

ISCAS '85 benchmarks and ALU circuit

	Mean Power			Mean + 3 Sigma Power		
switching probability	*minimum*	*maximum*	*average*	*minimum*	*maximum*	*average*
1%	**~0%**	**0.003%**	**~0%**	**~0%**	**0.036%**	**0.006%**
0.10%	**0.001%**	**0.004%**	**0.002%**	**0.005%**	**0.055%**	**0.021%**

- **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

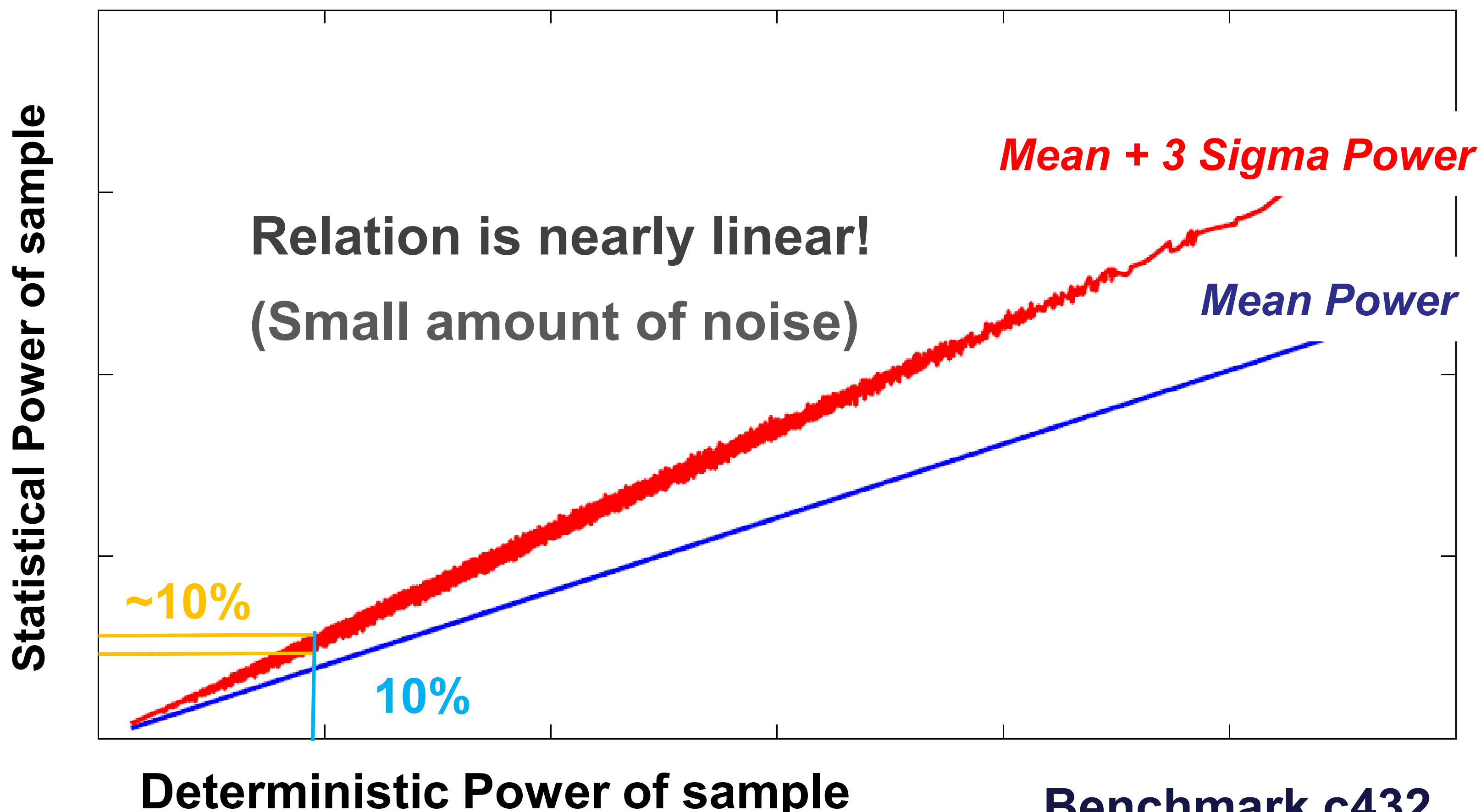
- **Suppose the deterministic solution is within the top 5% of all deterministic sizings**
- **Will this also be in the top 5% of all statistical solutions?**

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)



Benchmark c432

Quantifying the correlation

- **The deterministic and statistical powers are nearly linear relations:**

$$p_{\text{statistical}}(w, l_{eff}, V_t) = \boxed{\left(\alpha\, p_{\text{deterministic}}(w, l_{eff}, V_t) + \beta\right)} + \text{error}(w,\ l_{eff}, V_t)$$

linear

- **Error statistics:**

Leakage Power

	Mean Power			Mean + 3 Sigma Power		
variables	min	max	avg	min	max	avg
w	.004%	.03%	.01%	.07%	.65%	.19%
w, vt	.009%	.08%	.02%	.15%	1.8%	.46%
w, vt, l	.016%	.14%	.04%	.36%	3.5%	.86%

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