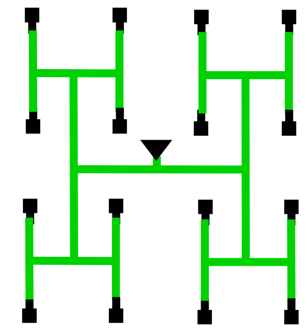
Clock Tree Synthesis with Data-Path Sensitivity Matching

Prof. Matthew R. Guthaus, UC Santa Cruz
Prof. Dennis Sylvester, University of Michigan
Prof. Richard B. Brown, University of Utah

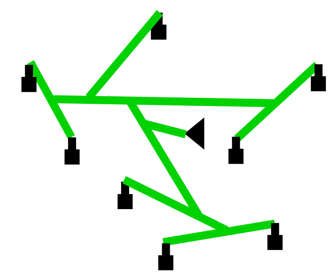
UC SANTA CRUZ
JACK BASKIN SCHOOL OF ENGINEERING
BIOTECHNOLOGY, INFORMATION TECHNOLOGY, NANOTECHNOLOGY

Clock Distribution Networks

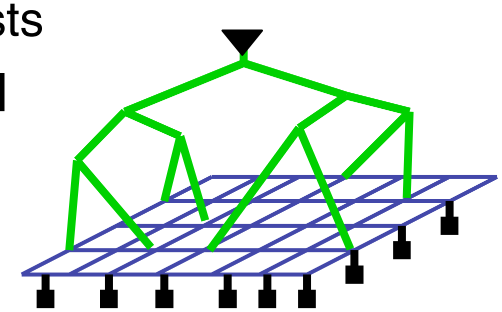
- Uniform (H-Tree)
  - Moderate power consumption
  - Fairly robust
  - Sinks are not usually uniform



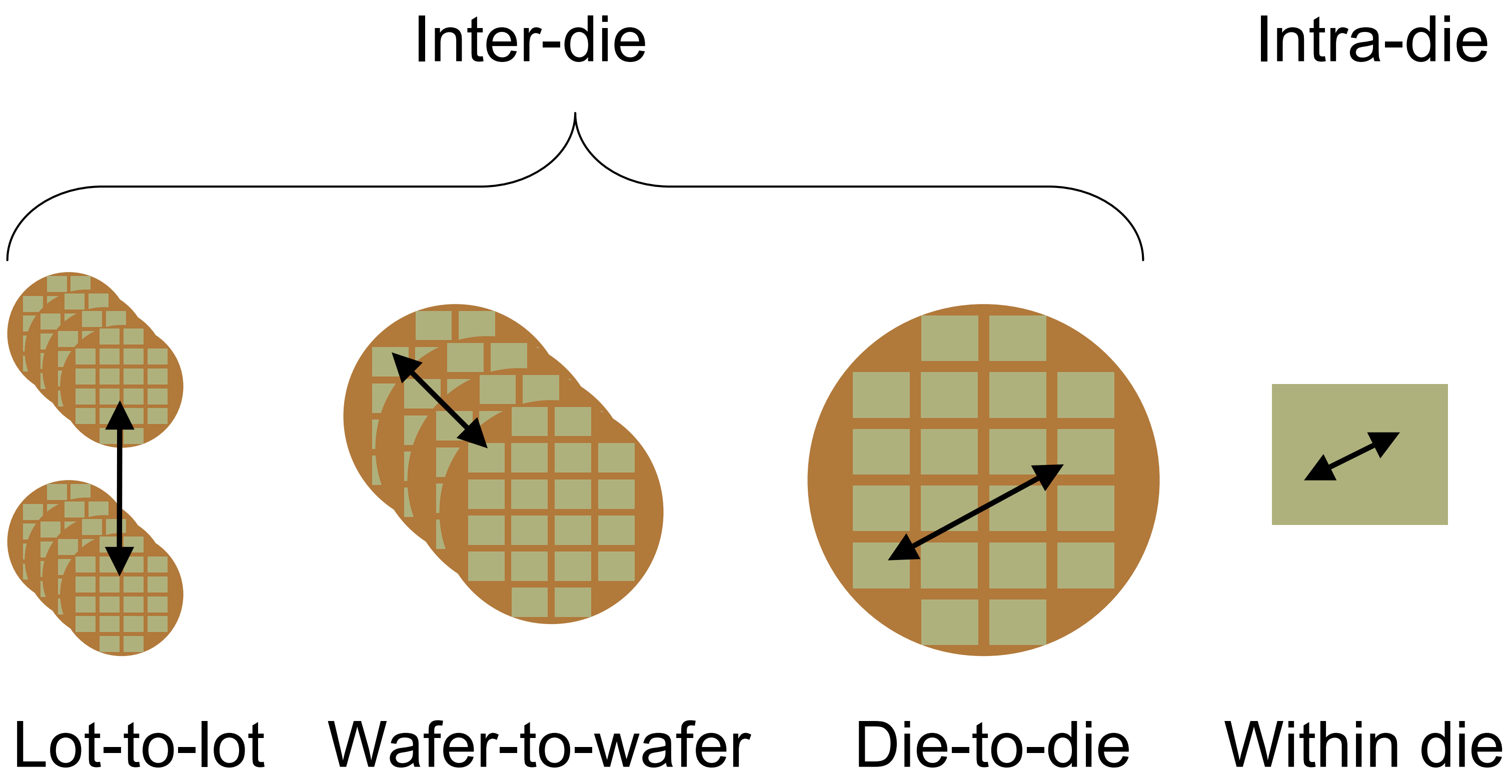
- Balanced Tree [Tsay ICCAD‘91, Chao et al. DAC‘92, Boese et al ASIC‘92]
  - Minimum wire length
  - Sensitive to process parameters



- Spines [Tam et al ISSCC’06]
  - Used by Intel (P6, Xeon MP)
  - Variations within and between spines still exists
- Grids [Anderson et al ISSCC’06, Golden et al ISSCC’06]
  - Used by IBM (Power4) and AMD (Hammer)
  - Low variation, but huge power overhead

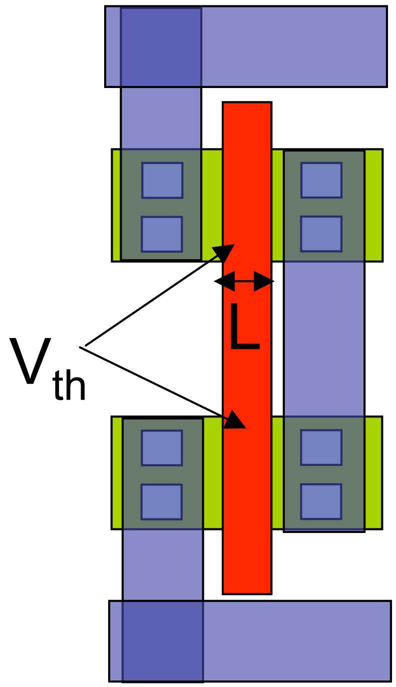


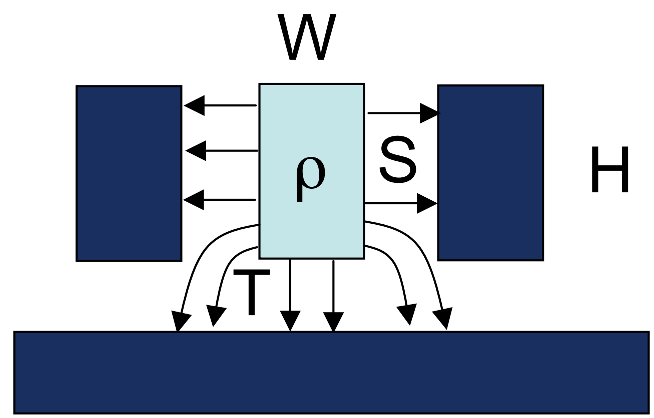
Types of Variation



- Environmental (temperature, voltage, etc.)
- Physical (lithography, materials, etc.)
- Fatigue (NBTI, metal migration, etc.)

Variation Source Assumptions



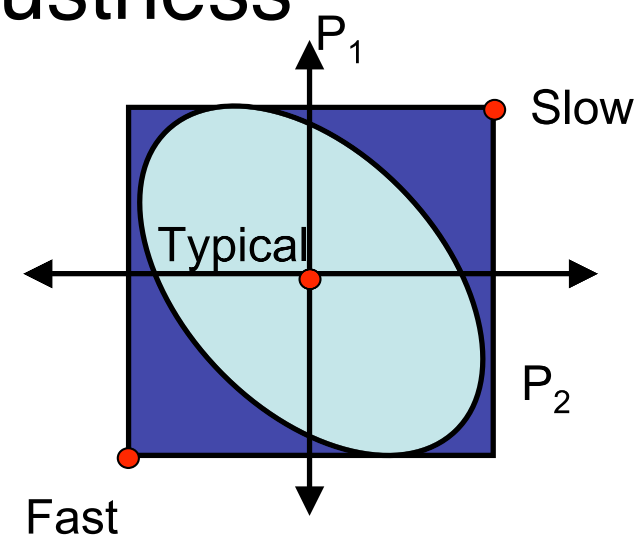


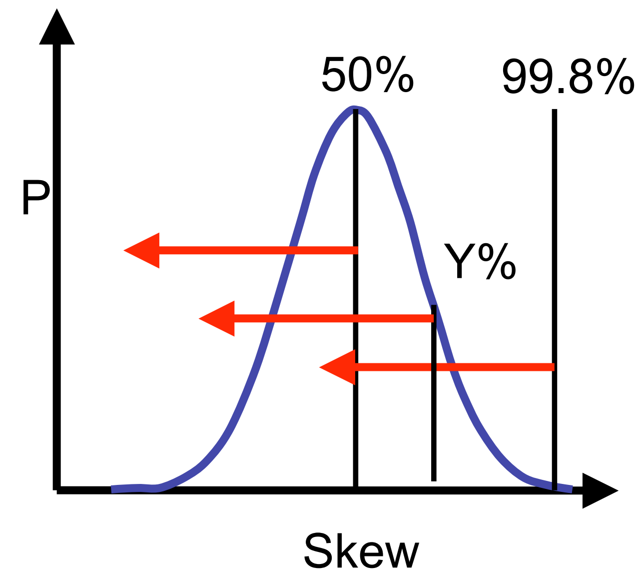
25% Corr.	L_{eff}= 53nm±16.7%
Indep.	$V_{thp=}$ 0.232±30% V_{thn}= -0.273±30%

Corr.	W =175nm±32nm S = 175nm±32nm H = 280nm±15% T = 280nm±10% ρ = 2.2e-8±30%

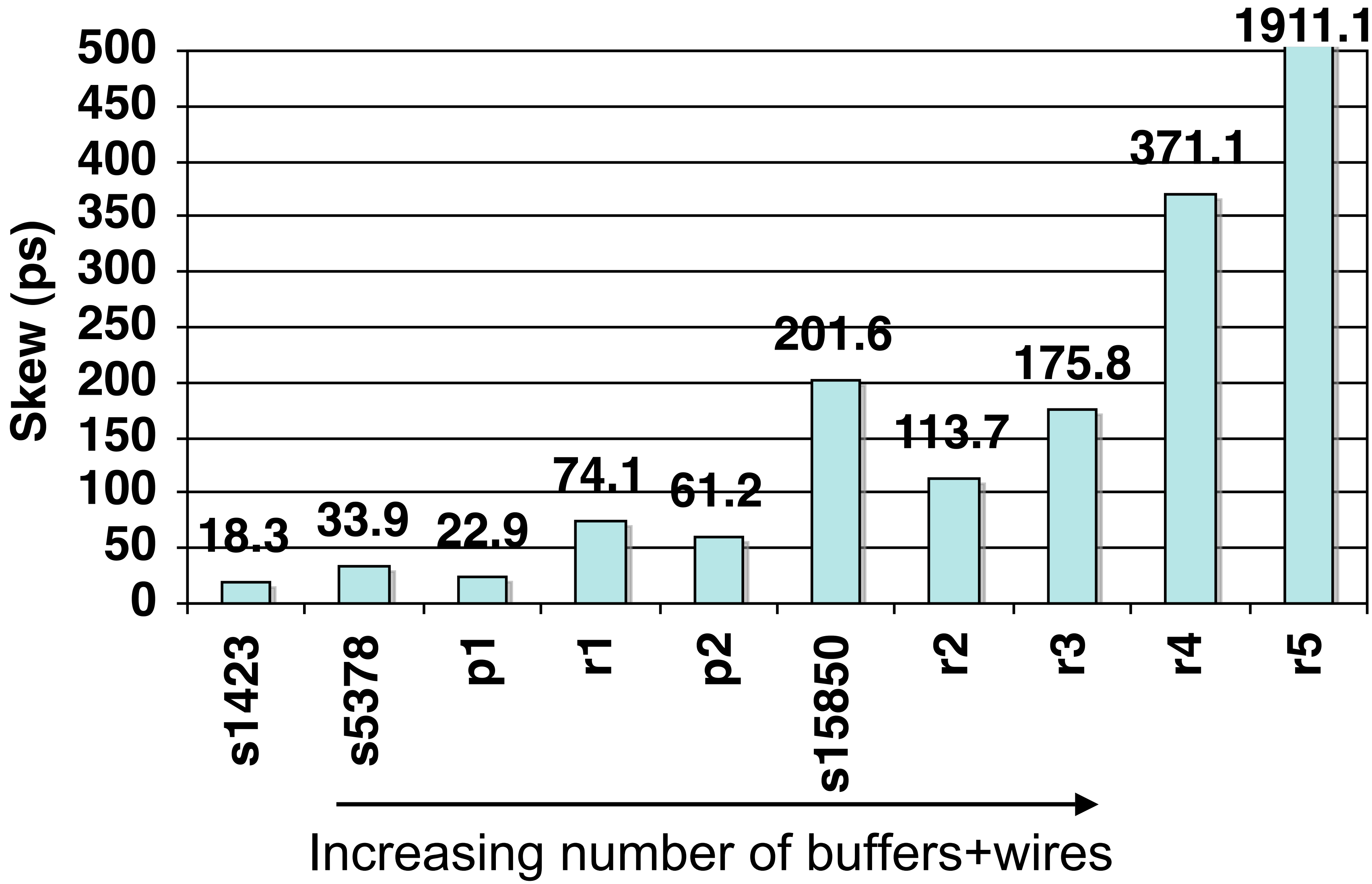
Improving Robustness

- Variation is a major concern in clock distribution
- Current Options
  - Corner-based optimization
    - Process-Voltage-Temp (PVT)
    - Risky, Pessimistic, etc.
  - Direct statistical optimization
    - Many simplifications or expensive to compute
- Can heuristics still help clock tree optimization?



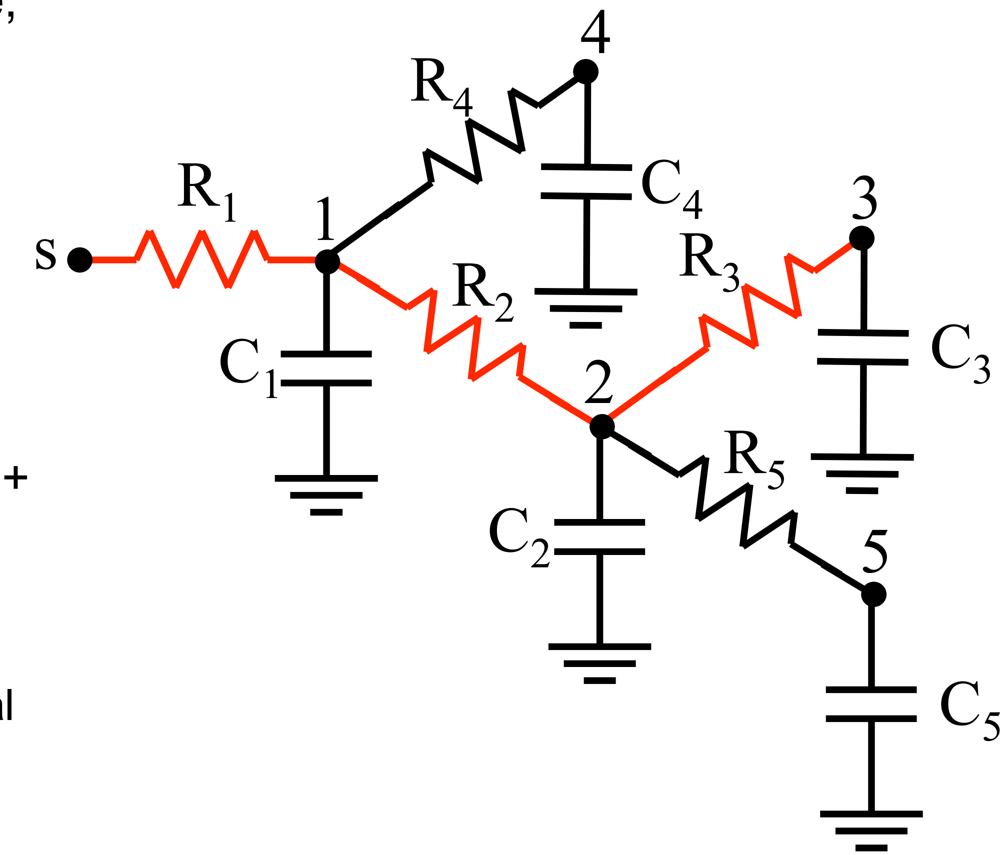


Expected Skew of “Zero Skew” Trees



Skew Calculation

- Elmore delay is not accurate, but high fidelity.
- Fast for optimization.
- S2M for slew calculation.
- Operations Required
  - Add/Subtract
  - Mult
  - Maximum/Minimum
- Delay(s,3)=0.69*(R1(C1+C2+C3+C4+C5)+R2(C2+C3+C5)+R3C3)
- Skew
  - Maximum Difference (global skew)
  - Maximum Path-Connected Difference (local skew)



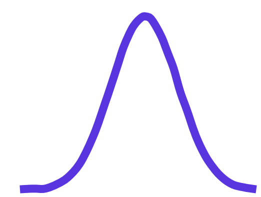
[Elmore, J. App. Physics 1948]

[Agarwal et al., TCAD 2004]

Parameterized Form

Nominal Correlated Independent

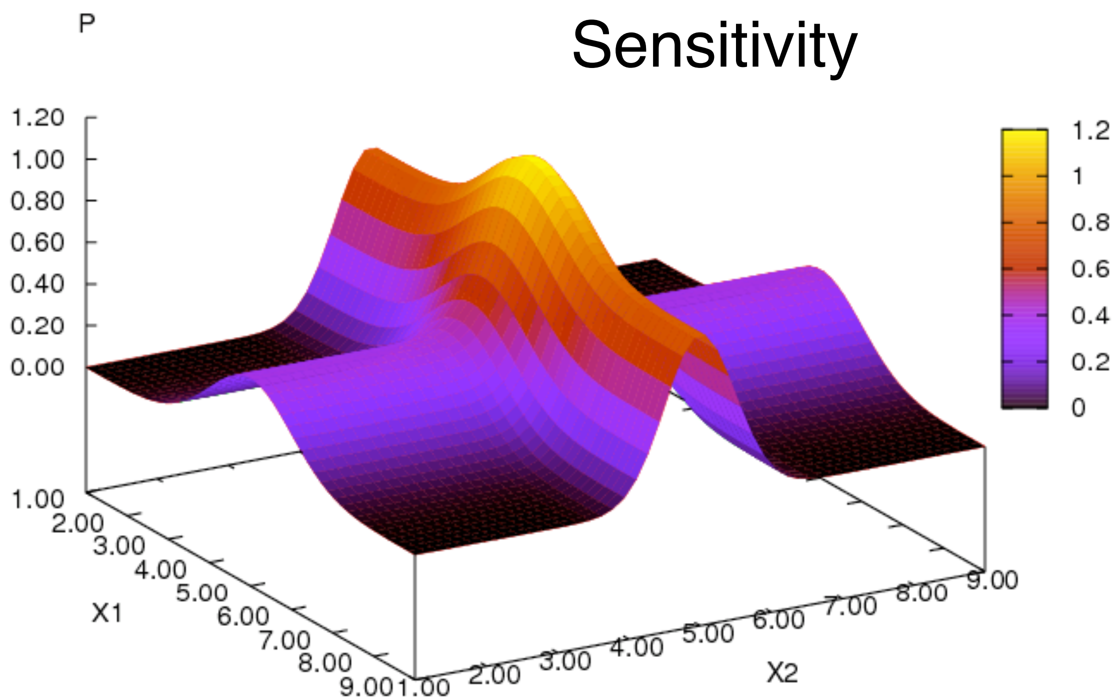
$$d = d_0 + \sum_i d_i X_i + d_r X_r$$



Sensitivity

Gaussian Random Variables

$N(\mu,\sigma) = N(0,1)$



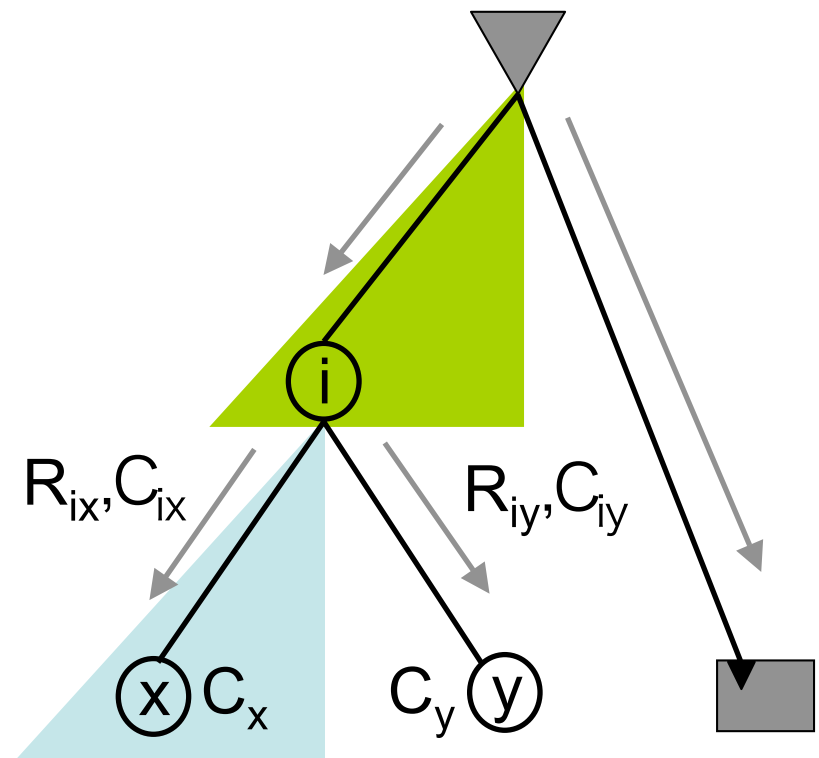
Parametric Operations

- Addition and Subtraction
  - Add/Sub the means
  - Correlated: Add/sub std. dev.
  - Independent: Root-sum-square std. dev.
- Multiplication
  - Many non-linear cross terms
  - Showed that approximating cross-terms as random variation works well
- Maximum and Minimum
  - First and second moments calculated analytically [Clark 1961,Cain 1994]
  - Sensitivities approximated by proportional weight [Visweswariah et al., DAC 2004]

Top-Down Statistical Analysis

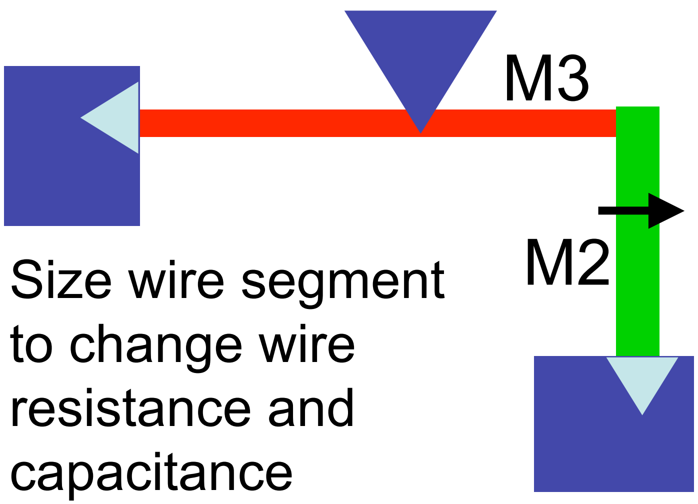
- Parameterized R, C, and D values.
- First bottom-up propagate total sub-tree capacitances, C_i
- Top-down propagate parameterized delays, D_i
- Skew is Max(D_i -D_j) for sinks i and j

$$D_x = D_i + R_{ix}\left(\frac{C_{ix}}{2} + C_x\right)$$

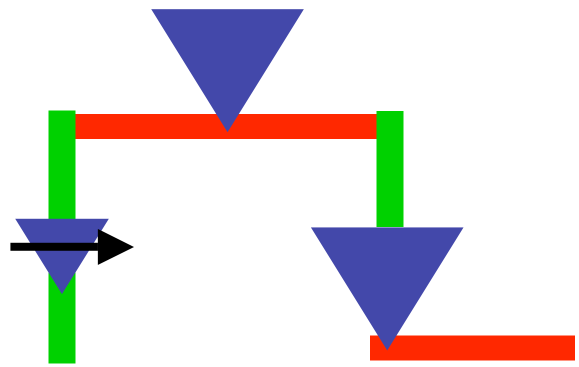


Clock Tree Tuning

- Start with DME + Buffered tree
- "In Place" Optimization
- Select Buffer Sizes and Wire Widths to Minimize Skew while Increasing Robustness
- Buffer/Wire Sizes
  - Two stage buffer with fixed internal gain
  - Continuous range of buffer output sizes
  - Continuous range of wire widths
  - Minimum and maximum limits for both sizes



Size wire segment to change wire resistance and capacitance



Size buffer to change load and drive strength

Sequential LP for Clock Skew

Minimum Skew Objective	$\min \quad s_{max} - s_{min}$
	$\text{s.t.} \quad s_i - s_{min} \geq 0, \forall i \in \text{Sinks}$
	$s_{max} - s_i \geq 0, \forall i \in \text{Sinks}$
Linear Delay Constraints	$D + G\Delta = S$
Power Bound	$P_{cur} + \beta\Delta \leq P_{max}$
Simple Bounds	$\max(L_i, x_i - \epsilon_i) \leq x_i + \delta_i \leq \min(U_i, x_i + \epsilon_i)$

Similar to Wang and Marek-Sadowska, DAC 2004, but for skew rather than power minimization.

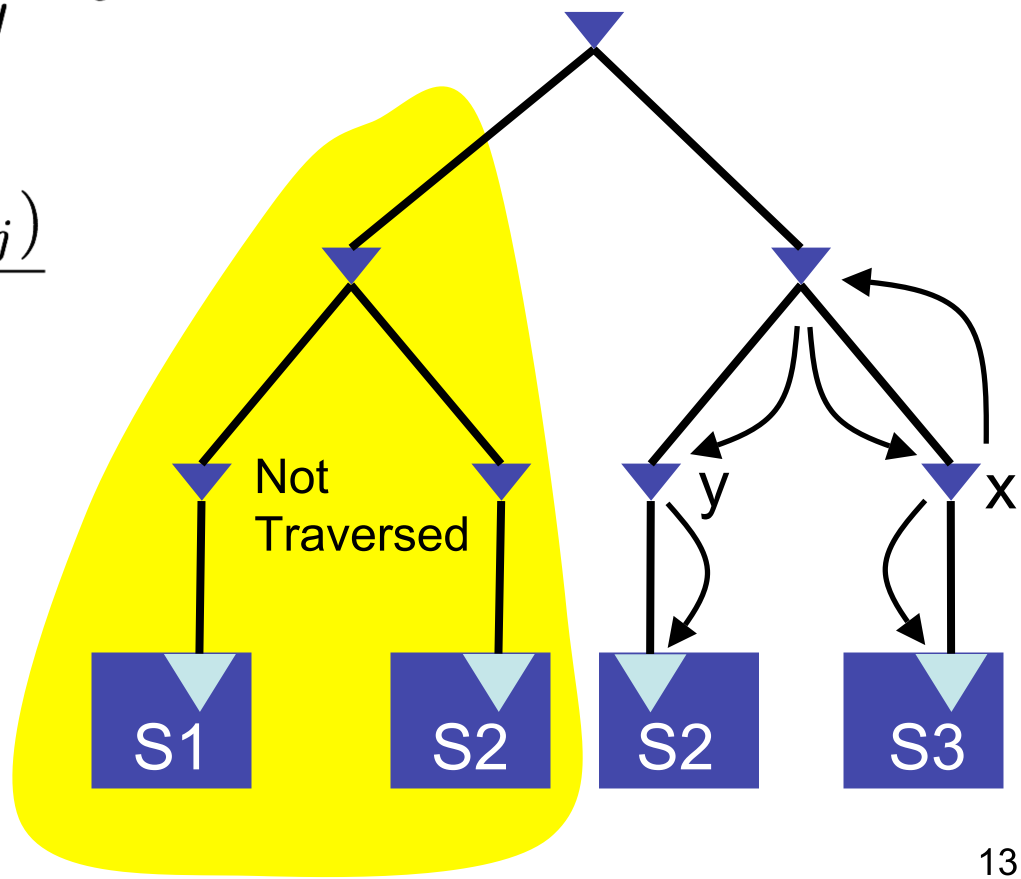
Linear Delay Constraints

Buffer/Wire Size Changes

$$s_3 = d_3 + \frac{\partial d_3}{\partial x}\Delta x + \frac{\partial d_3}{\partial y}\Delta y + \ldots$$

$$\frac{\partial d_i}{\partial x_j} \approx \frac{d_i(x_j+\epsilon) - d_i(x_j)}{\epsilon}$$

- Perturb & Difference can be used with any analysis
- More buffers provides better incremental analysis



Sequential Quadratic Formulation

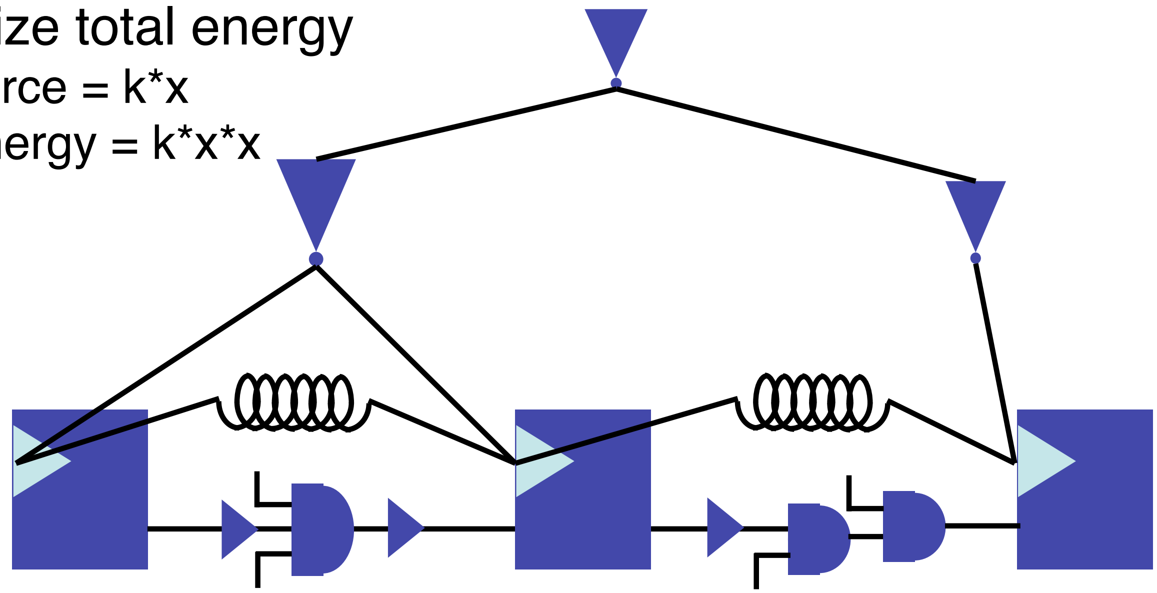
- We are NOT approximating skew or constraints with a second-order function
  - Indirect optimization
  - Convex cost function
- Minimize total energy
  - Force = k*x
  - Energy = k*x*x

$$\Phi(S) = \sum_{i>j} w_{ij}(s_i + b_{ij} - s_j)^2$$

Weight

Useful Skew

Sink Delays



[Guthaus et al., DAC 2006.]

Additional Constraints

- Power Bound
  - Dominated by dynamic power so capacitance rather than true power is bounded
  - Constraint ensures total size changes are still below power limit

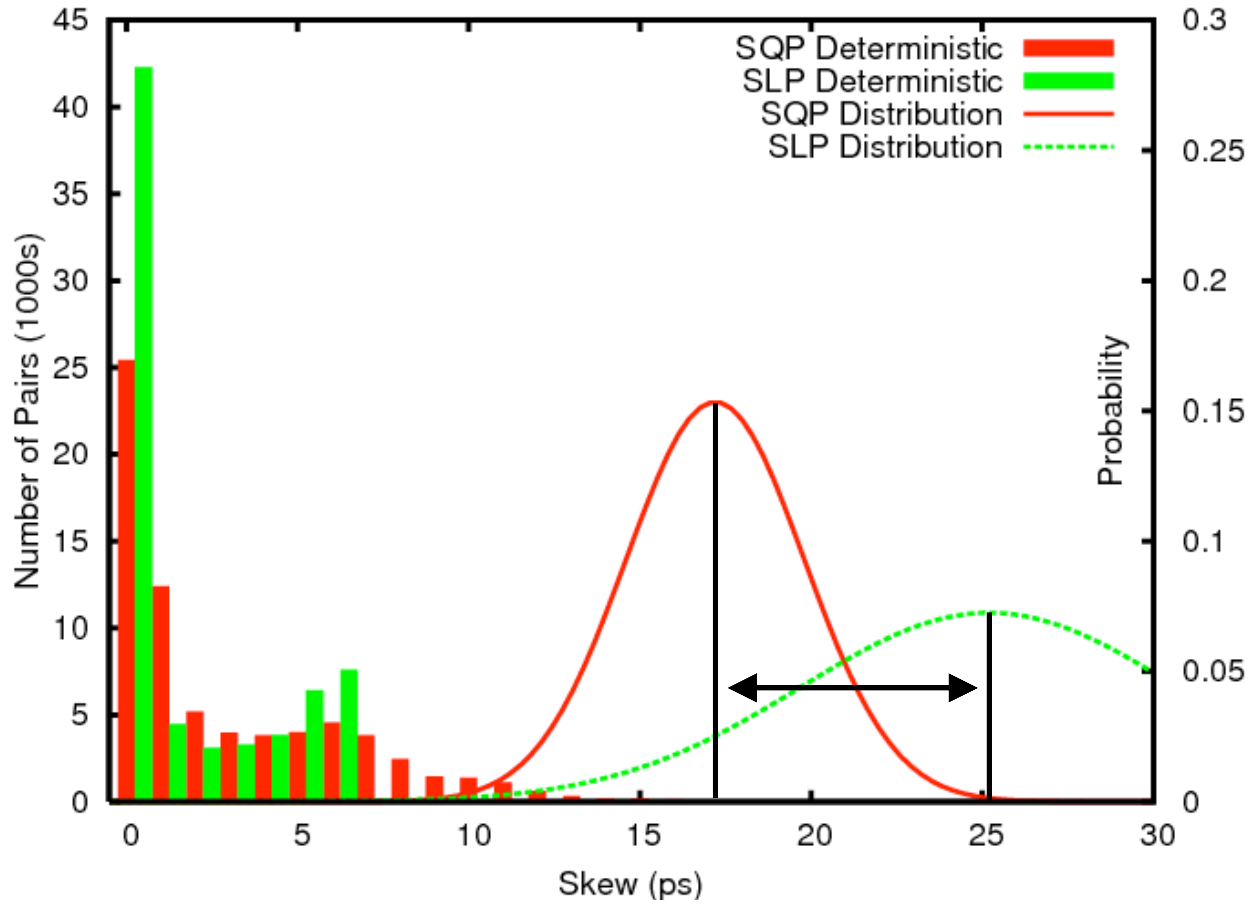
$$\sum_{x \in \{Gates, Wires\}} P(x + \Delta x) \leq P_{max}$$

- Simple Bounds
  - Linearity of sink delay is only valid in a small range so we restrict the size changes by epsilon
  - Technology places hard upper/lower limits on buffer and wire sizes

$$\max(L_i, x_i - \epsilon_i) \leq x_i + \delta_i \leq \min(U_i, x_i + \epsilon_i)$$

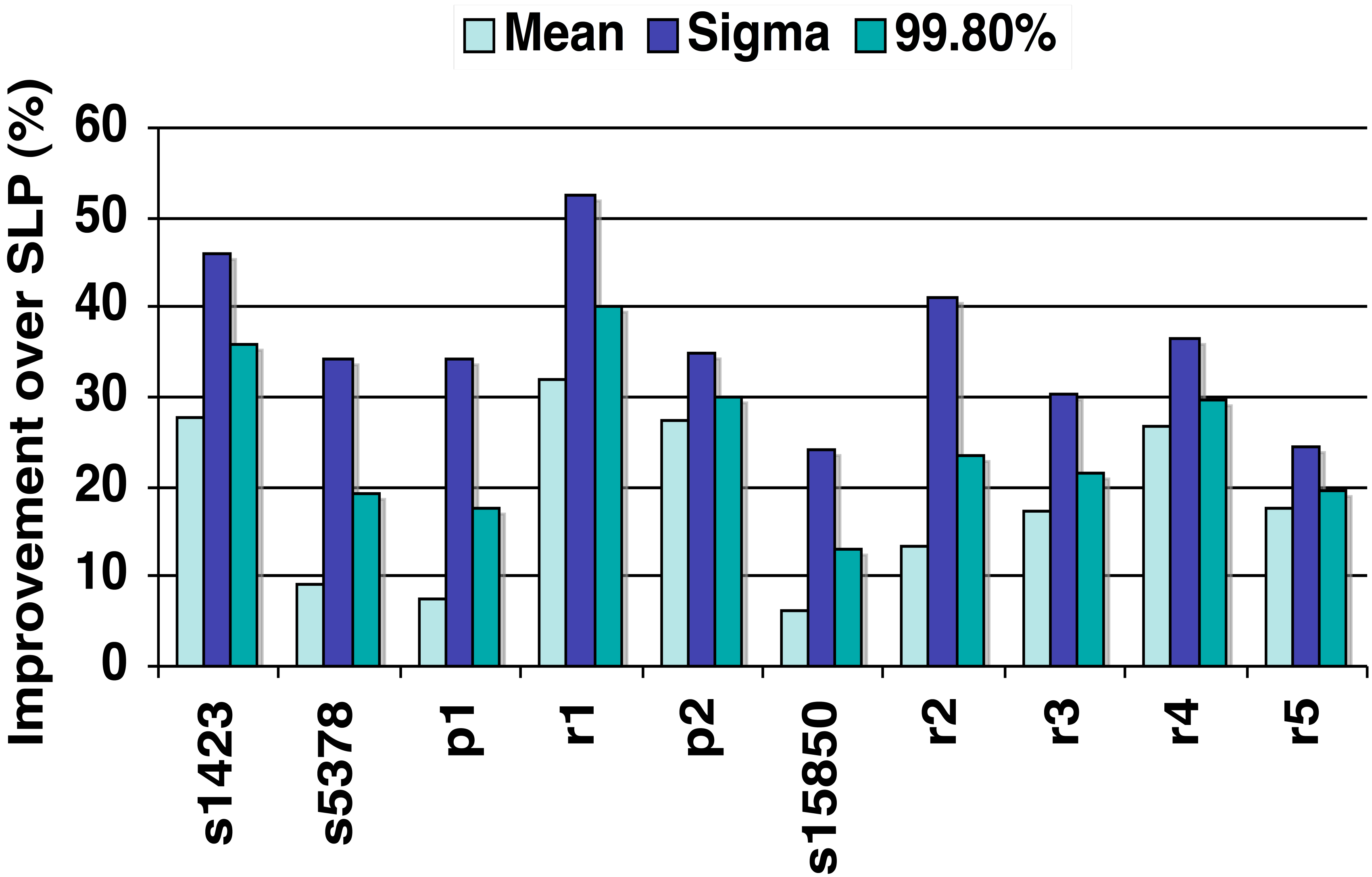
R1 Linear vs Quadratic "Push Out"

- Maximum Skew
  - SLP: 7ps
  - SQP: 15ps
- Pairs within 1ps of crititical
  - SLP: >7,000 pairs
  - SQP: 12 pairs
- Mean push out
  - Almost 8ps



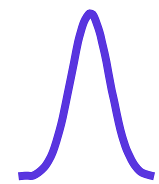
[Guthaus et al., DAC 2006.]

SLP vs. SQP Skew (50% Cap. Increase)



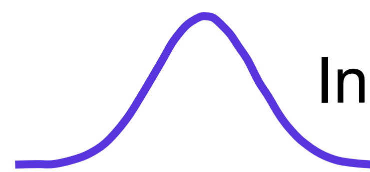
Why preserve sensitivities?

- Sensitivities attribute variability to a particular source
- Underlying sources of variation are defined as "correlated"
- Correlated sensitivities can "cancel out" whereas independent sensitivities accumulate as root-sum-of-squares



Correlated

$$N(\mu_1 - \mu_2, \sigma_1 - \sigma_2)$$



Independent

$$N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

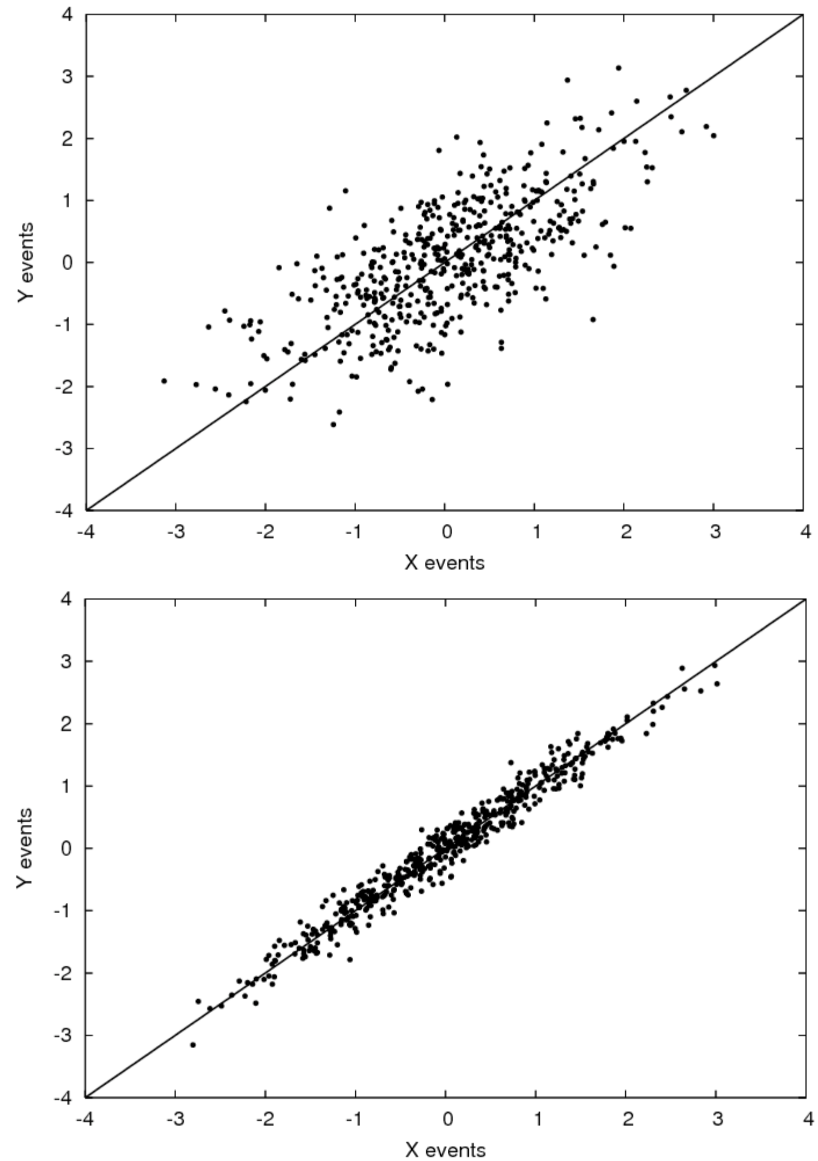
Correlation Definitions

- Defines tendency for events to track
- Formalized with the Pearson correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

- Can also be defined geometrically as cosine of angle between two event vectors

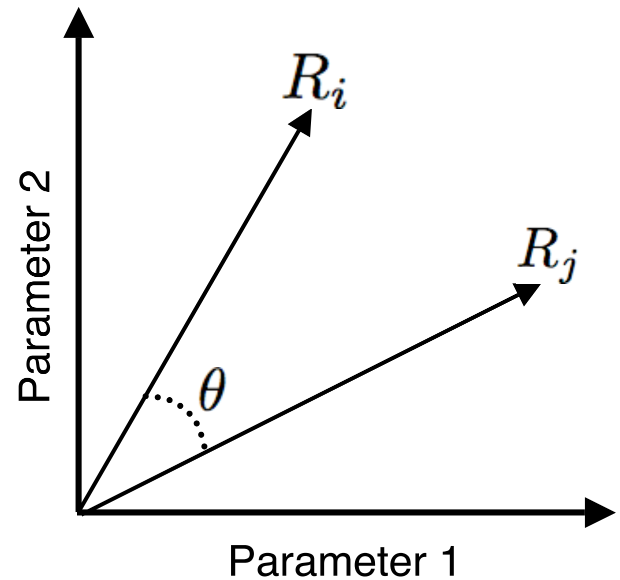
$$\cos(\theta) = \frac{X \cdot Y}{|X||Y|}$$



Geometric Interpretation of Correlation

- Parameterized form is already centered
- Sensitivity coefficients are linear
- Define the sensitivity vector, R:

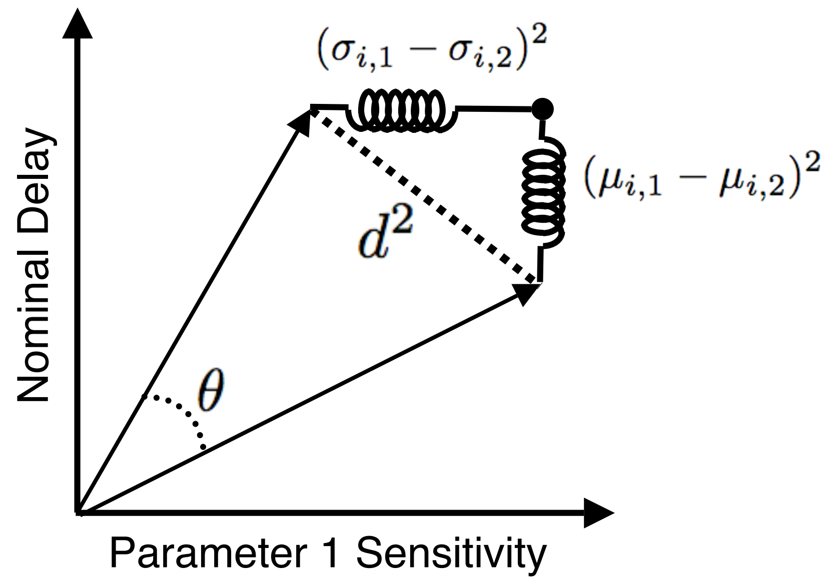
$$A = a_0 + \sum_{i=1}^{k} a_i X_i$$
$$= a_0 + R^T X$$



$$cos(\theta) = \frac{R_i \cdot R_j}{|R_i||R_j|}$$

Heuristic for Increasing Correlation

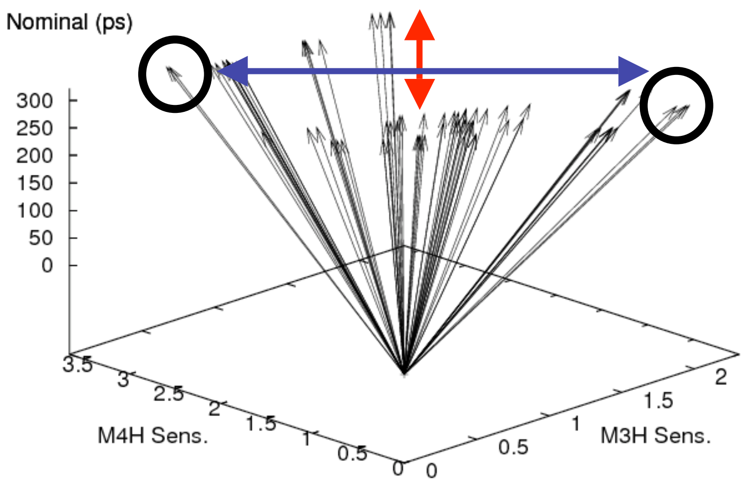
- Include Nominal as a parameter
- Angle is *approximately* proportional to distance squared
- Maximizing correlation, cos(theta), is same as minimizing angle
- SQP can be used for squared objectives



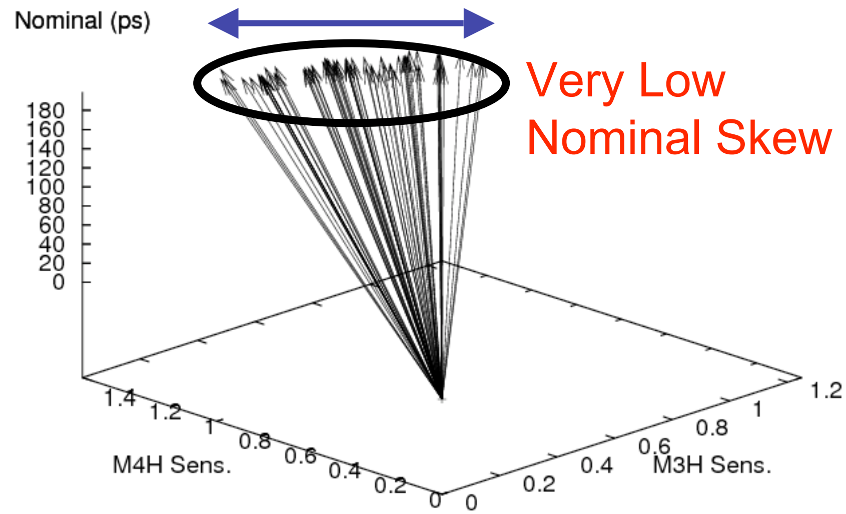
$$\Phi(S) = \sum_{i,j \in P} |S_i - S_j|^2$$

$$|S_i - S_j|^2 = (s_{i,0} - s_{j,0})^2 + \sum_k (s_{i,k} - s_{j,k})^2 + s_{i,r}^2 + s_{j,r}^2$$

Improvement of s1423

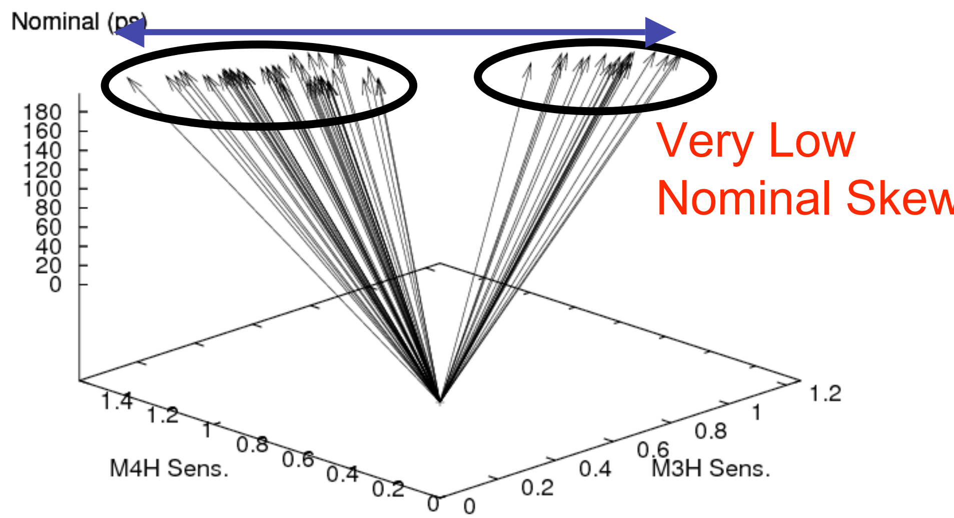


Pre-Tuning



Statistical Tuning

~5% improvement in 99.8% quantile

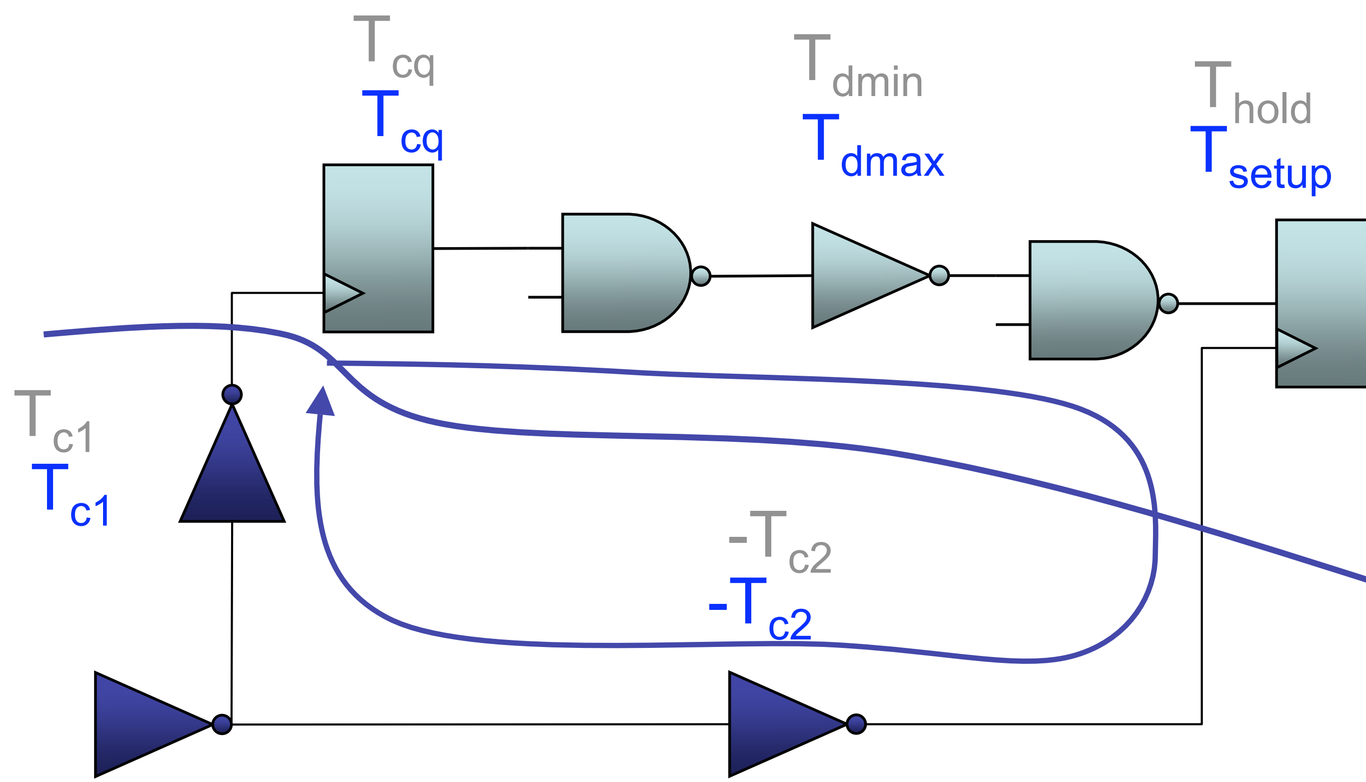


Deterministic Tuning

Infeasible Improvement

- Sometimes improvement is infeasible
  - Wire assignment is fixed
  - Contradiction of forces can result in zero improvement
  - Mutually exclusive sensitivities can result in zero improvement
- No improvement for other benchmarks
  - Same results as deterministic SQP heuristic
  - But still better than SLP
- Does this mean the idea is bad? No.
  - Consider local, not global, skew with data-path sensitivities.

Timing Constraints Revisited



Setup Constraint

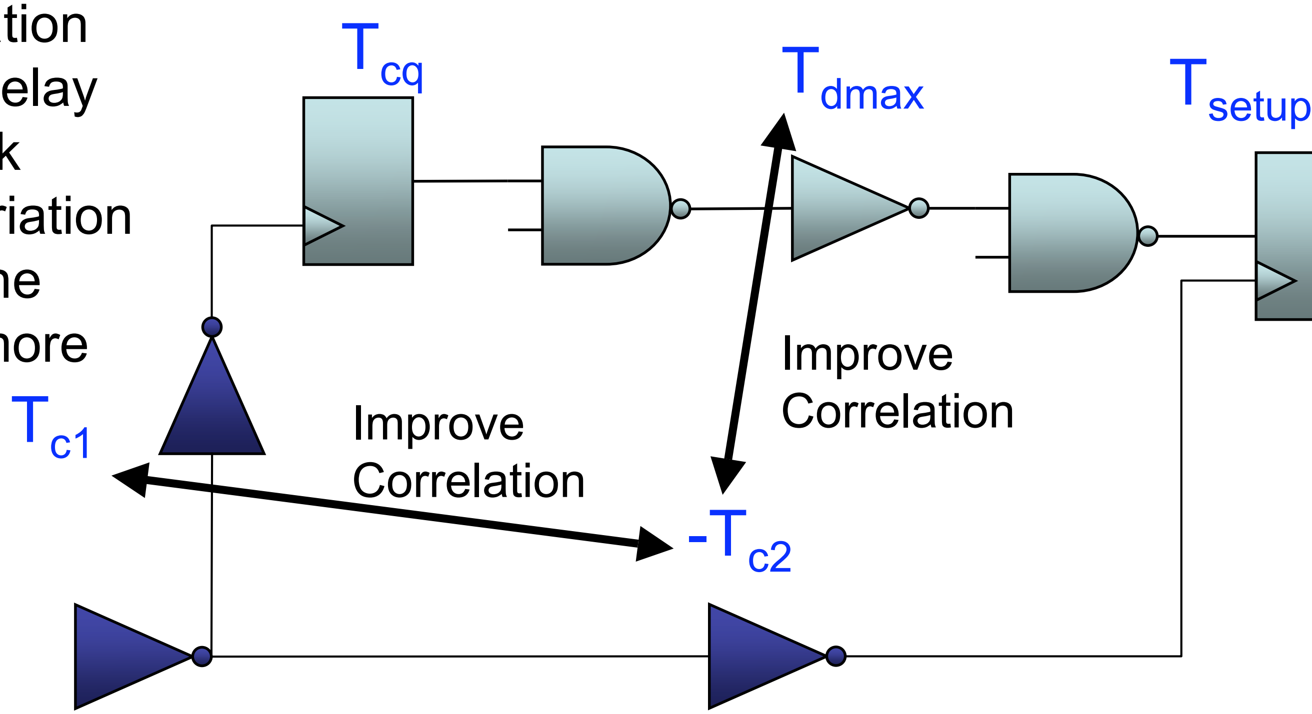
$T_{cq}+T_{dmax}+T_{setup}+T_{c1}-T_{c2}< P$

Hold Constraint

$T_{cq}+ T_{dmin}+T_{hold}+T_{c1}-T_{c2}>0$

Beyond Useful Skew: Useful Variation

Cancellation of max delay and clock skew variation makes the design more robust.

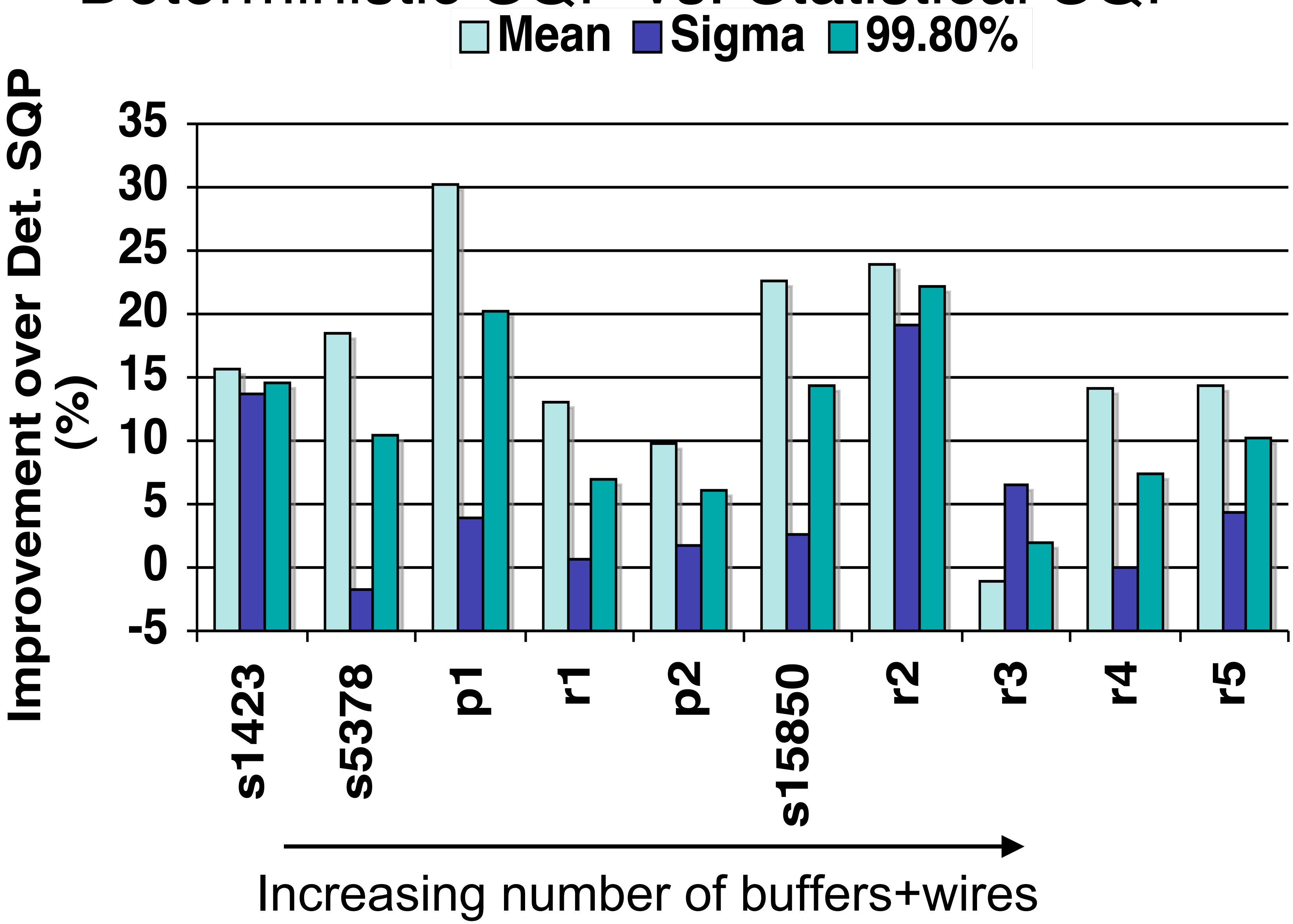


Setup Constraint

$T_{cq}+(T_{dmax}-T_{c2}+T_{c1})+T_{setup} < P$

$$\Phi(S) = \sum_{i,j \in P} |S_i - B_{i,j} + S_j|^2$$

Deterministic SQP vs. Statistical SQP



Run-Time Costs

- Up to 50x the run-time due to naïve gradient computation
- Evaluation of 12 random variables
- Performed all optimization using new method
- Can be used for “fine tuning” after deterministic optimization instead

Conclusions

- New technique for improved correlation
  - Uses distance between canonical vector delay representations
  - Matches nominal delay
  - Matches first order sensitivities
  - Minimizes uncorrelated sensitivity
- Data-path variation awareness
- Average of 16.3% better expected skew
- Average of 11.9% improved mean + 3-sigma

Thank you!