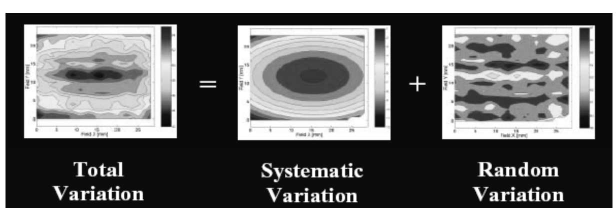
Statistical Timing Analysis

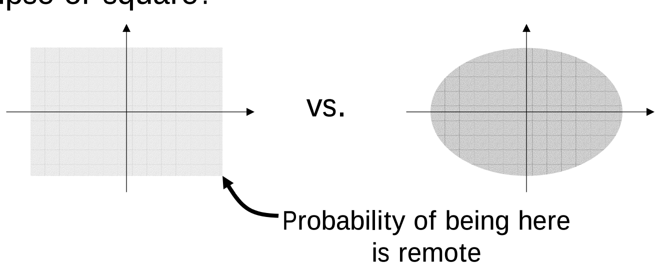
Sachin S. Sapatnekar
ECE Department
University of Minnesota
sachin@ece.umn.edu

Corner-based vs. statistical analysis



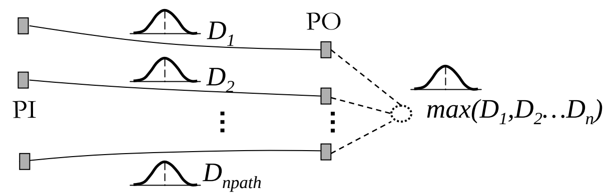
[Cain/Friedberg, UC Berkeley]

- Ellipse or square?



Statistical static timing analysis

- Intra-die variations in addition to inter-die variations
- Deterministic timing analysis ⇨ Statistical timing analysis
- Path-based analysis: find variability along a single path



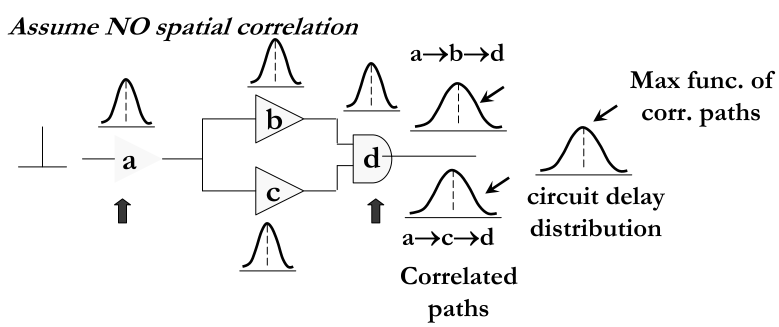
- Block-based analysis: Find the distribution (PDF/CDF) of:

$$D_{max} = \max(D_1, D_2, \ldots, D_{npaths})$$

D_i: *distribution of* i^{th} *path delay*

Difficulties in statistical timing analysis

- Path correlation due to reconvergent fanouts



- Spatial correlation makes the path correlation structure more complicated

Problem statement

- Find the PDF/CDF for circuit delay distribution:

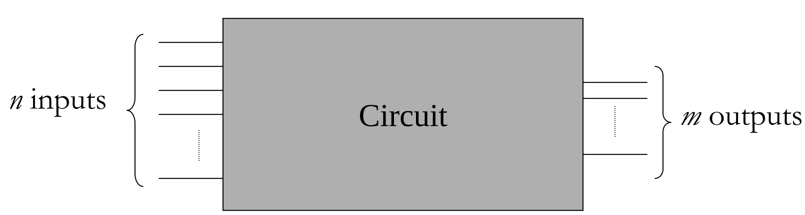
$$D_{max} = \max(D_1, D_2, \ldots, D_{npaths})$$

where D_i: delay distribution of i^{th} path in the circuit

- Assume normal distributions on process parameter values
  - Why?
  - Is this reasonable? If not, what is?
- Parameter correlations
  - L_{eff} shows high spatial correlations
  - T_{ox}, N_d are largely uncorrelated

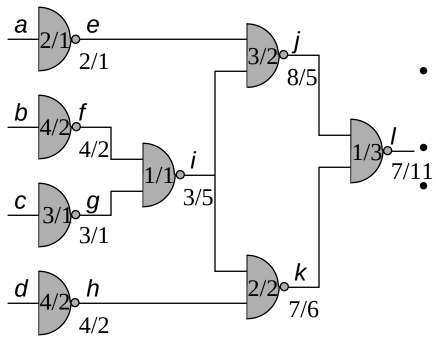
Basics of static timing analysis

- Main features
  - Fast
  - Input pattern-independent
  - Uses coarse delay models that are good for a first order approximation, and for optimization purposes



The critical path method (CPM)

- Often (incorrectly) called "PERT" in the literature



- Place PI's on a queue
- Process gate from head of queue
- Add gate to queue if all input delays known
- Complexity *O(E)*
- Can handle gate delay dependence on order of input arrivals

Interconnect and gate delay models

- Approximate a function by first-order Taylor expansion:

$d = f(\vec{P})$ $\quad \vec{P}$: Set of normally distributed random variables

First-order Taylor expansion →

$$d = \bar{d} + \sum_{\forall \text{ Parameter } P_i} \left[\frac{\partial f}{\partial p_i} \right]_{\overline{p_i}} \Delta p_i$$

Normal Distribution

Nominal Value of d

Sensitivity of f to p_i

$\Delta p_i = p_i - \overline{p_i}$

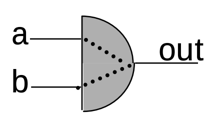
- *d* is a linear function of normally distributed random variables
- Sensitivities of gate/interconnect delay to the process parameters

Simple approaches

- Build in slacks to reduce variation
  - Bai *et al.*, DAC 02
- More exact methods
  - Perform statistical STA

Berkelaar's method

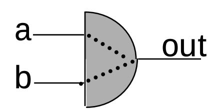
- Types of operations in STA
  - SUM: $T_{a \to out} = T_a + d_{a \to out}$; $T_{b \to out} = T_b + d_{b \to out}$
  - MAX: $T_{out} = \max(T_{a \to out}, T_{b \to out})$



- Gate delay modeled as a Gaussian
  - SUM is easy: sum of Gaussians = Gaussian
  - MAX of Gaussians is not a Gaussian
  - Approach: approximate max by a Gaussian

Tsukiyama's method

- No correlations in Berkelaar's method



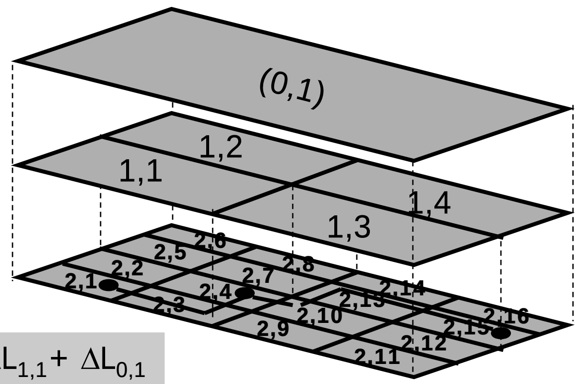
- Assume $d_{a \to out}$, $d_{b \to out}$ to be correlated Gaussians
- Basic operations:
  - SUM: still easy
  - MAX: use Clark's results from 1961 to find mean, variance of the max function

Incorporating spatial correlations

- In reality, there is a strong spatial correlation between process variables
- Orshansky's method (DAC02)
  - Path-based approach
  - Finds path pdf's, covariances, etc.
    - Uncorrelated distribution provides upper bounds on expected value
    - Uses results from probability theory to estimate pdf of the max function

Agarwal's path-based method

- Variational model
  - Regions of variation



$$\Delta L_{g1} = \Delta L_{2,1} + \Delta L_{1,1} + \Delta L_{0,1}$$
$$\Delta L_{g2} = \Delta L_{2,4} + \Delta L_{1,1} + \Delta L_{0,1}$$
$$\Delta L_{g3} = \Delta L_{2,15} + \Delta L_{1,4} + \Delta L_{0,1}$$

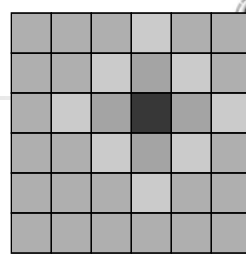
Agarwal's path-based method (contd.)

- For a small change
  - $\Delta D = S_L \Delta L + S_W \Delta W + \ldots$
- Delay variations

$$\Delta D_{g1} = K(\Delta L_{2,1} + \Delta L_{1,1} + \Delta L_{0,1})$$
$$\Delta D_{g2} = K(\Delta L_{2,4} + \Delta L_{1,1} + \Delta L_{0,1})$$
$$\Delta D_{g3} = K(\Delta L_{2,15} + \Delta L_{1,4} + \Delta L_{0,1})$$

- Add up delay variations on each path using the sum operator to get path delay distributions

Chang's method



- Variational model
  - Chip area divided into n squares
  - Nearby squares are correlated
- Basic idea
  - Compute sensitivities of Delay to process parameters

$$\Delta D = \Sigma_k S_{Lk}\Delta L_k + \Sigma_k S_{Wk}\Delta W_k + \ldots$$

  - Use principal component method to orthogonalize all variations

$$\Delta D = \Sigma S_i \Delta p_i$$

    - Calculating $\mathrm{cov}(d_i,d_j)$ is easy!
    - Used to calculate approximation of max function
    - Sum function is easy, as before
  - PERT-like traversal: complexity = n × STA complexity

Orthogonal transformation

- Principal Component Analysis (PCA):

Correlated (Original Set)	*Transform*	*Uncorrelated (Principal Components)*
$\vec{X} = \{x_1, \ldots, x_n\}$	→	$\vec{X}' = \{x'_1, \ldots, x'_n\}$
$x_i : \mu_i = 0, \sigma_i = 1$		$x'_i : \mu'_i = 0, \sigma'_i = 1$

- Element of original set can be expressed as linear function of the principal components (PC's)

$$\boxed{x_i} = \mu_i + \left(\sum_{j=1}^{k} \sqrt{\lambda_j}\,\alpha_{ij}\boxed{x'_j}\right) \cdot \sigma_i$$

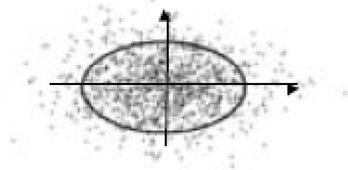
where λ_j : i^{th} eigenvalue of cov. matrix of $\vec{X}$

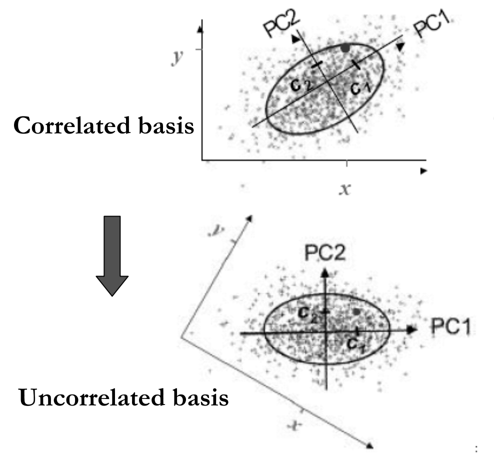
α_{ij} : i^{th} element of i^{th} eigenvector of cov. matrix of $\vec{X}$

Idea of orthogonal transformations

Set of equiprobable points for a 2D Gaussian = ellipse

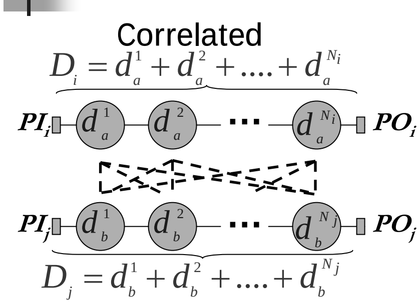
Gaussian with a diagonal covariance matrix

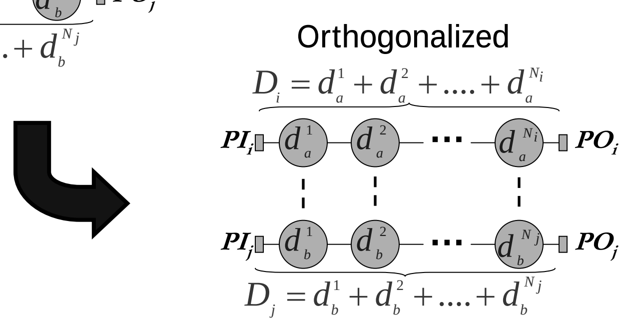




[Adapted from Tim Marks http://cogsci.ucsd.edu/~desa/pca.pdf]

Why?





Experimental results

- Verified with Monte-Carlo simulation: Avg. error **0.2%** for mean, **0.9%** for s.t.d. and 0.2% for 98 percentile point
- Fast run-time: < 200s

Circuit Name	Error% w.r.t. Monte-Carlo		Monte-Carlo	Proposed Method	
	Mean	S.T.D.	CPU (s)	CPU(s)	PCA(s)
s38417	-0.1%	-0.7%	15295	130.32	0.15
s38584	-0.1%	-0.9%	19024	132.08	0.15
s35932	-0.3%	-3.6%	48087	182.31	0.15
s15850	-0.2%	-0.3%	9932	56.00	0.15
s13207	-0.1%	0.8%	5082	50.48	0.15
s9234	-0.1%	0.4%	2952	9.42	0.02
s5378	-0.3%	-0.5%	1531	5.27	0.02
s1196	-0.3%	-0.3%	378	0.41	0.01
s27	-0.1%	-0.3%	67	0.00	0.00

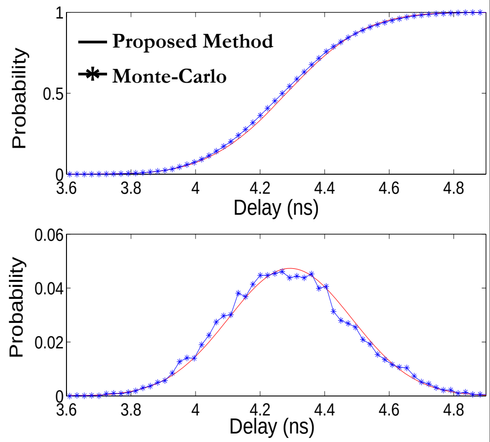
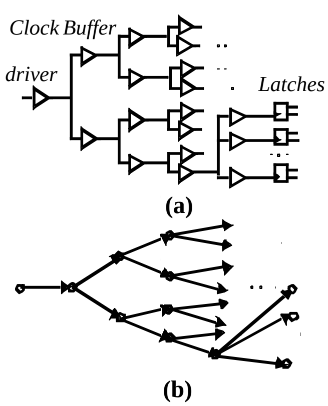


Fig 1: CDF and PDF Curves (ckt s38417)

Variability in the clock network



Routed clock network with tree topology composed of driver gates

- Timing tree
  - Root – clock driver
  - Sink nodes – Latches
  - Edges – gate + interconnect
- For a Deterministic Timing Tree:

$$d_{min} = min\,(d_{p,1}, d_{p,2}, \ldots\ldots, d_{p,n})$$

$$d_{max} = max\,(d_{p,1}, d_{p,2}, \ldots\ldots, d_{p,n})$$

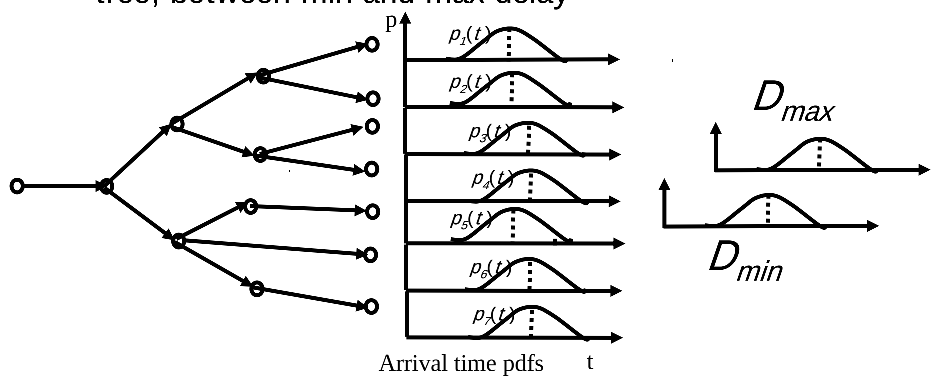
$$skew = d_{max} - d_{min}$$

- Aim : Determine the skew distribution for a Probabilistic Timing Tree (PTT)

[Agarwal, ICCAD03]

Statistical clock skew analysis

- Propagate arrival time distributions to each sink node
- Difficulties: correlations between shared sections of the tree, between min and max delay

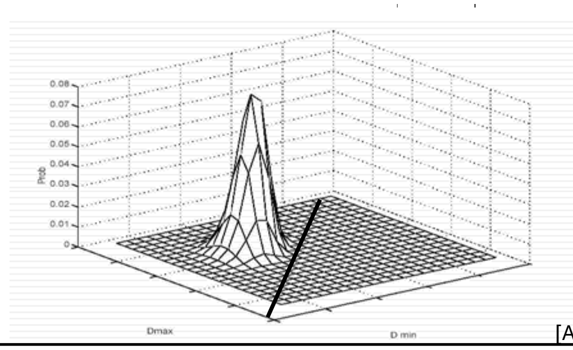


[Agarwal, ICCAD03]

Approach

- Key ideas
  - Compute the JPDF for D_{min} and D_{max}
  - Propagate bottom up
  - Details: see Agarwal *et al.*, ICCAD 2003

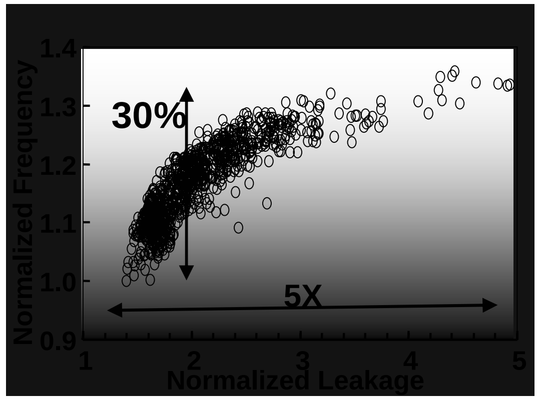
JPDF



[Agarwal, ICCAD03]

Statistical power analysis

- Leakage power
  - Most susceptible to variations
  - Components
    - Subthreshold leakage
    - Gate leakage
  - Problem amounts to summing up a set of correlated lognormals



What next?

- Statistical timing analysis methods will mature before statistical power analysis
  - The latter will be more “useful”
- Methodology issues
  - How does it all “come together” and work?
  - Statistical library characterization
  - Sources of uncertainty
    - Design uncertainty
    - Environmental uncertainty (Vdd/gnd levels, temperature, soft errors)
    - Process uncertainty
- Algorithmically, we still need solutions for
  - Nonlinear variations, non-Gaussian variations, multimodal distributions…

Conclusion

- Current approaches likely to be supplanted by statistical analysis
- Timing and power are both affected
- As a community, we now have an underlying algorithmic basis for this work
- Much more work to come…