Parameterized Block-Based Statistical Timing Analysis with Non-Gaussian Parameters and Nonlinear Delay Functions

H. Chang[1], V. Zolotov[2], S. Narayan[3], C. Visweswariah[2]

[1]Dept. of Computer Science & Engineering, University of Minnesota
[2]IBM T.J. Watson Research Center, Yorktown Heights, NY
[3]IBM Systems & Technology, Essex Junction, VT

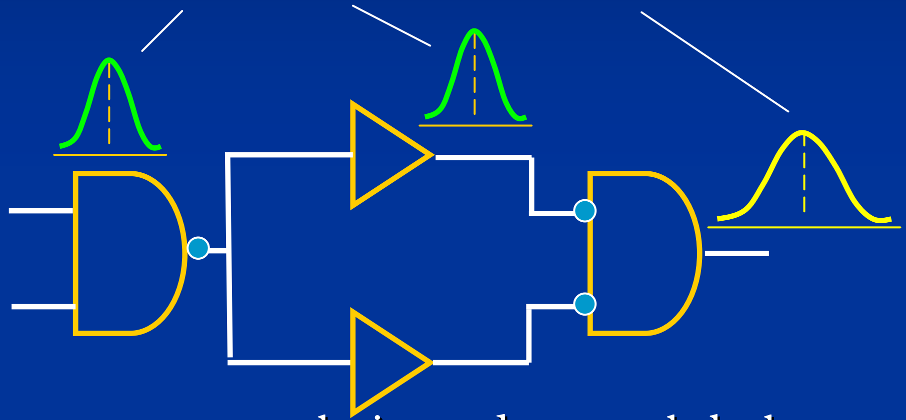
Statistical Static Timing Analysis

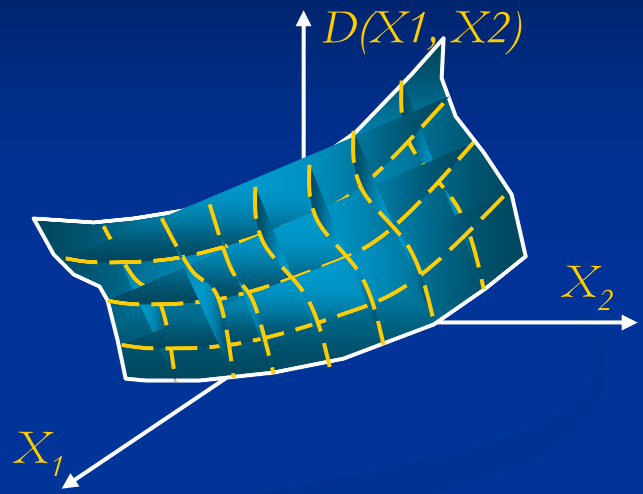
- Corner-based timing analysis: too slow and too pessimistic
- Statistical Static timing analysis (SSTA):
  - Computes PDF/CDF of arrival/required times
  - Predicts yield
  - Does not enumerate corners
- Block-based SSTA: fast run-time and incremental operation

Parameterized Block-based SSTA

- Parameterizes delays and arrival times by sources of variation:
  - L_{eff}, V_{th}, metal/dielectric thickness, etc

$D=D(X_1, X_2, \ldots)$ $A=A(X_1, X_2, \ldots)$





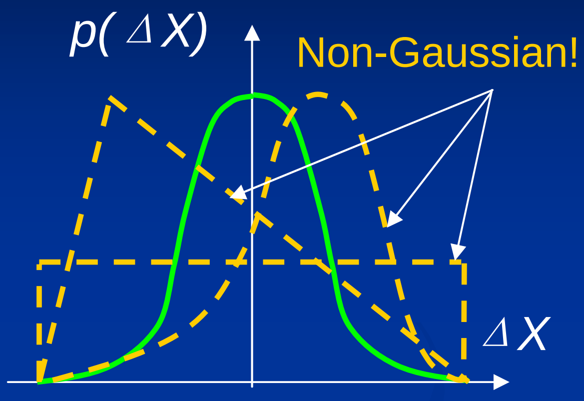
- Preserves correlations due to global sources of variation
- Predicts dependency of circuit delay on sources of variation
  - Computes circuit delay PDF/CDF and response surface
    - Convenient for circuit optimization and yield shaping
- Cannot handle non-Gaussian and nonlinear sources of variations
  - Require extension preserving all its functionality and flexibility
    - e.g. C. Visweswariah et. al. DAC'04 (IBM); Chang, Sapatnekar, ICCAD'03 (UMN)

Outline

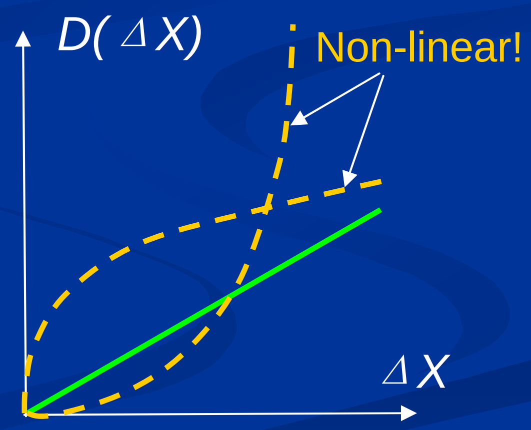
- Motivation and problem statement
- Parameterized SSTA for linear and Gaussian process parameters
  - First order canonical form of Gaussian variables
  - Linear approximations use analytical formula only
- Technique for non-linear and/or non-Gaussian parameters
  - Generalize canonical form of nonlinear & non-Gaussian parameters
  - Approximations extended on the same principles as for linear Gaussian case
  - Preserve full compatibility with linear Gaussian SSTA
- Experimental results
- Conclusions

Problem Statement

- Parameterized SSTA assumes:
  - All parameters have Gaussian distribution
  - Gate delays are linear function of parameters
- Non-Gaussian distributions cannot be approximated by Gaussian with required accuracy
  - Via resistance has asymmetric probability distribution
- Linearization is valid only for small variations
  - Delay depends non-linearly on many transistor parameters: L_{eff}, V_{th}, etc
- No known technique for efficiently handling arbitrary distributions and functions
- Accurate parameterized SSTA is required
  - For sign-off timing analysis
  - For validation of linear Gaussian approximation
  - For parameters with large variations



Distributions of parameters



Delay as a function of parameters

Proposed Approach

- Achieve high accuracy for nonlinear and non-Gaussian parameters
  - Required for validating of linear Gaussian approximation
- Handle mixture of linear Gaussian and arbitrary parameters
- Preserve merits of parameterized block-based SSTA
  - Maintain correlations
  - Compute circuit delay in parameterized form
  - Linear run time with respect to circuit size
  - Preserve high speed of handling linear Gaussian sources of variations
    - In practice only a few parameters are significantly non-linear or non-Gaussian
  - Achieve compatibility with existing implementation of IBM EinsStat tool

Delays and Arrival Times in Parameterized SSTA

- First-order canonical form of Gaussian sources of variations
  - Good for preserving correlations
- $\Delta X_i, \Delta R_a$: Gaussian independent random variables

Nominal value

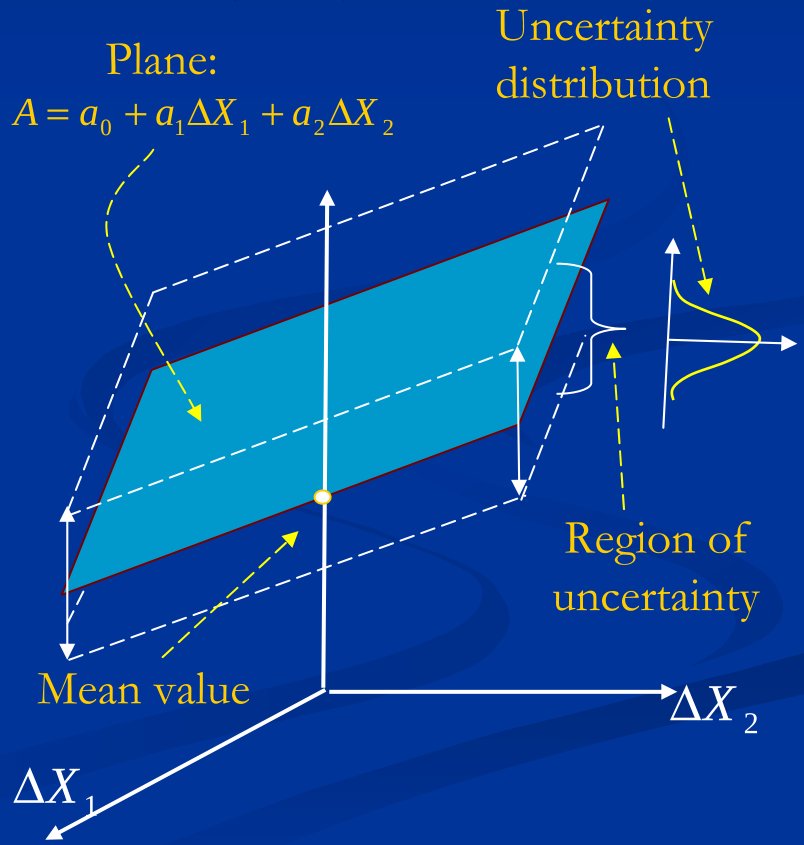
Global source of variability

Random Uncertainty

$$A = a_0 + \sum_{i=1}^{n} a_i \Delta X_i + a_{n+1} \Delta R_a$$

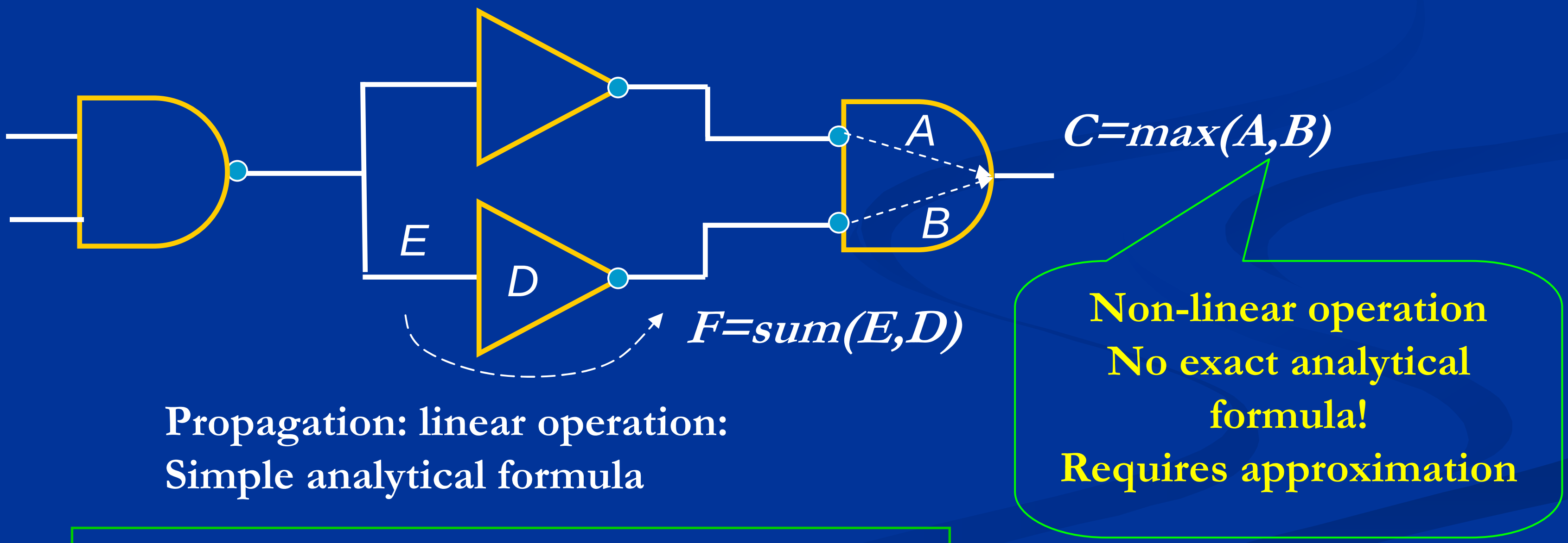
Sensitivity to global sources

Sensitivity to random component



Parameterized Block-based SSTA

- Propagation of arrival times through circuit in canonical forms
  - Similar to deterministic timing analysis
- Two timing operations on arrival times:
  - Propagating through a gate: incrementing arrival time by gate delay
  - Selection of the worst arrival time: maximum of two arrival times



Propagation: linear operation:
Simple analytical formula

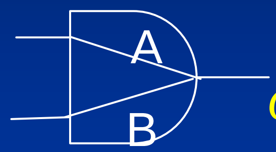
$$F = E + D = (e_0 + d_0) + \sum_{i=1}^{n} (e_i + d_i)\Delta X_i + \sqrt{e_{n+1}^2 + d_{n+1}^2}\,\Delta R_F$$

Linear Approximation of Max Operation

- Approximate *max(A,B)* with canonical form *C*

$$A = a_0 + \sum_{i=1}^{n} a_i \Delta X_i + a_{n+1} \Delta R_a$$

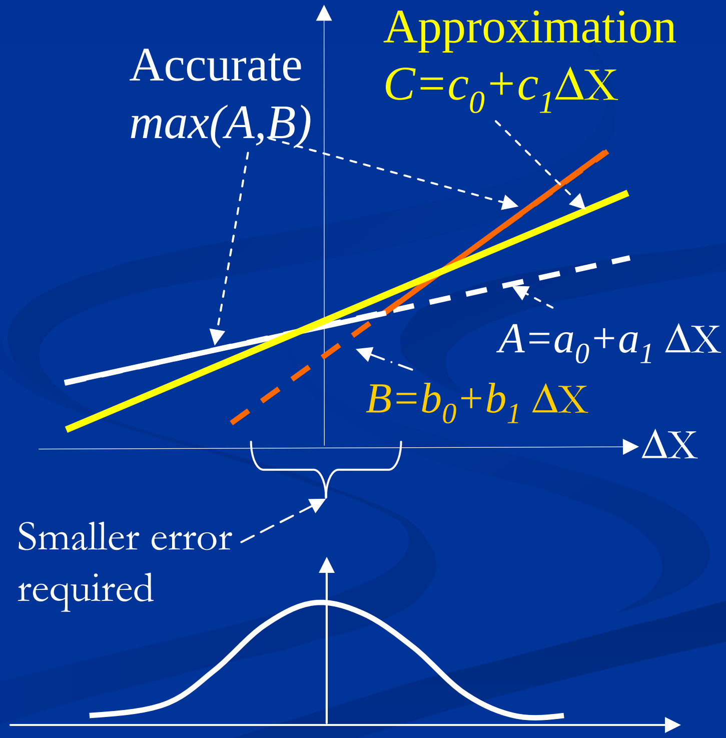
$$B = b_0 + \sum_{i=1}^{n} b_i \Delta X_i + b_{n+1} \Delta R_b$$



$$C \approx max(A, B)$$

$$C = c_0 + \sum_{i=1}^{n} c_i \Delta X_i + c_{n+1} \Delta R_c$$

- Matching and preserving:
  - Mean
  - Standard deviation (std)
  - Correlation of *C* with A, *B*
  - Correlation of *C* with all ΔX_i
- Linear approximation weighted with JPDF of ΔX_i's
  - Small error in high probability region
    - Similar to linear regression



PDF: weight function of approximation

Computing Linear Approximation of Max

- Linear approximation of max uses only analytical formulas
  - Linear complexity with respect to number of sources of variation

[C. Visweswariah et. al. DAC 2004; Chang, Sapatnekar ICCAD 2003]

$$A = a_0 + \sum_{i=1}^{n} a_i \Delta X_i + a_{n+1}\Delta R_a$$
$$B = b_0 + \sum_{i=1}^{n} b_i \Delta X_i + b_{n+1}\Delta R_b$$

$$\Rightarrow \quad C = c_0 + \sum_{i=1}^{n} c_i \Delta X_i + c_{n+1}\Delta R_c \approx max(A,B)$$

Variance of *A* and *B*, and their correlations

$$\sigma_A{}^2 = \sum_{i=1}^{n+1} a_i{}^2, \quad \sigma_B{}^2 = \sum_{i=1}^{n+1} b_i{}^2, \quad r = \sum_{i=1}^{n} a_i b_i$$

Tightness probability $T_A = P(A>B)$

$$T_A = \Phi\left(\frac{a_0 - b_0}{\theta}\right), \quad \text{where:} \quad \Phi(y) = \int_{-\infty}^{y} \phi(x)dx,$$
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad \theta = \sqrt{\sigma_A{}^2 + \sigma_B{}^2 - 2r}$$

Mean and second moment of *max(A,B)*

$$c_0 = a_0 T_A + b_0(1 - T_A) + \theta\phi\left(\frac{a_0 - b_0}{\theta}\right)$$
$$m_{2C} = (\sigma_A{}^2 + a_0{}^2)T_A + (\sigma_B{}^2 + b_0{}^2)(1 - T_A) + (a_0 + b_0)\theta\phi\left(\frac{a_0 - b_0}{\theta}\right)$$

- Sensitivities: $c_i = T_A a_i + (1 - T_A) b_i$
- Sensitivity c_{n+1} to random component to match accurate value of second moment

Overview of Proposed Technique

- Generalize canonical form of nonlinear & non-Gaussian parameters
- Represent gate delays in generalized canonical form
- Propagate arrival times in generalized canonical form
  - Extend main timing operations to generalized canonical forms
  - Preserve full compatibility with linear Gaussian SSTA
    - SSTA with linear Gaussian parameters is a special case of our approach
- Approximate max function of generalized canonical forms on the same principles as for linear Gaussian case
  - Use concept of tightness probability to preserve correlations
  - Compute sensitivities as linear combinations weighted by tightness probabilities
  - Match exact mean and variance values of max operation

Generalized Canonical Forms for Nonlinear and Non-Gaussian Parameters

- Generalize canonical form for representing delays and arrival times:
  - Introduce new term for non-linear and non-Gaussian parameters

$$A = \underbrace{a_0}_{\text{Mean}} + \underbrace{\sum_{i=i}^{n_{LG}} a_{LG,i} \cdot \Delta X_{LG,i}}_{\text{Linear Gaussian term}} + \underbrace{f_A(\Delta X_{N,1}, \Delta X_{N,2}, \ldots)}_{\text{Nonlinear non-Gaussian term}} + \underbrace{a_{n+1} \cdot \Delta R_a}_{\text{Random uncertainty}}$$

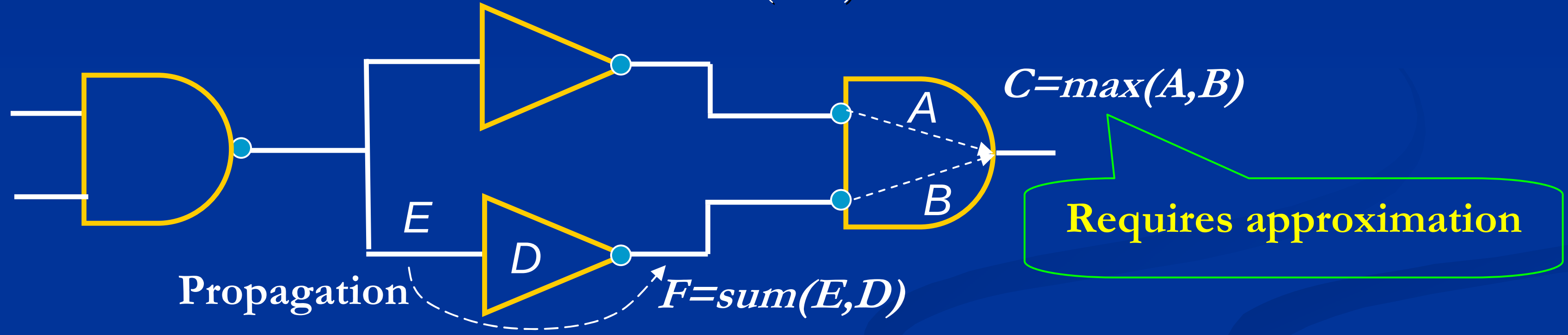
  - Good for preserving correlations
- No restrictions on non-linear function $f_A(\Delta X_{N,1}, \Delta X_{N,2}, \cdots)$
  - f_A can be arbitrary form specified by a table for numerical computation
- No restrictions on distribution of non-Gaussian parameters $\Delta X_{N,1}, \Delta X_{N,2}, \cdots$
  - Parameters can be mutually correlated with JPDF $p(\Delta X_{N,1}, \Delta X_{N,2}, \cdots)$
    - JPDF can be specified by a table for numerical computation
- In special case f_A is separable and all parameters are independent

$$f_A(\Delta X_{N,1}, \Delta X_{N,2}, \ldots) = f_A(\Delta X_{N,1}) + f_A(\Delta X_{N,2}) + \cdots$$

  - The most common and important case in practice
    - Simpler and more efficient implementation

SSTA with Nonlinear and Non-Gaussian Parameters

- Propagation of arrival times in the generalized canonical forms
  - Two timing operations on arrival times :
    - Arrival time propagation (sum): incrementing arrival time by gate delay
    - Selection of the worst arrival time (max): maximum of two arrival times



- Sum for generalized canonical form has analytical formula

$$F = E + D = (e_0 + d_0) + \sum_{i=1}^{N_{LG}} (e_{LG,i} + d_{LG,i})\Delta X_{LG,i} + (f_E(\Delta \vec{X}_N) + f_D(\Delta \vec{X}_N)) + \sqrt{e_{n+1}^2 + d_{n+1}^2}\Delta R_F$$

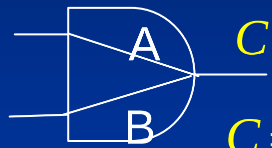
  - Nonlinear functions f_E and f_D can be summed numerically
- Approximation of *max(A,B)* in canonical form is difficult
  - Arbitrary functions and JPDFs cannot be handled analytically
  - Full numerical computation is too expensive

Max with Nonlinear and Non-Gaussian Parameters

- Approximate *max(A,B)* with generalized canonical form *C* :

$$A = a_0 + \sum_{i=1}^{n} a_i \Delta X_i + f_A(\Delta \vec{X}_N) + a_{n+1}\Delta R_a$$

$$B = b_0 + \sum_{i=1}^{n} b_i \Delta X_i + f_B(\Delta X_N) + b_{n+1}\Delta R_b$$



$$C \approx max(A, B)$$

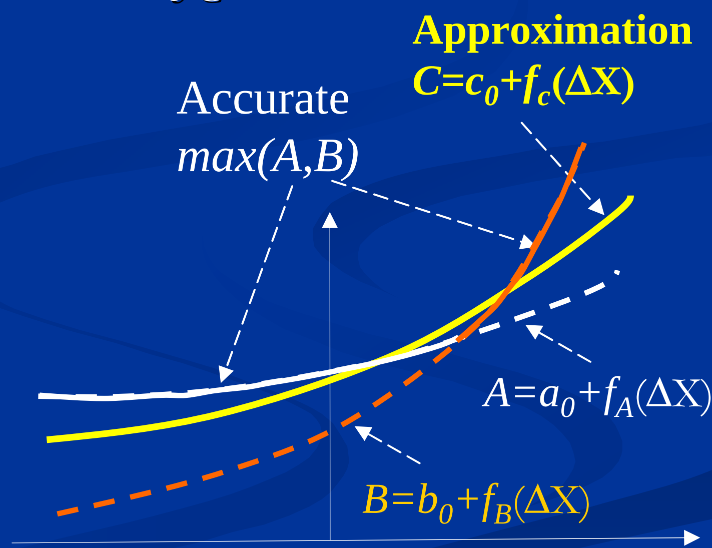
$$C = c_0 + \sum_{i=1}^{n_{LG}} c_i \Delta X_{LG,i} + f_c(\Delta \vec{X}_{N,i}) + c_{n+1}\Delta R_c$$

- Compute sensitivities c_i and non-linear function f_C
  - Use tightness probability $T_A=P(A>B)$

$$c_i = T_A a_{LG,i} + (1-T_A) b_{LG,i}$$

$$f_C(\Delta X_{N,i}) = T_A f_A(\Delta X_{N,i}) + (1-T_A) f_B(\Delta X_{N,i})$$

- Compute $c_0=E[max(A,B)]$
  - Match accurate mean of max function
- Compute c_{n+1} to match std of *max(A,B)*
  - Require second moment of *max(A,B)*



Approximation max for canonical forms with non-linear function

Computation of Tightness Probability

- Apply conditional probabilities technique
- Rewrite

$$A = a_0 + \sum_{i=i}^{n_{LG}} a_{LG,i} \cdot \Delta X_{LG,i} + f_A(\Delta \vec{X}_N) + a_{n+1} \cdot \Delta R_a$$

for fixed $\Delta \vec{X}_N$

Linear Gaussian Canonical Form

$$A_{cond}(\Delta X_N) = \underbrace{(a_0 + f_A(\Delta \vec{X}_N))}_{\text{New mean}} + \underbrace{\sum_{i=i}^{n_L} a_{LG,i} \cdot \Delta X_{LG,i} + a_{n+1} \cdot \Delta R_a}_{\text{Linear Gaussian part}}$$

- Compute conditional tightness probability for fixed values of $\Delta \vec{X}_N$

$$T_{A,Cond}(\Delta X_N) = P(A > B \mid \Delta X_N)$$

- Conditional JPDF of linear Gaussian parameters is Gaussian
  - Independence between linear Gaussian and nonlinear/non-Gaussian parameters

$$p(\Delta \vec{X}_{LG} \mid \Delta \vec{X}_N) = p(\Delta \vec{X}_{LG})$$

- Conditional tightness probability is tightness probability of linear Gaussian canonical forms A_{cond} and B_{cond}

$$T_{A,Cond}(\Delta X_N) = P(A > B \mid \Delta X_N) = T_A(A_{cond}(\Delta X_N), B_{cond}(\Delta X_N))$$

⇨ $T_{A,Cond}(\Delta X_N)$ has analytical expression !!!

  - C. Visweswariah et. al. DAC 2004; Chang, Sapatnekar, ICCAD 2003

Computation of Tightness Probability (cont.)

- Unconditional tightness probability $T_A = P(A>B)$

$$T_A = \int_{-\infty}^{\infty} T_{A,Cond}(\Delta\vec{X}_N)\cdot p(\Delta\vec{X}_N)\,d\Delta X_N$$

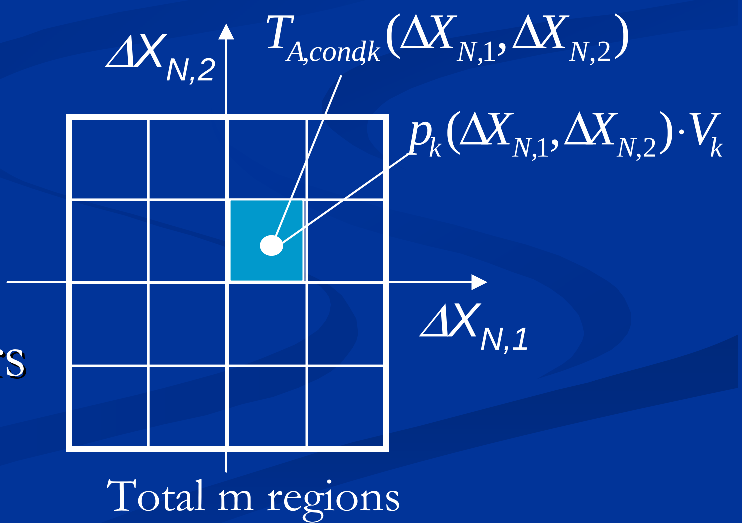
  - where $T_{A,Cond}(\Delta X_N)$ has analytical formula

- For "good" PDFs $p(\Delta\vec{X}_N)$, integration can be done analytically
  - Requires additional research
- For arbitrary PDFs $p(\Delta\vec{X}_N)$, apply numerical integration

$$T_A = \sum_{k=1}^{m} \underbrace{T_{A,Cond,k}} \cdot \underbrace{p_k(\Delta\vec{X}_N) V_k}$$

$T_{A,Cond}(\Delta X_N)$ in k_{th} grid cell — Joint probability in k_{th} grid cell

- Numerical integration is expensive, but feasible for 7-8 nonlinear parameters
  - According to experiments small number of points (e.g., 5) is enough



Mean and Second Moment Computation

- Apply conditional probabilities technique:

$$A_{cond}(\Delta X_N) = \underbrace{(a_0 + f_A(\Delta \vec{X}_N))}_{\text{New mean}} + \underbrace{\sum_{i=i}^{n_L} a_{LG,i} \cdot \Delta X_{LG,i} + a_{n+1} \cdot \Delta R_a}_{\text{Linear Gaussian part}}$$

Linear Gaussian Canonical Form

- Conditional JPDF of linear Gaussian parameters is Gaussian

$$p(\Delta \vec{X}_{LG} \,|\, \Delta \vec{X}_N) = p(\Delta \vec{X}_{LG})$$

- Conditional mean and second moment of *max(A,B)*:

$$c_{0,Cond}(\Delta X_N) = E[max(A,B) \,|\, \Delta \vec{X}_N] \qquad m_{2,Cond}(\Delta X_N) = E[(max(A,B))^2 \,|\, \Delta \vec{X}_N]$$

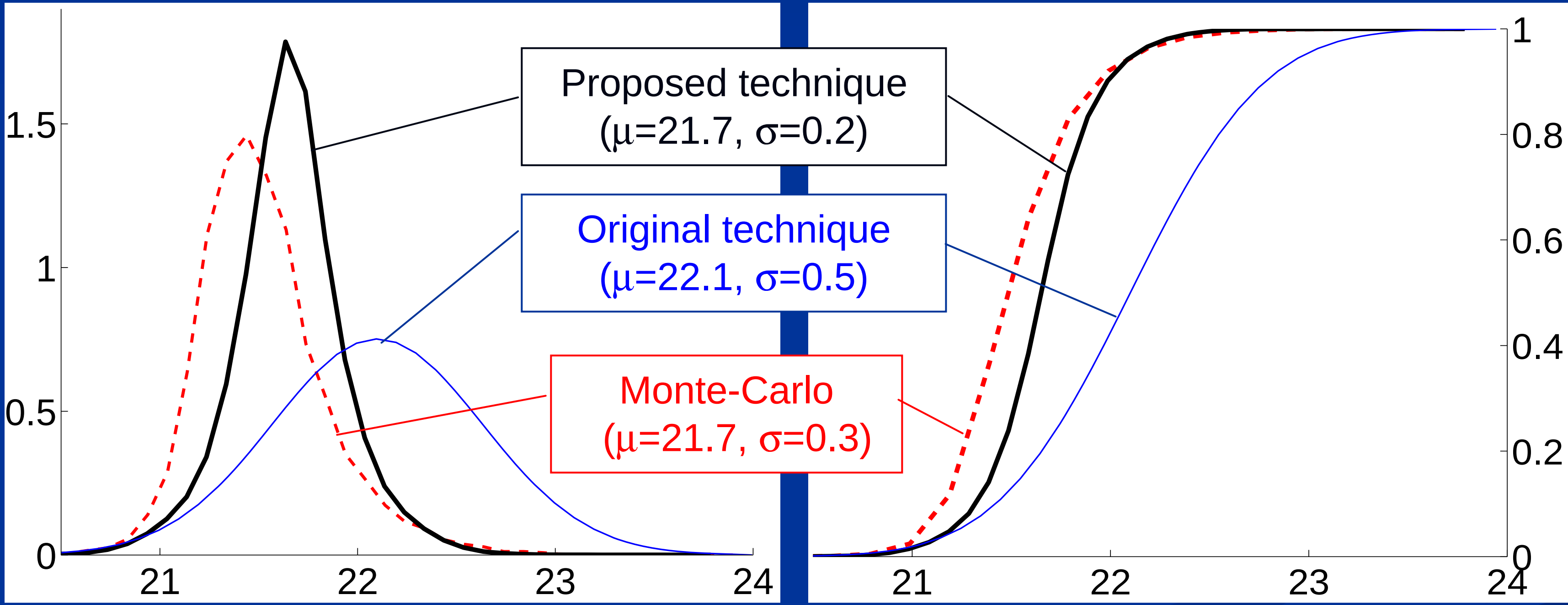
  - $c_{0,Cond}(\Delta X_N), m_{2,cond}(\Delta X_N)$ have analytical expression !!!
    - C. Visweswariah et. al. DAC 2004; Chang, Sapatnekar, ICCAD 2003
- Unconditional mean and second moment of *max(A,B)*:

$$c_0 = \int_{-\infty}^{\infty} c_{0,Cond}(\Delta \vec{X}_N) p(\Delta \vec{X}_N) d\Delta \vec{X}_N \qquad m_{2,C} = \int_{-\infty}^{\infty} m_{2,C,Cond}(\Delta \vec{X}_N) p(\Delta \vec{X}_N) d\Delta \vec{X}_N$$

- For "good" PDFs $p(\Delta \vec{X}_N)$, integration can be done analytically
- For arbitrary PDFs, use numerical integration

Test on Chip for Non-linearity

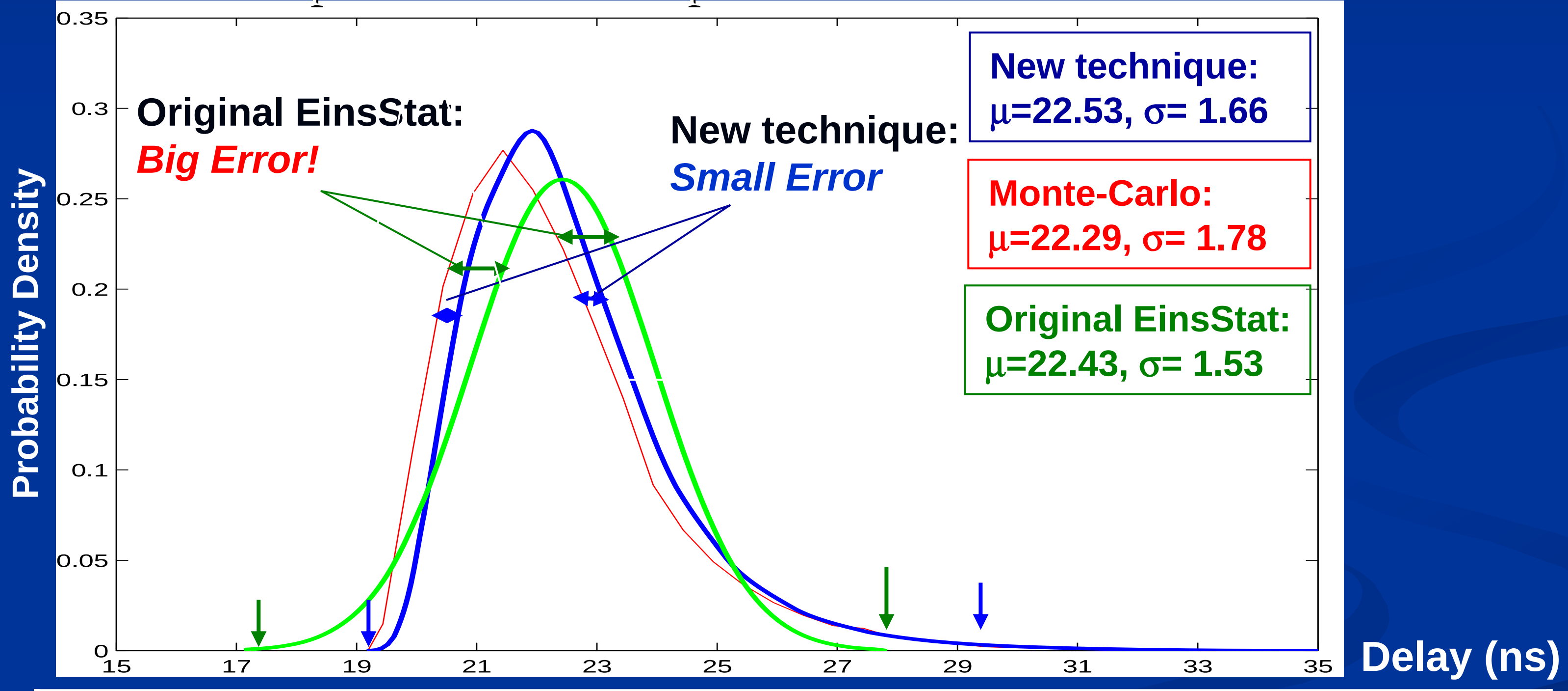
- Design A: 3,042 gates and 17,579 timing arcs
- 3 nonlinear (cubic) global parameters,
  - Each source has correlated σ=2%
  - Total independent Gaussian part: σ=6%



	Monte-Carlo	New Technique	Error%	Original EinsStat	Error%
99% Pt.	22.9ns	22.7ns	-0.9%	23.7ns	3.5%

Test on Chip for Non-Gaussian

- Design A: 3,042 gates and 17,579 timing arcs
- 3 global sources of variations
  - Each source has correlated 'Lognormal' distribution: σ=2%
  - Total independent Gaussian part: σ=6%



	Monte-Carlo	New Technique	Error%	Original EinsStat	Error%
99% Pt.	28.2ns	28.3ns	0.3%	26.0	-5.8%

Run-time

Run-time results w.r.t. number of non-Gaussian parameters

Design Name	Gates	Timing Arcs	CPU-time (second) v.s # of Non-Gaussians			
			3	2	1	0
A	3,042	17,579	3.82	1.54	1.40	1.38
B	11,937	57,151	12.32	5.53	4.27	3.07
C	53,317	392,097	79.07	35.77	27.34	18.71
D	70,216	363,537	93.25	41.28	30.52	19.68
E	1,085,034	5,799,545	2,083 (35 mins)	982 (16 mins)	788 (13 mins)	703 (12mins)

- Run-time with 3 non-Gaussian parameters is 3-5 times of that of pure-Gaussians
- Nonlinear parameters has similar dependence of run-time on
  - Number of discreitized points
  - Number of parameters
- The method is feasible to solve up to 6-8 non-Gaussian/nonlinear parameters

Conclusions

- Parameterized block-based SSTA is extended to handle:
  - Sources of variations with arbitrary non-Gaussian distributions
  - Nonlinear dependence of gates delays on process parameters
- New technique is implemented in industrial SSTA tool EinsStat
  - Can handle circuits with more than 1,000,000 gates
- Experimental results match closely with Monte-Carlo simulation
- It can handle up to 6-8 nonlinear/non-Gaussian parameters
  - It is much more efficient than Monte-Carlo simulation
- Run-time is linear w.r.t.:
  - Circuit size
  - Number of linear Gaussian parameters
  - Number of discretized points of non-Gaussian and nonlinear parameters
    - Can be improved by using analytical integration for "good" functions and PDFs
    - Multidimensional integration can be done by Monte-Carlo
- The technique is important and useful for:
  - Validation of linear Gaussian approximation
  - Accurate statistical timing analysis for sign-off