Static Timing Analysis

Statistical Timing Analysis: From Basic Principles to State of the Art

Vladimir Todorov

Moscow-Bavarian Joint Advanced Student School 2009

March 19, 2009

Overview

Where are we?

What is Static-Timing Analysis?

- A tool for analysing and computing delays for digital circuits.
- Advantageous over difficult-to-construct vector-based timing simulations.
- Provides conservative analysis of the delay.
- Traditional STA is deterministic and computes the delay for specific process conditions.
- Separate analysis for each condition is ran using corner files.

From Deterministic STA to Statistical STA

- Due to process scalling accross die variations are non-negligible.

From Deterministic STA to Statistical STA

- Due to process scalling accross die variations are non-negligible.
- Fundamental weakness of DSTA is the inability to model within-die variations.

From Deterministic STA to Statistical STA

- Due to process scalling accross die variations are non-negligible.
- Fundamental weakness of DSTA is the inability to model within-die variations.
- This results in either under- or overestimation of the delay (no longer conservative).

From Deterministic STA to Statistical STA

- Due to process scalling accross die variations are non-negligible.
- Fundamental weakness of DSTA is the inability to model within-die variations.
- This results in either under- or overestimation of the delay (no longer conservative).
- Further increase in process parameters increases the runtime of DSTA

From Deterministic STA to Statistical STA

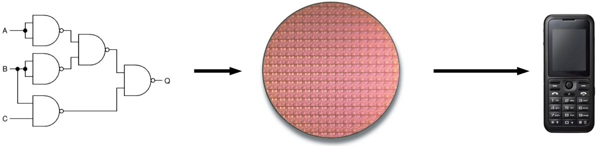
- Due to process scalling accross die variations are non-negligible.
- Fundamental weakness of DSTA is the inability to model within-die variations.
- This results in either under- or overestimation of the delay (no longer conservative).
- Further increase in process parameters increases the runtime of DSTA
- These give a rise to Statistical Static-Timing Analysis (SSTA)

Where are we?

Sources of Timing Variation

There are three sources of timing variation that need to be considered:

- **Model Uncertainty**
- **Process Uncertainty**
- **Environment Uncertainty**

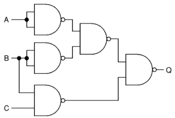


Model Uncertainty **Process Uncertainty** **Environment Uncertainty**

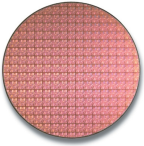
Sources of Timing Variation

There are three sources of timing variation that need to be considered:

- **Model Uncertainty**
  Inaccuracy in device models used, extraction and reduction of interconnect parasitics and timing-analysis algorithms
- **Process Uncertainty**
- **Environment Uncertainty**



Model Uncertainty

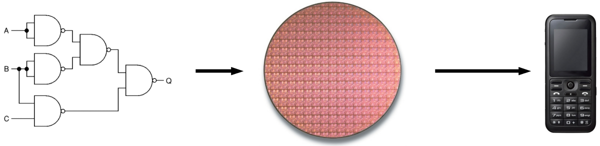


Process Uncertainty

Environment Uncertainty

There are three sources of timing variation that need to be considered:

- **Model Uncertainty**
- **Process Uncertainty**
  Uncertainty in the parameters of fabricated devices and interconnects (die-to-die and within-die)
- **Environment Uncertainty**

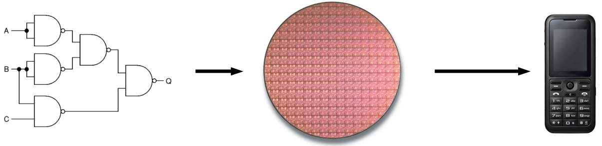


Model Uncertainty **Process Uncertainty** **Environment Uncertainty**

Sources of Timing Variation

There are three sources of timing variation that need to be considered:

- **Model Uncertainty**
- **Process Uncertainty**
- **Environment Uncertainty**
  Uncertainty in the operating environment of a device: temperature, operating voltage, mode of operation, lifetime wear-out



Model Uncertainty **Process Uncertainty** **Environment Uncertainty**

Sources of Process Variation

Physical Parameter Variations

Critical Dimension
Oxide Thickness
Channel Doping
Wire Width
Wire Thickness

Sources of Process Variation

Physical Parameter Variations	$\longrightarrow$	Electrical Parameter Variations
Critical Dimension		Saturation Current
Oxide Thickness		Gate Capacitance
Channel Doping		Treshold Voltage
Wire Width		Wire Capacitance
Wire Thickness		Wire Resistance

Sources of Process Variation

Physical Parameter Variations ⟶ Electrical Parameter Variations ⟶ Delay Variations

Physical Parameter Variations	Electrical Parameter Variations	Delay Variations
Critical Dimension	Saturation Current	Gate Delay
Oxide Thickness	Gate Capacitance	Slew Rate
Channel Doping	Treshold Voltage	Wire Delay
Wire Width	Wire Capacitance	
Wire Thickness	Wire Resistance	

Sources of Process Variation

Physical Parameter Variations → Electrical Parameter Variations → Delay Variations

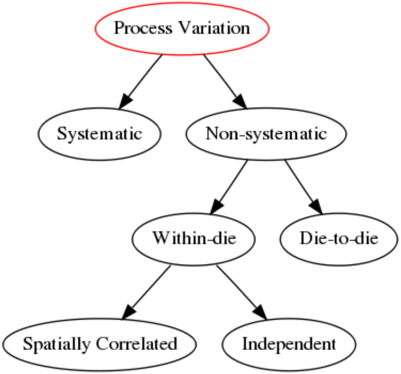
Physical Parameter Variations	Electrical Parameter Variations	Delay Variations
Critical Dimension	Saturation Current	Gate Delay
Oxide Thickness	Gate Capacitance	Slew Rate
Channel Doping	Treshold Voltage	Wire Delay
Wire Width	Wire Capacitance	
Wire Thickness	Wire Resistance	

Variations are Correlated

- Physical variations are result of process variations and may be correlated as one process variation may result in many physical.
- Electrical variations are also correlated as one physical variation may result in more than one electrical variation. (Ex. R and C are negatively correlated with respect to wire width.)

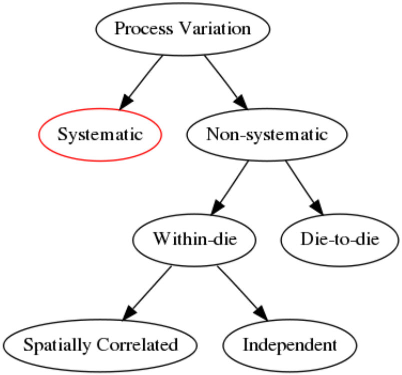
Classification of Physical Variations

Physical variations may be characterized as either deterministic or statistical.



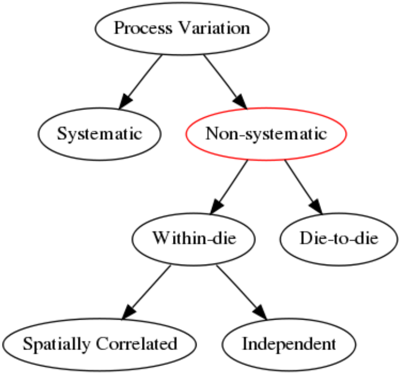
Classification of Physical Variations

- Follow well understood behavior.
- Predicted upfront by analyzing the design layout
- Arise due to proximity effects and chemical metal polishing.
- Deterministic treatment at later stages of the design.
- Advantageous to treat them statistically at early stages.



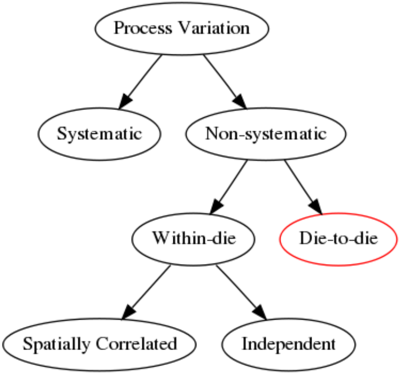
Classification of Physical Variations

- Truly uncertain component of variation.
- Result from processes orthogonal to design implementation.
- Only statistics are known at design time.
- Modeled as random variables (*RVs*).



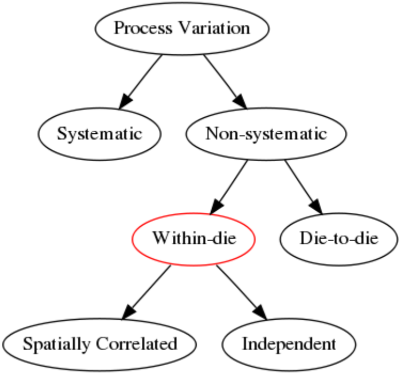
Classification of Physical Variations

- Affect all devices on the same die in the same way.
- Result in shifts that occur from lot to lot, wafer to wafer, reticle to reticle and accross a reticle.



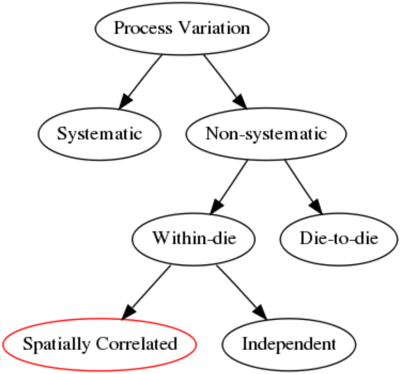
Classification of Physical Variations

- Affect each device on the same die differently.
- Only caused by within reticle variations in the confines of a single chip layout
- Divided into correlated and independent variations.



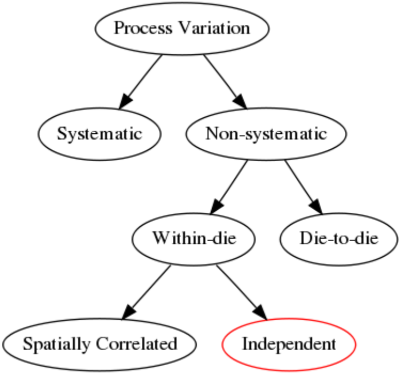
Classification of Physical Variations

- Processes that cause within die variation change gradually with position
- Affect closely spaced devices in similar manner (these exhibit similar characteristics)
- Referred to as Spatially Correlated Variation.



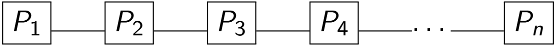
Classification of Physical Variations

- The residual variability of a device that is statistically independent of all other devices
- Increasing effect with continued process scalling.
- Examples are line-edge roughtness and random-dopant fluctuations.



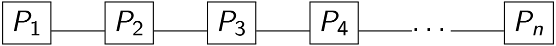
Impact of Variation on Circuit Delay

Single Path Delay



- Gates connected in series with delay probabilities P_i
- Delays-normally distributed and independent with mean μ and variance σ^2

Single Path Delay

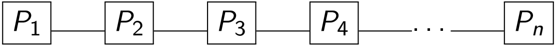


- Gates connected in series with delay probabilities P_i
- Delays-normally distributed and independent with mean μ and variance σ^2
- The total delay distribution results in the sum of the distributions on the path

$$\sum_{i=1}^{n} \mathcal{N}(\mu, \sigma^2) = \mathcal{N}\left(\sum_{i=1}^{n} \mu, \sum_{i=1}^{n} \sigma^2\right) = \mathcal{N}(n\mu, n\sigma^2)$$

Impact of Variation on Circuit Delay

Single Path Delay



- Gates connected in series with delay probabilities P_i
- Delays-normally distributed and independent with mean μ and variance σ^2
- The total delay distribution results in the sum of the distributions on the path

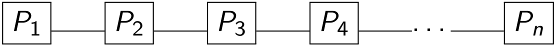
$$\sum_{i=1}^{n} \mathcal{N}(\mu, \sigma^2) = \mathcal{N}\left(\sum_{i=1}^{n} \mu, \sum_{i=1}^{n} \sigma^2\right) = \mathcal{N}(n\mu, n\sigma^2)$$

- Which results in total coefficient of variation

$$\left(\frac{\sigma}{\mu}\right)_{path} = \frac{\sqrt{n\sigma^2}}{n\mu} = \frac{1}{\sqrt{n}}\left(\frac{\sigma}{\mu}\right)_{gate}$$

Impact of Variation on Circuit Delay

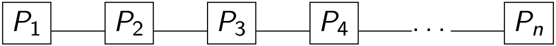
Single Path Delay



- Gates connected in series with delay probabilities P_i
- Delays-normally distributed and correlated with mean μ, variance σ^2 and correlation coefficient ρ

Impact of Variation on Circuit Delay

Single Path Delay



- Gates connected in series with delay probabilities P_i
- Delays-normally distributed and correlated with mean μ, variance σ^2 and correlation coefficient ρ
- The total delay distribution results in the sum of the distributions on the path

$$\mu_{path} = n\mu$$

$$\sigma_{path}^2 = \sum_{i=1}^{n} \sigma^2 + 2\rho \sum_{i=1}^{n} \sum_{j>i}^{n} \sigma_i \sigma_j = n\sigma^2(1 + \rho(n-1))$$

Impact of Variation on Circuit Delay

Single Path Delay

P_1 — P_2 — P_3 — P_4 — ... — P_n

- Gates connected in series with delay probabilities P_i
- Delays-normally distributed and correlated with mean μ, variance σ^2 and correlation coefficient ρ
- The total delay distribution results in the sum of the distributions on the path

$$\mu_{path} = n\mu$$

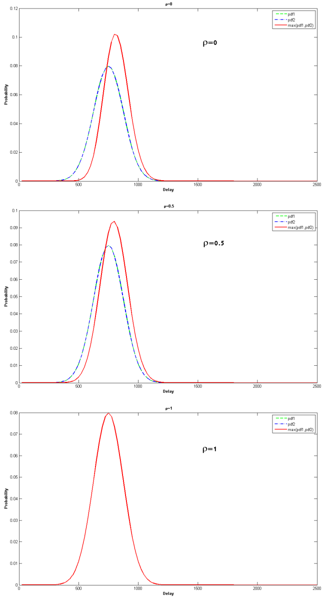
$$\sigma_{path}^2 = \sum_{i=1}^{n} \sigma^2 + 2\rho \sum_{i=1}^{n} \sum_{j>i}^{n} \sigma_i \sigma_j = n\sigma^2(1 + \rho(n-1))$$

- Which results in total coefficient of variation

$$\left(\frac{\sigma}{\mu}\right)_{path} = \frac{\sqrt{n\sigma^2(1+\rho(n-1))}}{n\mu} = \sqrt{\frac{1+\rho(n-1)}{n}} \left(\frac{\sigma}{\mu}\right)_{gate}$$

Impact of Variation on Circuit Delay

Maximum Delay of Multiple Paths



- Two paths with equal delay distributions
- Three cases are considered:
  - $\rho = 0$
  - $\rho = 0.5$
  - $\rho = 1$
- The independent case overestimates the delay

Impact of Assumptions

- *The independent assumption will underestimate the spread of single path delay and will overestimate the maximum of delay of multiple paths.*
- *The correlated assumption will overestimate the spread of single path delay and will underestimate the maximum delay of multiple paths.*
- *Assumptions may be based on circuit topology.*

Problem Formulation

- DAG $G = \{N, E, n_s, n_f\}$
  - N set of nodes (in/out pins)
  - E set of edges with weights d_i
  - n_s, n_f source and sink
- If d_i are RVs then the total delay is RV as well

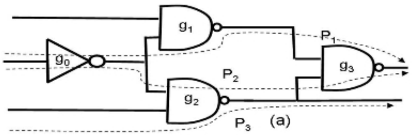


Figure taken from the original paper Static-Timing Analysis: From Basic Principles to State of the Art

Problem Formulation

- DAG $G = \{N, E, n_s, n_f\}$
  - N set of nodes (in/out pins)
  - E set of edges with weights d_i
  - n_s, n_f source and sink
- If d_i are RVs then the total delay is RV as well

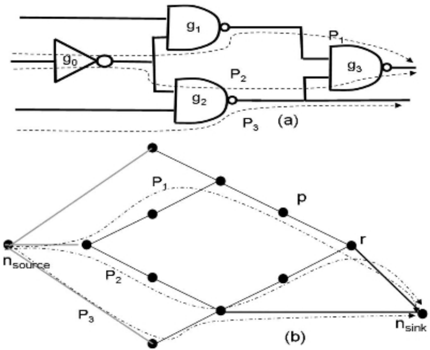


Figure taken from the original paper Static-Timing Analysis: From Basic Principles to State of the Art

- DAG $G = \{N, E, n_s, n_f\}$
  - N set of nodes (in/out pins)
  - E set of edges with weights d_i
  - n_s, n_f source and sink
- If d_i are RVs then the total delay is RV as well

Definition:

Let p_i be a path of ordered edges from source to sink in G.
Let $D_i = \sum^k d_{i_j}$ be the path length of p_i.
Then $D_{max} = max(D_1, \ldots, D_i, \ldots, D_n)$ is referred as the SSTA problem of the circuit.

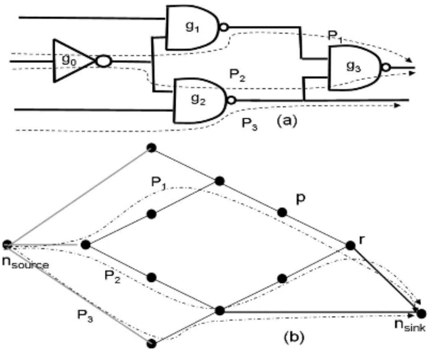


Figure taken from the original paper Static-Timing Analysis: From Basic Principles to State of the Art

Problem Formulation

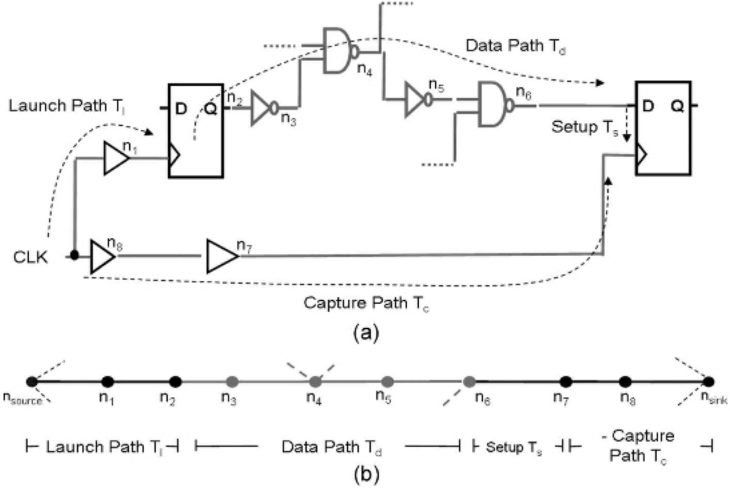
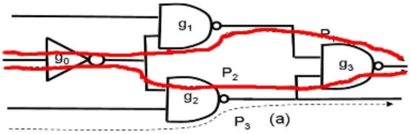


Figure taken from the original paper Static-Timing Analysis: From Basic Principles to State of the Art

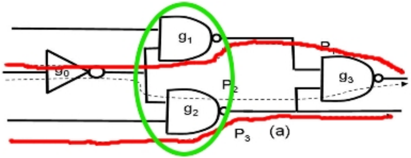
Challenges in SSTA

- Topological Correlation
  - Caused by reconvergent paths
  - Complicates the *max*(...)



Challenges in SSTA

- Topological Correlation
  - Caused by reconvergent paths
  - Complicates the $max(\ldots)$
- Spatial Correlation
  - Due to device proximity
  - How to model gate delays and arrival times in order to express the parameter correlation?
  - How to propagate and preserve the correlation information?



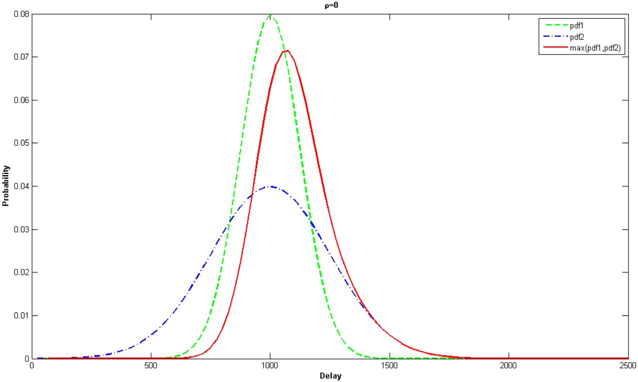
Challenges in SSTA

- Nonnormal Process Parameters and Non-linear Delay Models
  - Nonnormal physical variations exist.
  - Dependence of electrical parameters can be nonlinear.
  - Due to reduction of geometries linear assumption no longer applies.

Challenges in SSTA

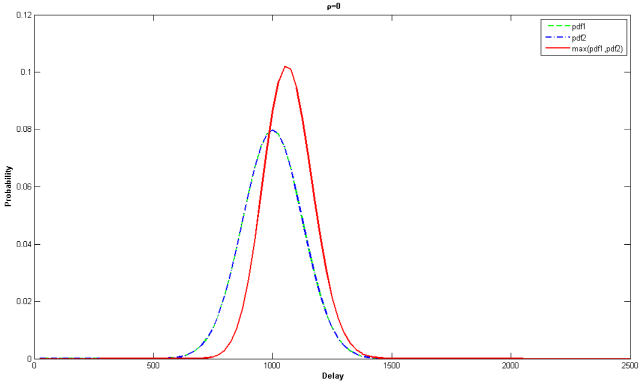
- Nonnormal Process Parameters and Non-linear Delay Models
  - Nonnormal physical variations exist.
  - Dependence of electrical parameters can be nonlinear.
  - Due to reduction of geometries linear assumption no longer applies.
- Skewness due to Maximum Operation
  - Maximum operation is nonlinear, hence
  - Normal arrival times will result in nonnormal delay
  - Maximum operation which can operate on nonnormal delays is desired
  - Simply ignoring the skewness introduces error

Skewness due to Maximal Operation



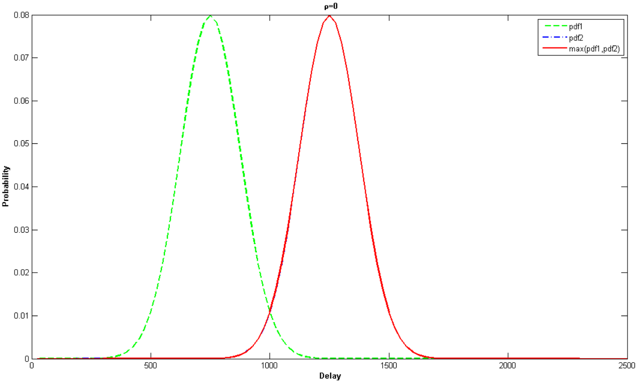
Two arrival times with same mean, but different variance result in a positively skewed maximum delay.

Skewness due to Maximal Operation



Identically distributed arrival times result in slightly positively skewed distribution.

Skewness due to Maximal Operation



The maximum can safely be assumed to be the distribution with the greater mean.

SSTA Solution Approaches

- Numerical Integration
  - Most general
  - Computationally expensive
- Monte Carlo Methods
  - Statistical sampling of the sample space
  - Perform deterministic computation for each sample
  - Agregate these results into a final result
- Probabilistic Analysis Methods
  - Path-based Approach
  - Block-based Approach

Probabilistic Analysis Methods

Path-based Approach

How it works...

- Set of likely to become critical paths
- Compute the delay for each path
- Perform a statistical maximum

Block-based Approach

How it works...

- For all fan-in edges the edge delay is added to the arrival time at the source node
- Given the resulting times the final arrival time at the node is computed using maximum operation
- Propagates exactly 2 arrival times(rise and fall)

Probabilistic Analysis Methods

Path-based Approach → *Block-based Approach*

How it works... (Path-based)

- Set of likely to become critical paths
- Compute the delay for each path
- Perform a statistical maximum

Problems...

- Difficult to construct set of suitable paths
- High computational efforts for balanced circuits

How it works... (Block-based)

- For all fan-in edges the edge delay is added to the arrival time at the source node
- Given the resulting times the final arrival time at the node is computed using maximum operation
- Propagates exactly 2 arrival times(rise and fall)

Distribution Propagation Approaches

Propagation of Sampled and Renormalized Distributions

- Computed using discrete sum and maximum
- Summation is done by combining multiple shifted values of the delay

$$z = x + y$$

- Maximum is taken by evaluatig the probability

$$z = max(x, y)$$
$$f_z(t) = F_x(\tau < t)f_y(t) + F_y(\tau < t)f_x(t)$$

Distribution Propagation Approaches

Propagation of Sampled and Renormalized Distributions

- Computed using discrete sum and maximum
- Summation is done by combining multiple shifted values of the delay

$$f_z(t) = f_x(1)f_y(t-1) + f_x(2)f_y(t-2) + \cdots + f_z(n)f_y(t-n)$$

- Maximum is taken by evaluatig the probability

$$z = max(x, y)$$
$$f_z(t) = F_x(\tau < t)f_y(t) + F_y(\tau < t)f_x(t)$$

Distribution Propagation Approaches

Propagation of Sampled and Renormalized Distributions

- Computed using discrete sum and maximum
- Summation is done by combining multiple shifted values of the delay

$$f_z(t) = \sum_{i=-\infty}^{\infty} f_x(i) f_y(t-i)$$

- Maximum is taken by evaluatig the probability

$$z = max(x, y)$$
$$f_z(t) = F_x(\tau < t) f_y(t) + F_y(\tau < t) f_x(t)$$

Distribution Propagation Approaches

Propagation of Sampled and Renormalized Distributions

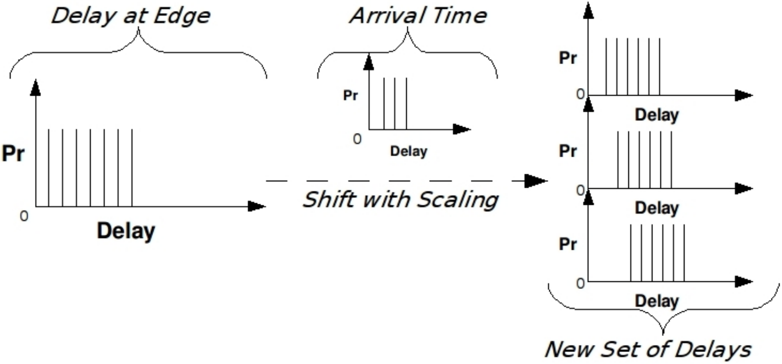
- Computed using discrete sum and maximum
- Summation is done by combining multiple shifted values of the delay

$$f_z(t) = \sum_{i=-\infty}^{\infty} f_x(i) f_y(t-i) = f_x(t) * f_y(t)$$

- Maximum is taken by evaluatig the probability

$$z = max(x, y)$$
$$f_z(t) = F_x(\tau < t) f_y(t) + F_y(\tau < t) f_x(t)$$

Discrete Distribution Propagation

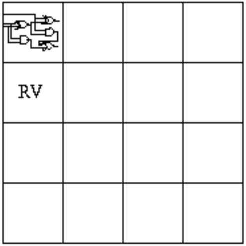


Handling Topological Correlation

- Super-Gate
  - Statistically independent inputs
  - Single fan-out
  - Separate propagation of discrete events (enumeration) $\in O(c^n)$
- Ignoring Topological Correlations
  - Exists a pdf $Q(t)$ which upper-bounds $P(t)$ for all t
  - Results in pessimistic analysis
  - Original $P(t)$ can be approximated by upper and lower bounds

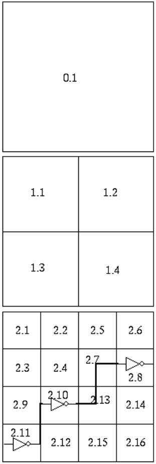
Correlation Models (Parameter Space)

- Grid Model
  - Divide the die by a square grid
  - Each square corresponds to a group of fully correlated devices
  - Each square is a RV and is correlated to all other squares
  - Construct a new set of RVs by whitening
  - Express the old set as a linear combination of the new one



Correlation Models (Parameter Space)

- Quadtree Model
  - Recursively divide the die area into 4
  - Each partition is assigned to an independent RV
  - Express the correlation variation by summing the RV of the gate with the ones from higher levels
  - Correlation arises from sharing RVs on higher levels



Propagation of Delay Dependence (Parameter Space)

- Canonical form of the delay

$$d_a = \mu_a + \sum_{i}^{n} a_i z_i + a_{n+1} R$$

Propagation of Delay Dependence (Parameter Space)

- Canonical form of the delay

$$d_a = \mu_a + \sum_{i}^{n} a_i z_i + a_{n+1} R$$

- Express the sum in a canonical form

$$\begin{aligned} C &= A + B \\ \mu_c &= \mu_a + \mu_b \\ c_i &= a_i + b_i \quad \text{for} \quad 1 \leq i \leq n \\ c_{n+1} &= \sqrt{a_{n+1}^2 + b_{n+1}^2} \end{aligned}$$

Propagation of Delay Dependence (Parameter Space)

- Express the maxium in a canonical form

1 Compute variance and covariance of A and B

$$\sigma_a^2 = \sum_i^n a_i^2 \quad \sigma_b^2 = \sum_i^n b_i^2 \quad r = \sum_i^n a_i b_i$$

Propagation of Delay Dependence (Parameter Space)

- Express the maxium in a canonical form

1. Compute variance and covariance of A and B

$$\sigma_a^2 = \sum_i^n a_i^2 \quad \sigma_b^2 = \sum_i^n b_i^2 \quad r = \sum_i^n a_i b_i$$

2. Compute the tightness probability $T_A = Pr(A > B)$

$$T_A = \Phi\left(\frac{\mu_a - \mu_b}{\theta}\right)$$

$$\Phi(x') = \int_{-\infty}^{x'} \phi(x)dx$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{x^2}{2}}$$

$$\theta = \sqrt{\sigma_a^2 + \sigma_b^2 - 2r}$$

Propagation of Delay Dependence (Parameter Space)

3 Compute mean and variance of $C = max(A, B)$

$$\mu_c = \mu_a T_A + \mu_b(1 - T_A) + \theta\phi\left(\frac{\mu_a - \mu_b}{\theta}\right)$$

$$\sigma_c^2 = (\mu_a + \sigma_a^2)T_A + (\mu_b + \sigma_b^2)(1 - T_A) + (\mu_a + \mu_b)\theta\phi\left(\frac{\mu_a - \mu_b}{\theta}\right) - \mu_c^2$$

Propagation of Delay Dependence (Parameter Space)

3 Compute mean and variance of $C = max(A, B)$

$$\mu_c = \mu_a T_A + \mu_b(1 - T_A) + \theta\phi\left(\frac{\mu_a - \mu_b}{\theta}\right)$$

$$\sigma_c^2 = (\mu_a + \sigma_a^2)T_A + (\mu_b + \sigma_b^2)(1 - T_A) + (\mu_a + \mu_b)\theta\phi\left(\frac{\mu_a - \mu_b}{\theta}\right) - \mu_c^2$$

4 Compute sensitivity coefficients c_i

$$c_i = a_i T_A + b_i(1 - T_A) \quad for \quad 1 \leq i \leq n$$

Propagation of Delay Dependence (Parameter Space)

3 Compute mean and variance of $C = max(A, B)$

$$\mu_c = \mu_a T_A + \mu_b(1 - T_A) + \theta\phi\left(\frac{\mu_a - \mu_b}{\theta}\right)$$

$$\sigma_c^2 = (\mu_a + \sigma_a^2)T_A + (\mu_b + \sigma_b^2)(1 - T_A) + (\mu_a + \mu_b)\theta\phi\left(\frac{\mu_a - \mu_b}{\theta}\right) - \mu_c^2$$

4 Compute sensitivity coefficients c_i

$$c_i = a_i T_A + b_i(1 - T_A) \quad for \quad 1 \leq i \leq n$$

5 Compute c_{n+1} of C_{approx} to get a consistent estimate

Propagation of Delay Dependence (Parameter Space)

- C_{approx} is only approximation and is not conservative
- Therefore, by the use of the relationship

$$max\left(\sum_{i}^{n} a_i, \sum_{i}^{n} b_i\right) \leq \sum_{i}^{n} max(a_i, b_i)$$

- C_{bound} can be constructed which is conservative

$$\mu_c = max(\mu_a, \mu_b)$$
$$c_{bound_i} = max(a_i, b_i)$$

Nonlinear and Nonnormal Approaches

- Nonlinear device dependencies

$$d_a = \mu_a + \sum_i^n a_i z_i + \sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{b_{ij} z_i z_j} + a_{n+1} R$$

- Nonnormal physical or electrical variations

$$d_a = \mu_a + \sum_i^n a_i z_i + \sum_{\mathbf{j}}^{\mathbf{m}} \mathbf{a_{n+j} z_{n+j}} + a_{n+m+1} R$$

- Generalized in

$$d_a = \mu_a + \sum_i^n a_i z_i + \mathbf{f(z_{n+1}, \ldots, z_{n+m})} + a_{n+1} R$$

- Handled by numerical computations and tightness probabilities

Where are we?

Conclusion

- SSTA has gained excessive interest in recent years
- Currently a number of commercial efforts are underway
- However, state-of-the-art SSTA does not address many of the issues taken for granted in DSTA
  - Coupling noise
  - Clock issues
  - Complex delay modelling
- SSTA must move beyound analysis into optimization

Discussion

. . .