A Probabilistic Analysis of Pipelined Global Interconnect Under Process Variations

N. Kankani, V. Agarwal & Dr. J. M. Wang
Digital VLSI Design Laboratory
Electrical & Computer Engineering
University of Arizona
Speaker: Vineet Agarwal

Overview

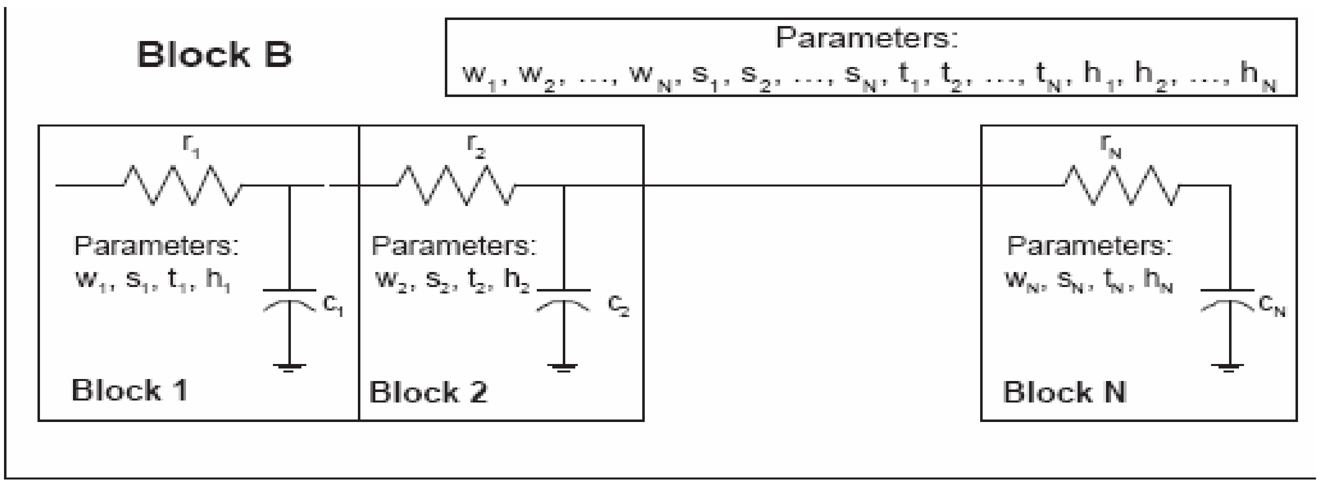
- A Novel Delay Metric based on ANOVA
- Statistical Timing Analysis of Pipelined Global Interconnect
- Reliability Aware Global Interconnect Sharing
- Experimental Results

Overview

- A Novel Delay Metric based on ANOVA
- Statistical Timing Analysis of Pipelined Global Interconnect
- Reliability Aware Global Interconnect Sharing
- Experimental Results

Our Approach – Delay Metric based on Anova (DMA)

- Given the uncertainty in parameters and the degree of required model, DMA returns a polynomial expression for delay
- A distributed RC model for interconnect is assumed



RC –Tree Model for Global Interconnect

Algorithm for DMA Implementation

{B1, n} = DIVIDE(Block B);	//divide Block B into 'n' smaller //identical blocks B1
{M} = FindModel(B1, p);	//find model M of degree 'p' for block //B1 using PCM

{P } = ANOVA(M, R^2);	//run ANOVA on M to find
	//insignificant variables such that
	//reduced model is R^2 % accurate

{E} = Extrapolate(P);	//extrapolate the insignificant variables
	//for other identical blocks of B1 to find
	//all the insignificant variables. E

{RM} = VarRed(E, Block B);	//using E, eliminate variables from //block B to get reduced model RM
{μ, σ} = FindMoments(RM);	//find moments of the reduced model //RM

Algorithm for DMA Implementation – Step 3 ANOVA

- What is ANOVA (Analysis of Variance)?
- As its name suggests – " Analyzes Variances"
- Main Idea - Decomposition of total variance

$$\sigma^2 = \sum_i \sigma_i^2$$

- Mean response due to a particular input - Keep that input constant and vary all other inputs

$$\hat{\mu}_i(\xi_i) \equiv \int \dots \int \hat{y}(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 \dots d\xi_{i-1} d\xi_{i+1} \dots d\xi_n$$

ANOVA Basics

- Variance due to design variable ξ_i

$$\hat{\sigma}_i^2 = \int [\hat{\mu}(\xi_i) - \mu]^2 d\xi_i$$

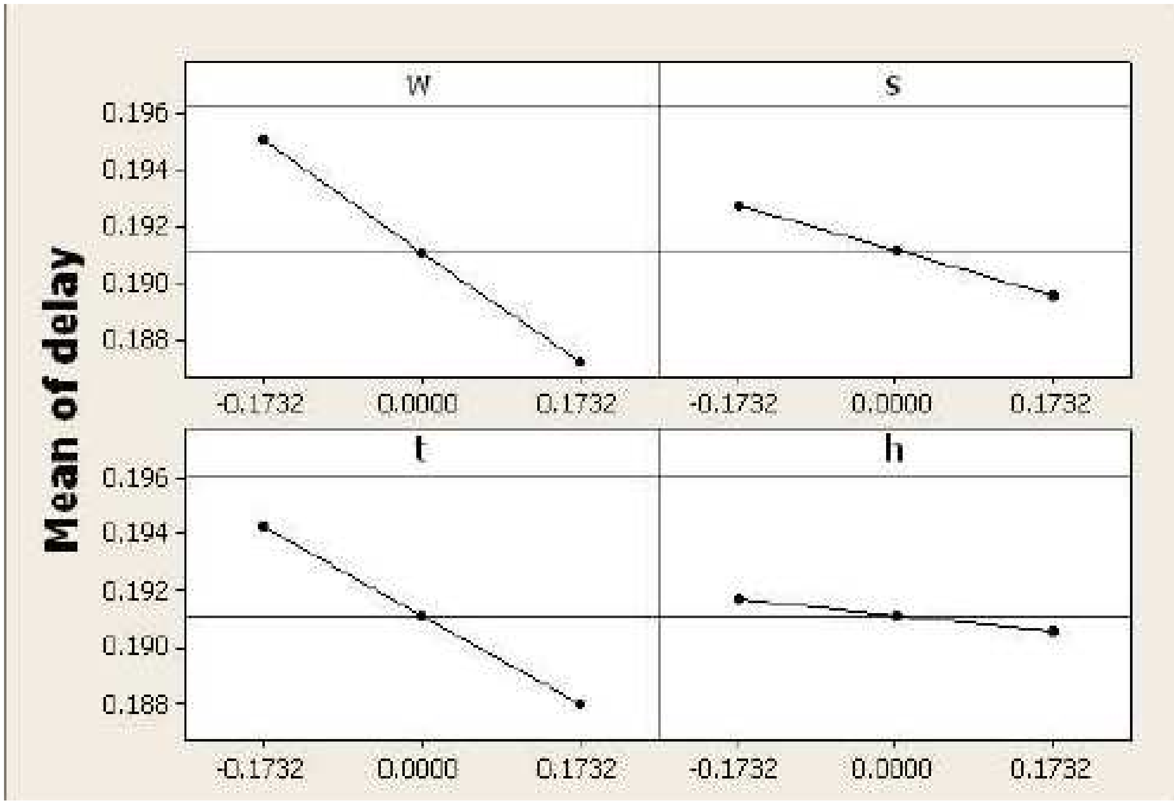
- Statistical Significance parameter (F):

$$\frac{\int [\hat{\mu}(\xi_i) - \mu]^2 d\xi_i}{\sigma^2}$$

- We calculate the "F" parameter using ANOVA
- Another Important parameter found using ANOVA is: R^2
- Based on these parameters, the algorithm decides whether the input parameter is significant or not.

Running ANOVA

- Input - Delay equation from PCM
- Apply Primary Screening – Compute delay gradient



Algorithm for DMA Implementation – Step 5: Variable Reduction

- Secondary Level of Screening - ANOVA
- Identify Insignificant terms based on F- value
- Remove insignificant terms
- Generate reduced analytical equation such that R^2 is at least 98.5%

An Example

Delay for a single RC segment of a global interconnect for 0.13um technology

$$
\begin{aligned}
delay = \;& 19.65 - 2.28\xi_1 - 0.9\xi_2 - 1.82\xi_3 - 0.32\xi_4 \\
& +0.28(\xi_1^2 - 1) + 0.1(\xi_2^2 - 1) + 0.12(\xi_3^2 - 1) \\
& +0.05(\xi_4^2 - 1) + 0.17(\xi_1\xi_2) + 0.03(\xi_1\xi_4) \\
& +0.2(\xi_2\xi_3) - 0.17(\xi_2\xi_4) + 0.17(\xi_3\xi_4)\ ps
\end{aligned}
$$

Mean = 19.62ps
Variance = 3.15ps

In this case, ANOVA gives us terms that are insignificant as follows:

$$\xi_4, \xi_2^2, \xi_4^2, \xi_1\xi_2, \xi_1\xi_3, \xi_1\xi_4, \xi_2\xi_4, \xi_3\xi_4$$

After removing these terms, the reduced equation is:

$$
\begin{aligned}
delay = \;& 19.65 - 2.28\xi_1 - 0.9\xi_2 - 1.82\xi_3 + 0.28(\xi_1^2 - 1) \\
& +0.12(\xi_3^2 - 1) + 0.2(\xi_2\xi_3)\ ps
\end{aligned}
$$

Mean = 19.64ps
Variance = 3.13ps

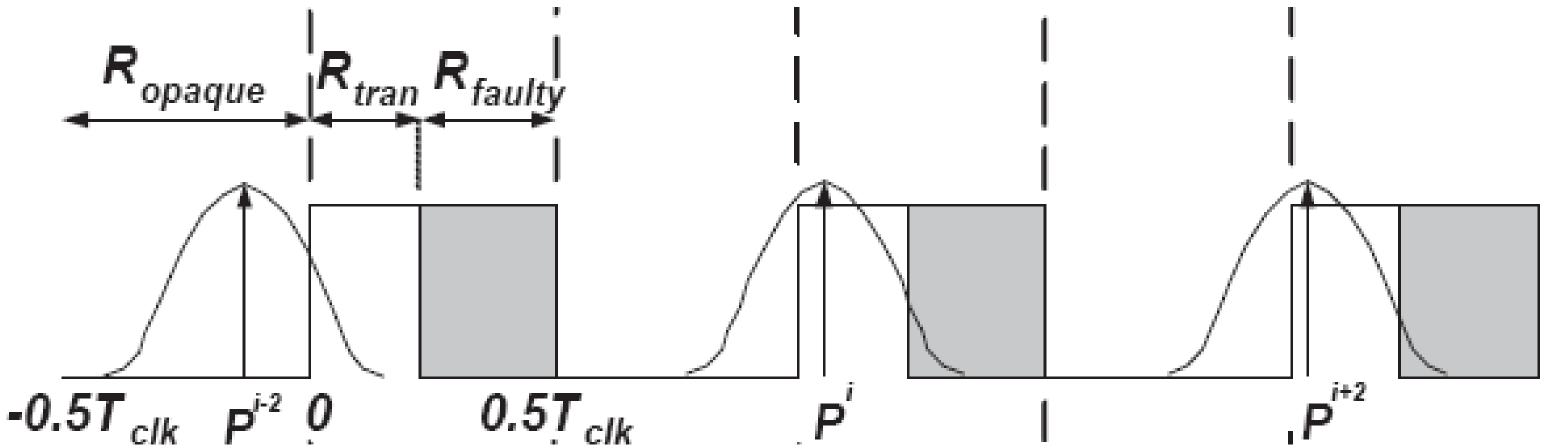
Overview

- A Novel Delay Metric based on ANOVA
- Statistical Timing Analysis of Pipelined Global Interconnect
- Reliability Aware Global Interconnect Sharing
- Experimental Results

Statistical Timing Analysis of Pipelined Global Interconnect

- Global Interconnect spans several clock cycles
- Interconnect Pipelining is mainly used to increase throughput.
- Two types of pipelining
  - Register Based (Edge Triggered)
  - Latch Based (Level Triggered)
- Latch pipelining has performance advantage over Register based pipelining.

Statistical Timing Analysis of Pipelined Global Interconnect

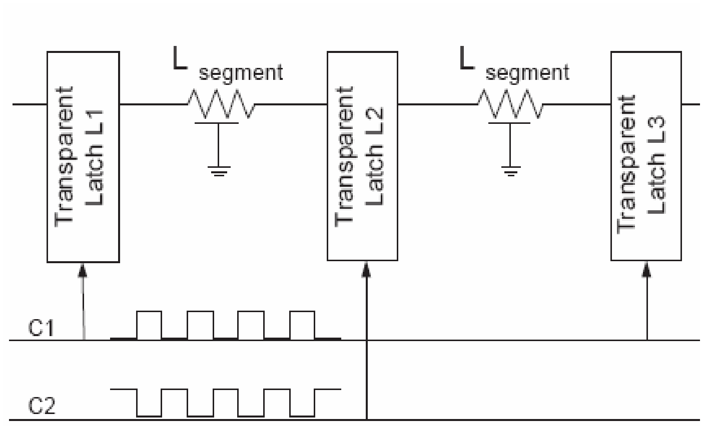


- Propagation delay (p) of present stage is not independent of previous stage
- Therefore, based on region where previous stage lies, p of previous stage lies, p of present stage is written as:

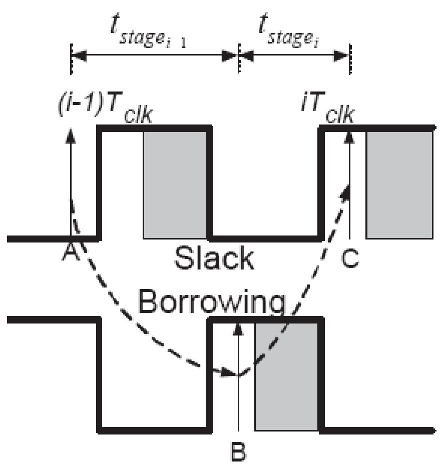
$$p_i = \begin{cases} \tau_{wire} + \tau_{data} - 0.5T_{clk} & p_i \in R_{opaque} \\ p_{i-1} + \tau_{wire} + \tau_{prop} - 0.5T_{clk} & p_i \in R_{tran} \end{cases}$$

Dual- Phase Clocking Scheme

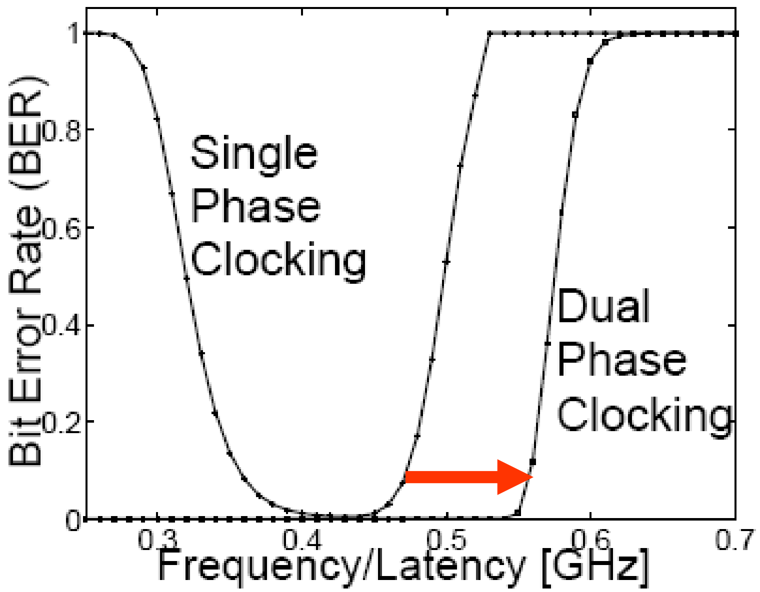
- Two sequentially adjacent clocks are driven using two clocks with 180 phase difference

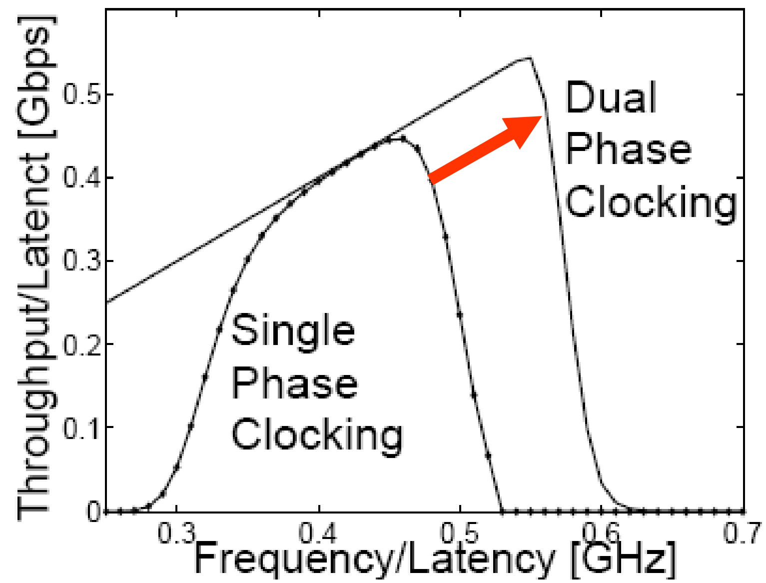


- Advantages – More Flexibility in timing, Higher performance, Robust to clock variations,



Result showing advantages of dual- phase clocking





- Less Bit Error Rate
- More Throughput

Overview

- A Novel Delay Metric based on ANOVA
- Statistical Timing Analysis of Pipelined Global Interconnect
- Reliability Aware Global Interconnect Sharing
- Experimental Results

Reliability Aware Global Interconnect Sharing – Basics

- Different transfers needs to be scheduled as when they will be sent on the interconnect
- A feasible schedule - Transfer starts after its arrival time and completes before specified deadline
- To ensure correct transmission, BER must be kept low.
- Thus, need for a probabilistic methodology to find a schedule for sampling transfers on the interconnect

Reliability Aware Global Interconnect Sharing – Preliminaries

- Total Confidence Level(Ψ_f) → Sampling Confidence (ψ_s)
  → Transmission Confidence (ψ_b)

- Transmission Confidence: $\psi_b = (1 - BER)$
- Sampling Confidence: $\psi_s = 100 * \int_{-\infty}^{\pi_x} f(x)\,dx$

  where f(x) is pdf of x and π_x is ψ^{th} percentile of x
- $\Psi_f = \psi_s * \psi_b$

Reliability Aware Global Interconnect Sharing – Problem Statement

- Problem Statement: "Given a set of arrival times and their corresponding deadlines, and a global interconnect clock period, the sampling time set is to be found such that total confidence level (Ψ_f) is maximized".

$$\begin{array}{lrcll} & \max & & \Psi_f & \\ \texttt{subject to:} & \omega_i & \leq & \omega_{min} & \forall i \in n \\ & \mathrm{d}_i & \leq & \lambda_i & \forall i \in n \\ & \psi_{s_i} & \geq & 50\% & \forall i \in n \end{array}$$

Where ω is the BER, d is the delivery time, λ is the deadline time

Scheduling Algorithm

p = FindLatestArrivalTime(B);	//find latest arrival time of an event st //BER is less than B

S = ComputeSlack(ω_{min}, AT);	//compute slack of each event according
	// to mininum BER and mean arrival
	// time (AT)

C = FindSamplingClockCycle(S);	//find latest clock cycles (C) a event can //be sampled

ST = FindSamplingTime(C);	//Find sampling times according to C
	//such that there is no conflict in
	//scheduling

ω = FindBER(ST);	//compute BER according to the //sampling time
Ψ_f = ComputeConfidence(ω, ST);	//compute total confidence for an event

Overview

- A Novel Delay Metric based on ANOVA
- Statistical Timing Analysis of Pipelined Global Interconnect
- Reliability Aware Global Interconnect Sharing
- Experimental Results

Experimental Results - DMA

Number of segments	Mean delay (ps)			Delay Variation (ps)			Number of Spice Runs			Error%	
	MC	PCM	DMA	MC	PCM	DMA	MC	PCM	DMA	μ	σ
2	40.06	39.92	40.11	3.59	3.45	3.63	10000	45	38	0.12	1.11
4	163.67	163.94	164.04	9.24	9.37	9.42	15000	153	86	0.22	1.94
8	496.36	496.66	494.62	26.59	26.36	27.16	20000	561	261	0.35	2.15
16	1616.29	1617.15	1625.82	43.45	42.81	44.86	25000	2145	912	0.59	3.26

- MC: Monte Carlo
- PCM: Probabilistic Collocation Method

Experimental Results - Scheduling

Table 5.4: Comparison of Deterministic and probabilistic Scheduling

	μ_A ps	λ ps	δ		ζ		Ψ_f						Gain %			
							μ Sampling		$\mu + 3\sigma$ Sampling		probabilistic Sampling		μ Sampling		$\mu + 3\sigma$ Sampling	
			SP	DP	SP	DP	SP	DP	SP	DP	SP	DP	SP	DP	SP	DP
Case 1																
E1	70	2850	1	1	385	650	44.64	43.13	69.99	70.62	99.47	99.79	122.77	131.37	42.09	41.29
E2	395	3700	3	1	1240	1110										
E3	690	3450	1	2	955	1570										
E4	1020	4100	2	2	1810	2030										
Case 2																
E1	70	2600	0	0	90	90	49.47	40.65	0	81.54	73.52	99.74	48.62	145.36	-	22.30
E2	395	2900	0	0	405	650										
E3	690	3150	0	1	695	1110										
Case 3																
E1	70	2600	0	0	-	90	×	43.13	×	0	×	89.38	-	107.23	-	-
E2	395	3450	2	2	-	1570										
E3	690	2850	-1	0	-	710										
E4	1020	3100	-2	0	-	1165										

SP → Single Phase clocking DP → Dual Phase clocking × → Deadline Constraint Violation

Slack > 0

Slack = 0

Slack < 0

Questions?

Thank you

Contact Information:

Navneeth Kankani

kankani@ece.arizona.edu