Convergence-Provable Statistical Timing Analysis with Level-Sensitive Latches and Feedback Loops

Lizheng Zhang, Weijen Chen, Yuhen Hu, Charlie Chen

Cadence, University of Wisconsin Madison,

National Taiwan University

DFM: Design for Manufacturability

Designer?

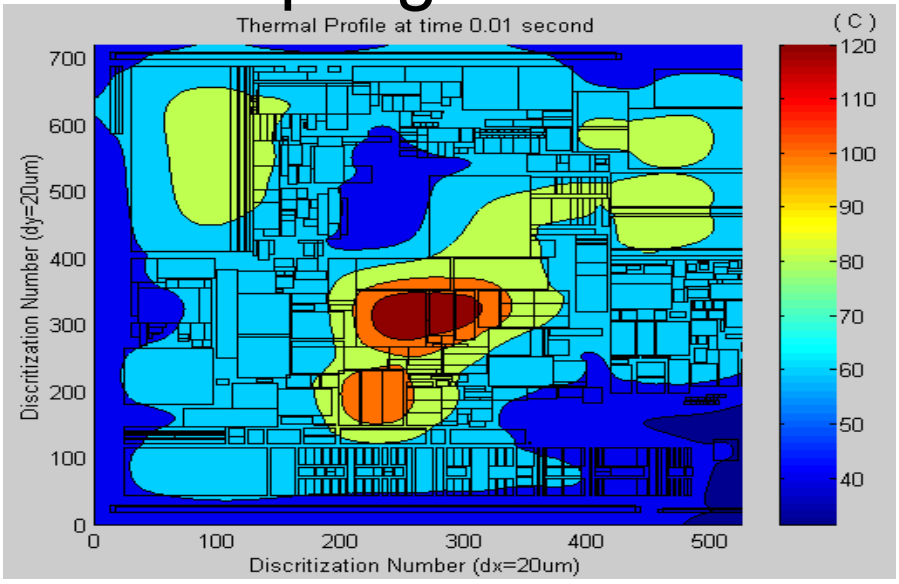
????

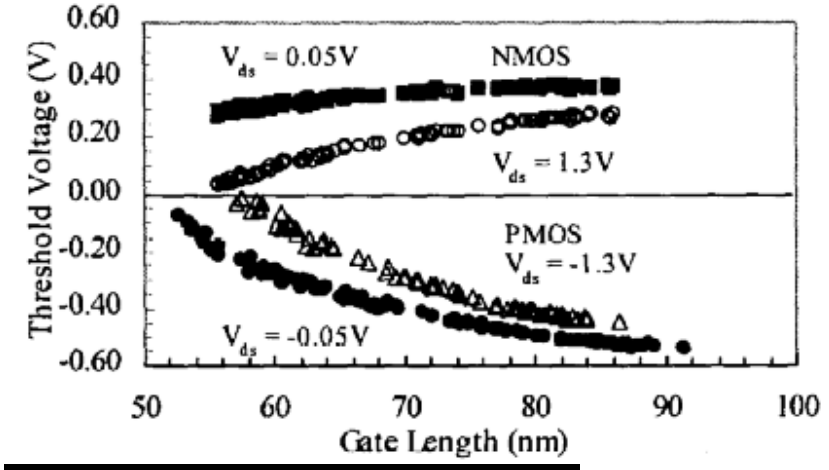
whatever, let's throw a die

(Monte Carlo)

General Variation Sources

- Inter- and Intra-die Process Variation-W, L,…..
- PVT Variations:
  - Power Fluctuation: IR-drop,….
  - Non-uniform Temperature Distribution
  - Vt Variation
- Crosstalk: capacitive, inductive, and substrate coupling

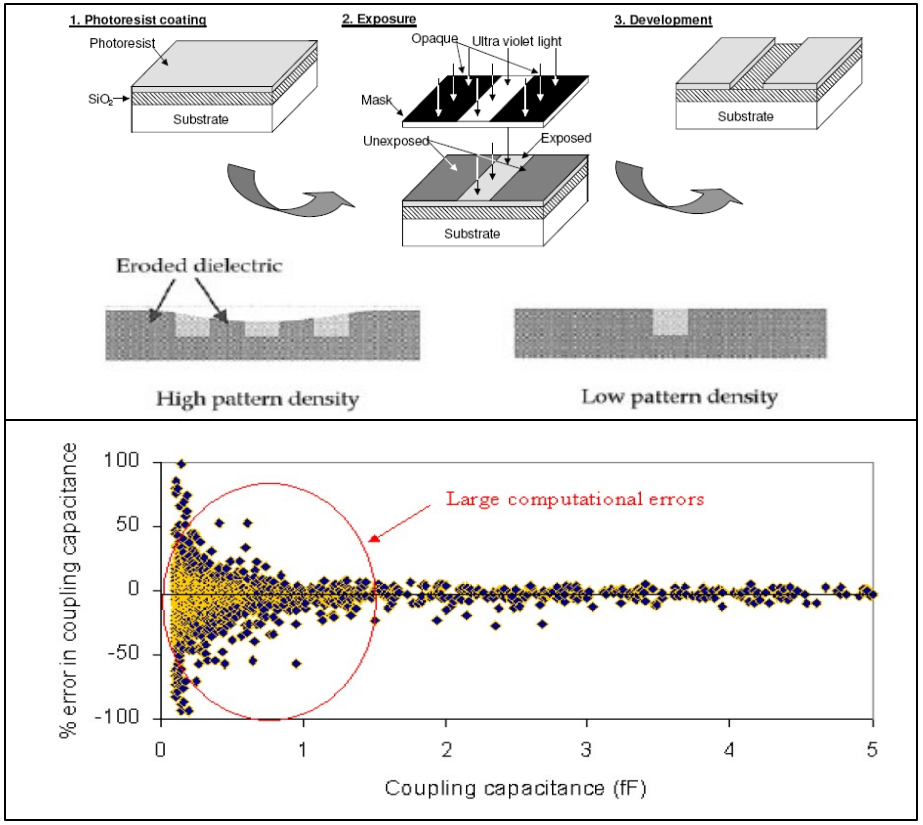


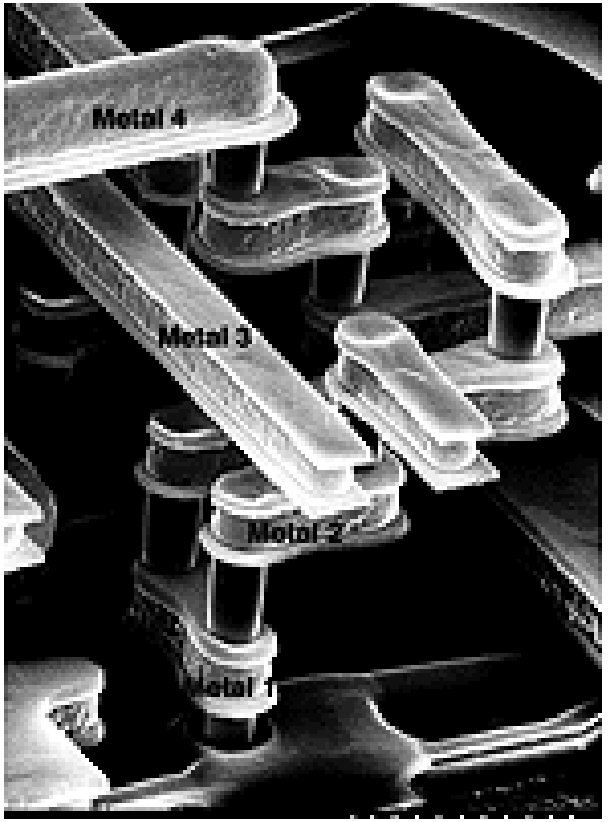


Courtesy from Intel

Interconnect Geometry Variations

- Process induced variations significantly change the interconnect geometry, which directly affects interconnect models, especially parasitic capacitances.
  - Optical lithography (optical proximity effects)
  - Chemical-mechanical planarization (CMP)





Picture courtesy of TSMC, Intel Corporation.

Variations Trends

(courtesy from IBM)

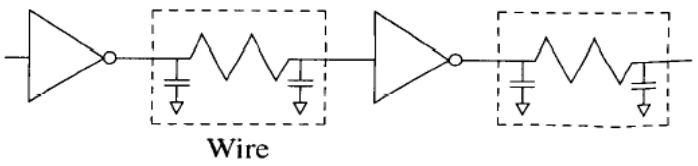
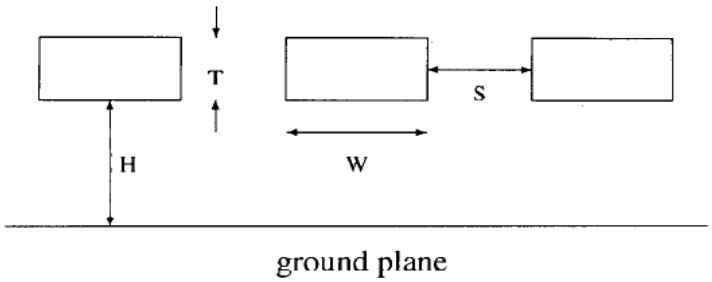
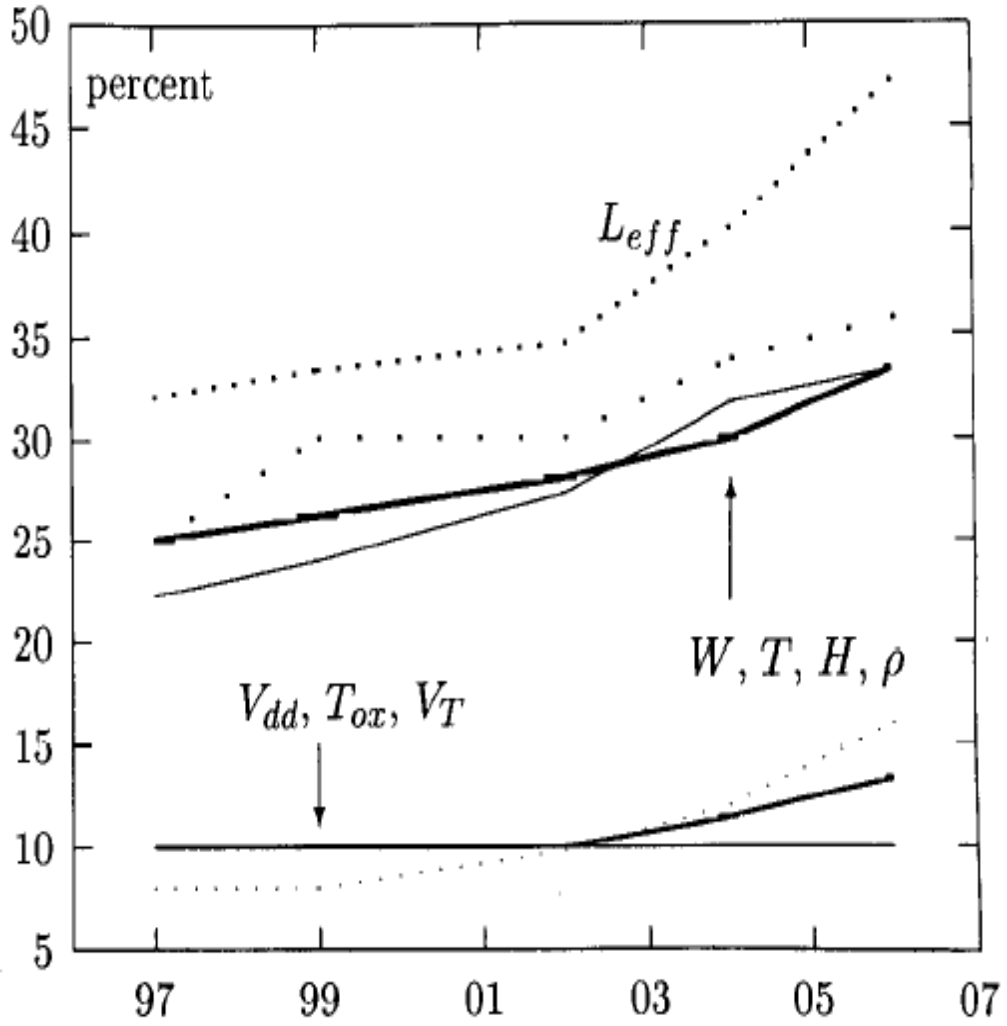


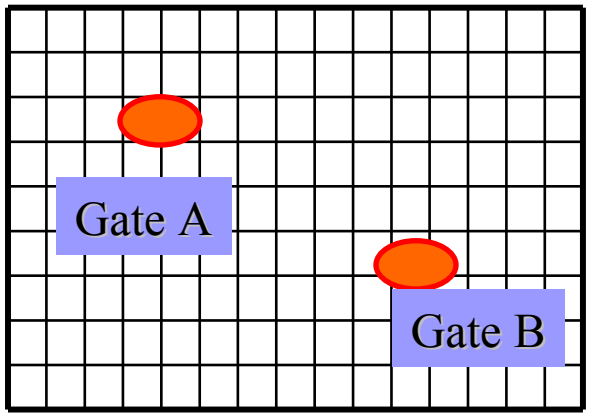
Figure 1: Canonical circuit.

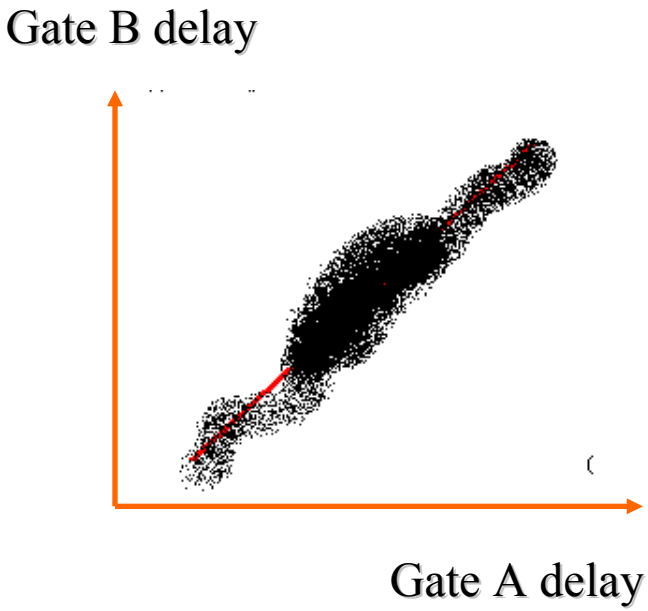


Year	L_{eff} nm	T_{ox} nm	V_{dd} V	V_T V	W μ	H μ	ρ $\frac{m\Omega}{\square}$	L_{max}
1997	250	5	2.5	0.5	0.8	1.2	45	*2123*
1999	180	4.5	1.8	0.45	0.65	1.0	50	*1920*
2002	130	4	1.5	0.4	0.5	0.9	55	*1670*
2005	100	3.5	1.2	0.35	0.4	0.8	60	*1526*
2006	70	3	0.9	0.3	0.3	0.7	75	*1303*

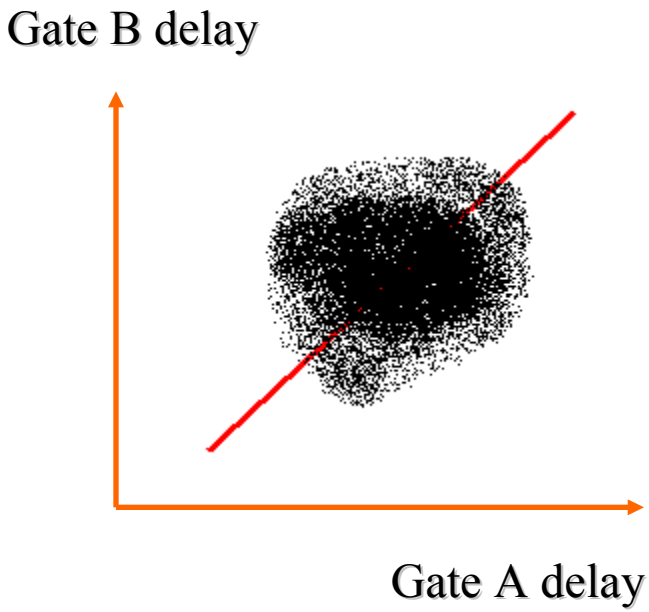


Correlated and Uncorrelated (independent) Variation



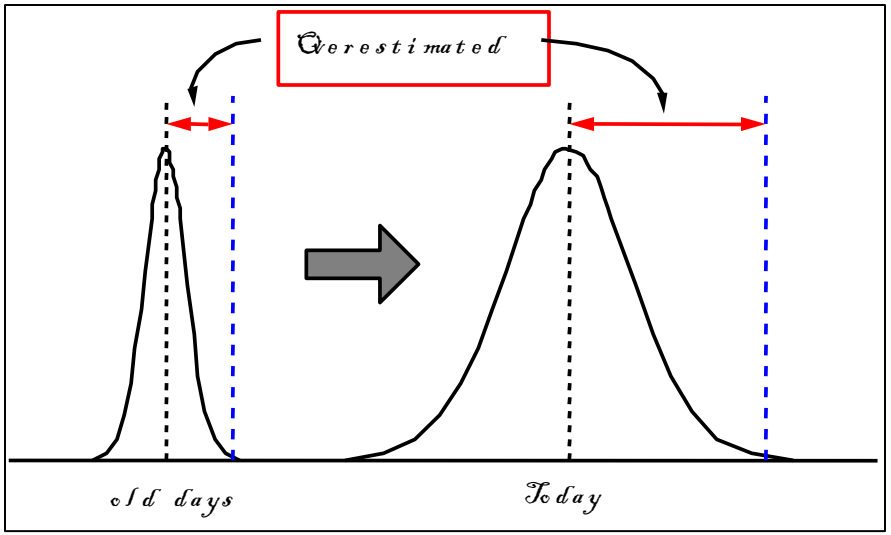


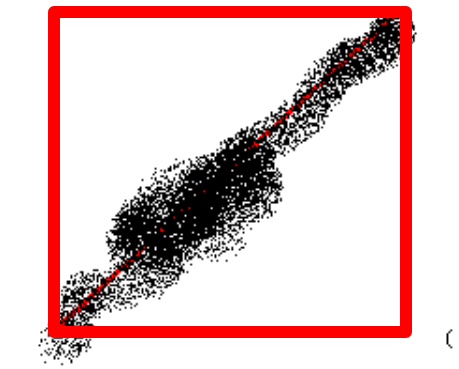
Correlated



Uncorrelated

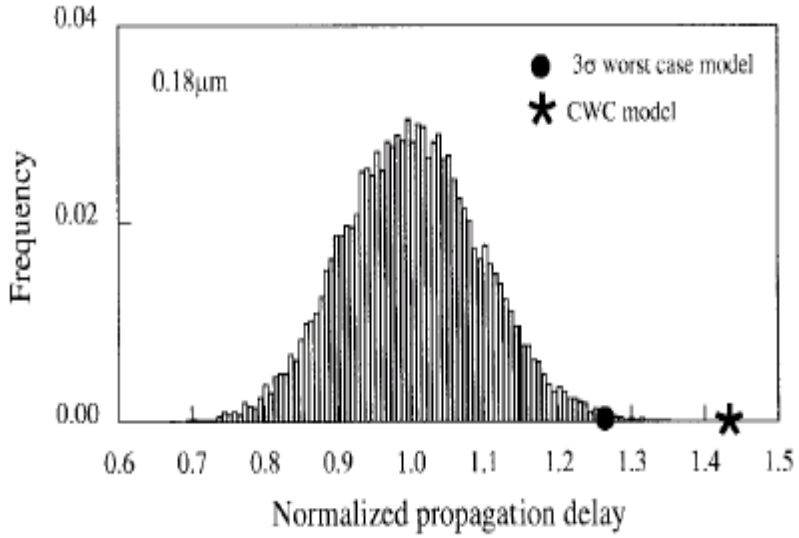
Pessimistic Worst Case Corner Analysis (ignore correlation)





- Pessimistic estimation:
  - It is very rare to have all gates behave as the worst case
  - Larger variation make overestimation more significant

Worst-case Timing Analysis Example



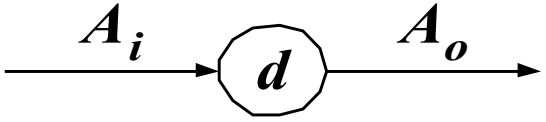
- 16-bit Adder
  - Longest path Delay
- Monte Carlo
  - $3\sigma = 1.25$
- Worst-case Timing
  - Delay=1.45
- Too **pessimistic**!
  - 20% overestimation

"Impact of Unrealistic Worst Case Modeling on the Performance of VLSI circuits in Deep Submicron Region CMOS Technologies"
A.Nardi, A.Neviani, E.Zanoni,M.Quarantelli, *IEEE '99*

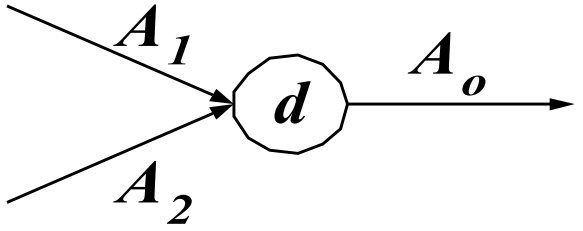
Existing Statistical Timing Analysis

- To reduce the pessimism, statistical method is applied to treat variation-induced delay variation
- Existing statistical timing method is
  - mainly focused on combinational circuit
  - flip-flop based sequential circuit
  - Latched circuit without feedback loops
- In this paper, we focus on SSTA of Latched circuit *with feedback loops*

Basic Timing Analysis



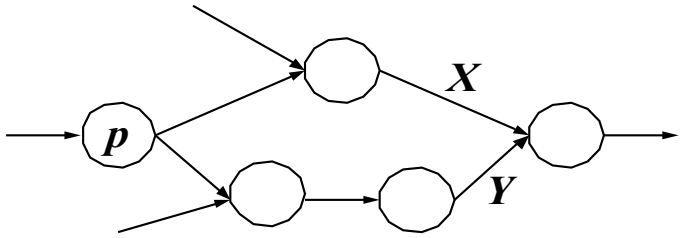
$$A_o = A_i + d$$



$$A_o = \max(A_1, A_2) + d$$

"Statistical timing analysis using bounds and selective enumeration" *Agarwal, A.; Zolotov, V.; Blaauw, D.T.;* TCAD, Vol. 22(9) , Sept. 2003, P1243 - P1260

Path correlation



$$X = \mu_X + \alpha_X R_X + \sum \beta_{Xi} G_i$$

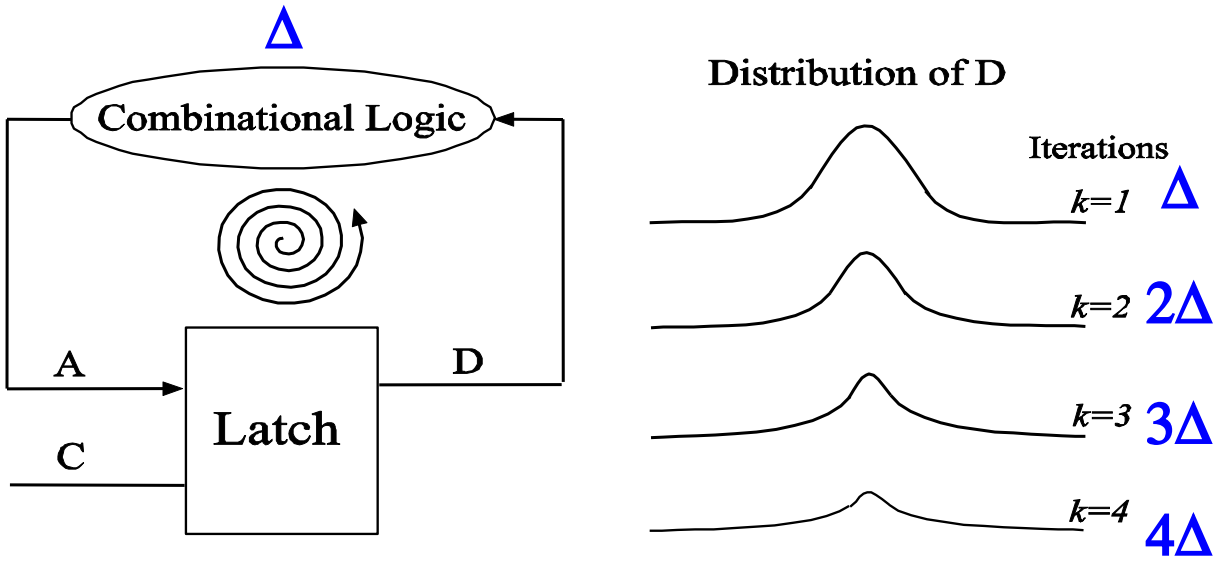
$$Y = \mu_Y + \alpha_Y R_Y + \sum \beta_{Yi} G_i$$

$$Z = aX + bY = \mu_Z + \alpha_Z R_Z + \sum \beta_{Zi} G_i$$

$\alpha_Z = \sqrt{a^2\alpha_X^2 + b^2\alpha_Y^2}$ since R_X and R_Y are asummed to be independent

- But R_X and R_Y will have correlation since they are both dependent on R_p
- We remember the past history

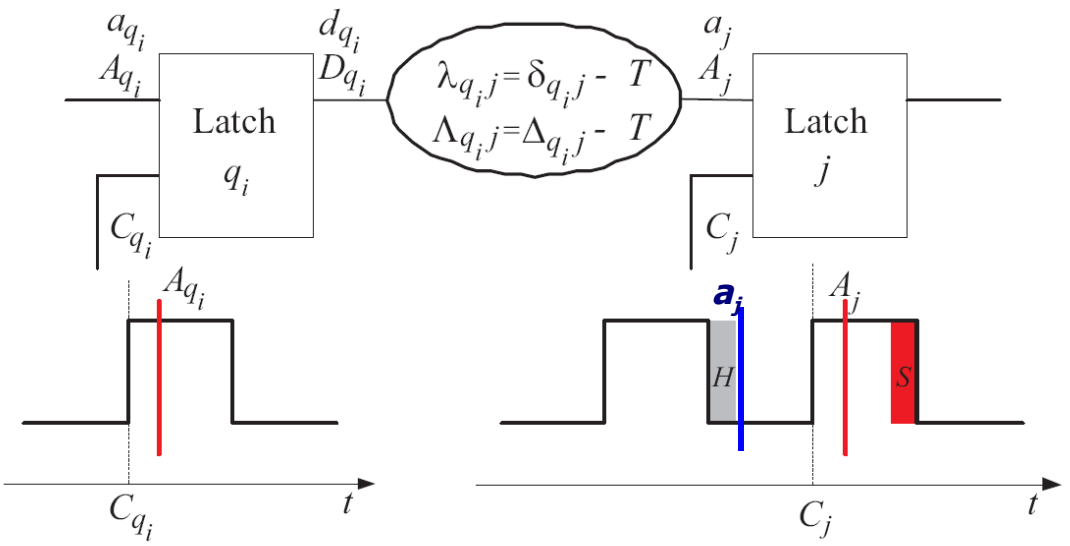
Latch with Feedback Loop



- Convergence is a problem for iterative latch timing:
  - Even if mean of **D** converges, variance of **D** will still be possible to diverge
  - No solution available in the statistical timing so far

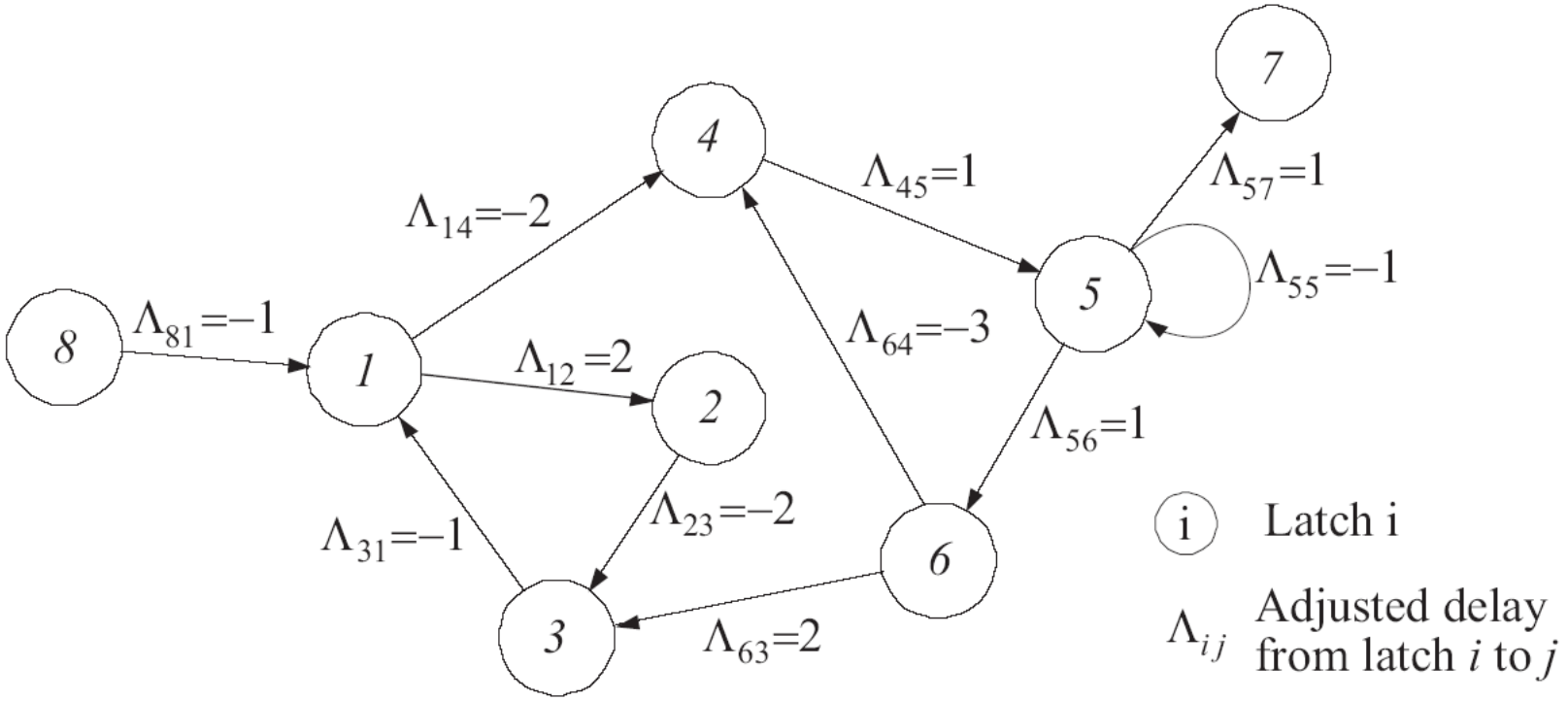
Latch (active high) Timing Diagram

- q_i is one of the input latch for latch $_j$
- C_j is clock rising edge time
- A_j is latest arrival time
- a_j is earliest arrival time
- Λ is the maximum delay
- λ is the minimum delay
- A_j is the maximum of all possible quantity of
  - $max(A_{qi},C_{qi})+\Lambda_{qi}$ (delay)
- Set time constraints
  - $A_j<C_j+T_h-s$
- Hold time constraints
  - $a_j>C_j-T_l+h$



Reduced Timing Graph

Delay Adjusted by Clock Cycle Λ=D-T



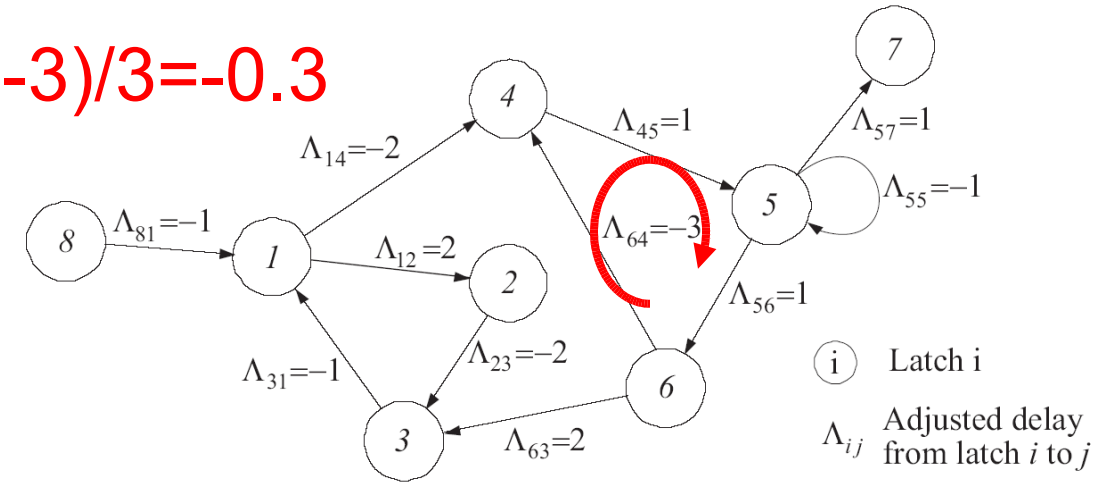
Key Concept(1)—Cycle Mean

Every loop m with p_m latch nodes in it will have a **cycle mean**(G_m) defined as the average edge weight in the loop:

$$G_m = \frac{1}{p_m} \sum_{e_{ij} \in m} \Lambda_{ij}$$

and p_m is usually called *cycle length*.

Cycle mean=(1+1-3)/3=-0.3



Key Concept(2)—Critical Cycle Mean

The **critical cycle mean** of the reduce timing graph, G_c, is the larger value between 0 and the largest cycle mean among all possible loops:

$$G_c = max(0, G_1, G_2, ...) \geq 0$$

where $G_1, G_2, ...$ are cycle means for all loops $1, 2, ...$ in the reduced timing graph.

Key Concept(3)—Iteration Mean

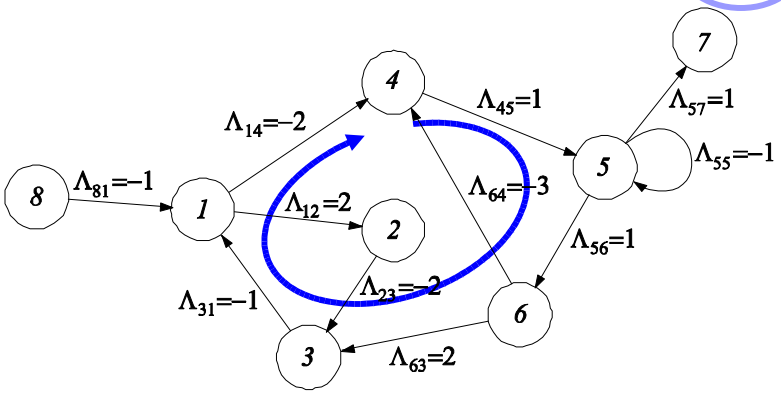
At every iteration k, each latch node i in the reduced timing graph will have an **iteration mean** defined as the latch's average latest data arrival time per iteration:

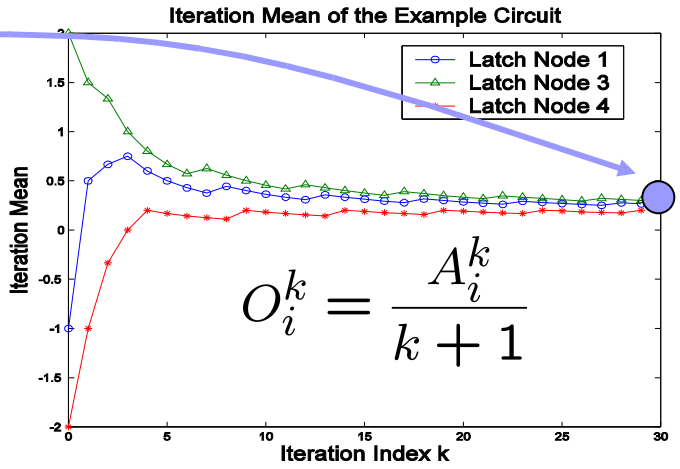
$$O_i^k = \frac{A_i^k}{k+1}$$

Iteration Mean Converges into Critical Cycle Mean

Critical cycle mean:

$G_c = (-1-2+1+1+2)/5 = 0.2$





- Guarantee converging to the critical cycle mean

$$G_c = max(K, O_1^\infty, O_2^\infty, ..., O_N^\infty)$$

Timing Yield

- Circuit will have non-zero yield only if its critical cycle mean is non-positive
  - If positive critical cycle mean, there will be latches whose arrival time will go to infinity and violate the setup time constraints after sufficient iterations
- True yield of the circuit will have to consider setup and hold time violation probability when all latches have finite arrival time

Algorithm

- Timing core is statistical
- Iteration mean is random variable
- Simultaneous propagate mean and variance
- Proven to converge in finite iteration

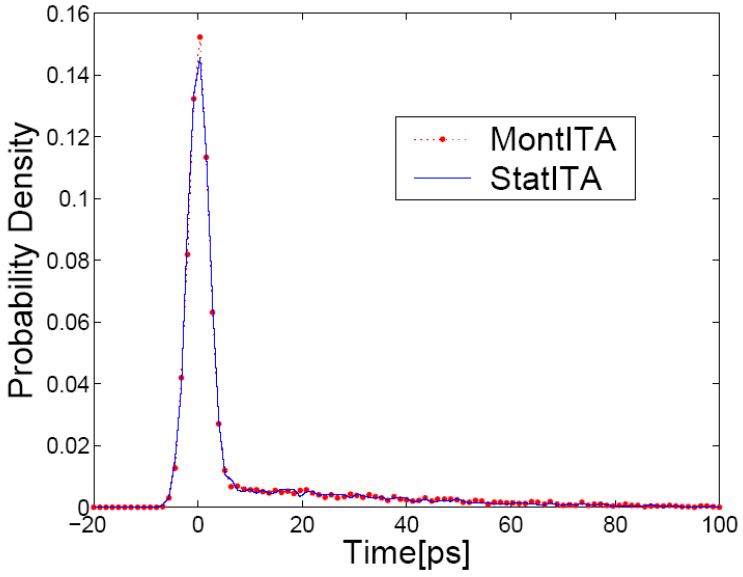
```
1: procedure StatITA(ClockCycle T)
2:     for (each latch i) do                                  ▷ initialization
3:         C_i = setClockArrivalTimeForLatch(i)
4:         a_i^0 = 0; A_i^0 = 0; O_i^0 = 0;
5:     end for
6:     k = 1;
7:     repeat                                                 ▷ iteration starts
8:         {a_i^k, A_i^k}=statistiticalTiming(a_i^{k-1}, A_i^{k-1});
9:         done = true;
10:        for (each latch i ) do
11:            O_i^k = A_i^k/(k + 1);
12:            if |O_i^{k-1} − O_i^k| ≥ threshold then
13:                done = false;                              ▷ Not converged yet
14:            end if
15:        end for
16:        k = k + 1;
17:    until done                                             ▷ iteration ends
18:    k = k-1;
19:    G_c = 0; s_c^∞ = −∞; h_c^∞ = −∞;
20:    for (each latch i) do
21:        G_c = max(O_i^k, G_c);
22:        s_c = max(A_i^k − (C_i + 0.5T − S), s_c^∞);
23:        h_c = max((C_i − 0.5T + H) − a_i^k, h_c^∞);
24:    end for
25:    Y = Pr{max(G_c, s_c^∞, h_c^∞) = 0};                    ▷ circuit yield
26: end procedure
```

Accuracy and Speed-up Compared with Monte Carlo

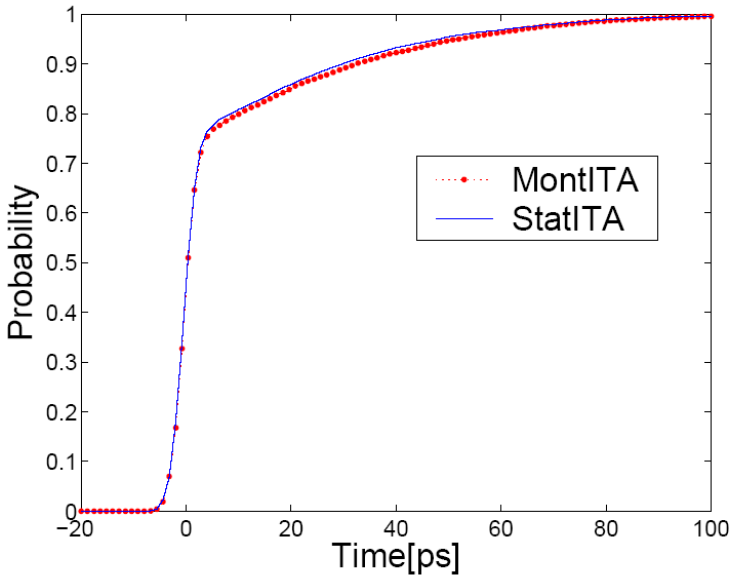
Circuits	Gates	Latches	T_{97}[ps]			CPU Time[s]		
			StatITA	MontITA	Error	StatITA	MontITA	Speedup
s298	130	14	443	452	2.0%	2.14	320	150x
s344	137	15	673	663	1.5%	0.72	295	410x
s382	164	21	548	559	2.0%	1.28	379	296x
s526	196	21	465	469	0.9%	5.76	694	120x
s641	173	19	999	998	0.1%	1.17	372	320x
s820	279	5	777	788	1.4%	1.35	692	513x
s953	401	29	862	858	0.5%	3.32	1041	314x
s1423	616	74	2088	2051	1.8%	16.0	2083	130x
s5378	1517	179	764	780	2.1%	106	12372	117x
s9234	1827	211	859	858	0.1%	101	19073	189x
s13207	3516	638	1242	1246	0.3%	231	41571	180x
s15850	3889	534	1189	1199	0.8%	540	61044	113x
s38417	11543	1636	1544	–	–	1468	200hr*	490x*
s38584	12389	1426	1430	–	–	1209	303hr*	903x*
Average	–	–	–	–	**1.1%**	–	–	**303x**

Table 1: 97% yield clock cycle(T_{97}) and CPU time comparison between *StatITA* and *MontITA*. (*)Estimation is from 100 repetitions and the accuracy of StatITA is not evaluated for these circuits.

Critical Cycle Mean

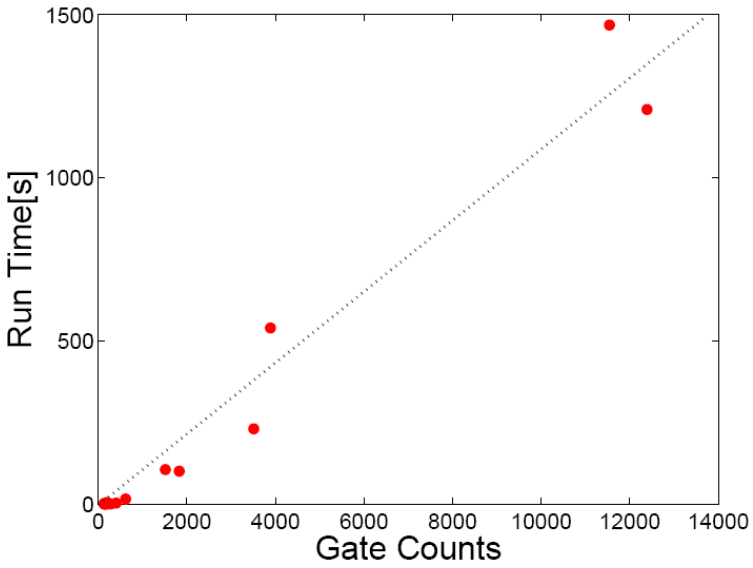


(a) $p.d.f.$ of G_c for s526

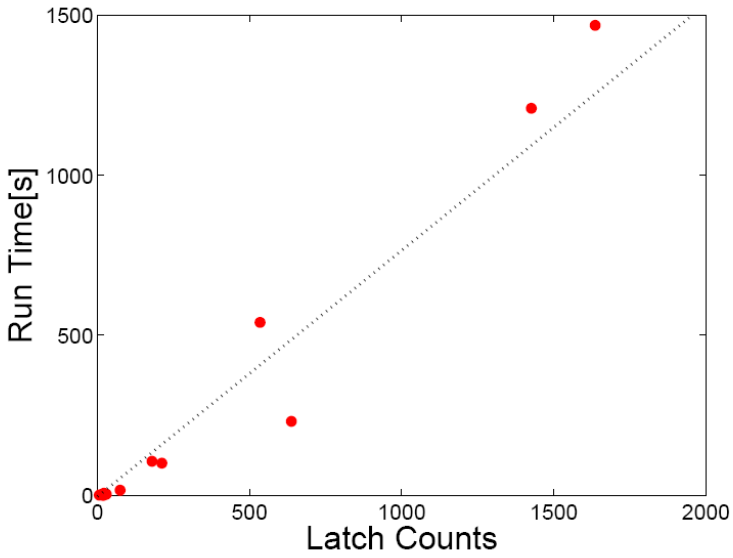


(b) $c.d.f.$ of G_c for s526

Complexity



(a) CPU time v.s. gate number



(b) CPU time v.s. latch number

Conclusion

- Statistical timing algorithm for level-sensitive latches with feedback loops
  - Iterative algorithm with guaranteed convergence
  - Simple relationship to predict the circuit timing yield