Timing Yield Estimation with Clock Network Correlations by Propagating Discrete Probability Distributions

Lee-eun Yu[‡], Changsik Shin[‡], Jing-Jia Liou[†] and Youngsoo Shin[‡]
[‡]Department of Electrical Engineering, KAIST, Daejeon 305-701, Korea
[†]Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan

***Abstract*— Timing yield, in conjunction with other types of yield, directly affects profit; under-estimation is as bad as over-estimation, because large amount of time is unnecessarily spent to increase small amount of timing yield. The correlation that stems from clock network, when ignored, turns out to be one of the reasons of under-estimation in clocked sequential circuit. Three sources of topological correlation are identified; the key problem is to determine the correlations we can ignore without sacrificing accuracy so that we keep run time within control, which is addressed in this paper. A prototype tool was implemented with gate delay modeled as discrete probability distribution; experiments with benchmark circuits show that, compared to Monte Carlo simulation, speedup is 75× with 0.53% difference of timing yield on average.**

I. Introduction

Process variations, which naturally increase as technology scales down, are classified into die-to-die (D2D) and within-die (WID) variations. Traditionally, D2D variations have been handled by process corners and WID variations by linear combination of process corners, e.g. setup time is checked at worst corner (WC) by comparing earliest clock arrival time at $0.5 \cdot \mathrm{NC} + 0.5 \cdot \mathrm{WC}$, where NC is nominal corner, and latest data arrival time at WC. This approach is pessimistic since device-wise variations are not taken into account; it becomes excessively pessimistic as WID variations take increasing proportion of total variations, e.g. 35% in 130-nm and 60% in 70-nm technology [1].

Statistical static timing analysis (SSTA) [2]–[8] has been proposed to alleviate this limitation of static timing analysis (STA). All the timing parameters such as gate delays and arrival times (ATs) are modeled as random variables, thus associated with probability distribution functions (PDFs). Propagating arrival times is the same as STA except that adding and taking maximum (or minimum) is done in random variables. PDF is either modeled analytically or as a discrete distribution. Analytical approach typically assumes normal distribution for the sake of computational convenience, which however is a limitation; the approach using a discrete PDF [2] does not assume any particular type of distribution, thus is robust, but generally slower than analytical approach.

Since PDF is derived for maximum arrival time at primary outputs (for checking max-time constraints), for a given required arrival time (RAT), SSTA reports the percentage of AT that is lower than RAT, which is called a timing yield (other constraints such as min-time, clock pulse width, etc have to be taken into account as well). In estimating timing yield of clocked sequential circuits, the main challenge is how to incorporate correlations, which arise from various sources as described in Section II-C. The correlation that arises from a clock network, i.e. due to common clock path, is typically ignored, especially in discrete PDF approach due to excessive run time. It however yields non-negligible error. Fig. 1 shows the curves of timing yield, while we vary clock cycle time, with and without clock network correlation in b03, which is one of the ITC benchmark circuits. Maximum error is 13.5% when clock cycle time is 495 ps. Similar experiments were performed in 45-nm technology with various circuits, which are summarized in Table I; maximum error is 5.27% on average. Note that this is not small error, even though it may sound to be. Timing yield, in conjunction with other types of yield, directly affects profit. Large amount of time is typically spent to increase small amount of timing yield, which however may be actually unnecessary since full amount of timing yield is not extracted without considering clock network correlation.

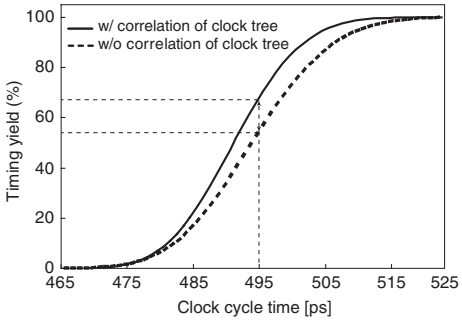

Fig. 1. Difference of timing yield with and without clock network correlation in b03 benchmark.

TABLE I
Maximum difference of timing yield with and without clock network correlation

Circuits			Max yield
Name	# CLs	# F/Fs	difference (%)
b03	171	30	13.50
b11	609	31	4.54
b13	381	53	6.89
s208	96	9	4.04
s298	134	14	3.61
s838	340	32	4.47
s1423	713	74	1.08
s5378	1386	163	6.80
s13207	3182	490	2.48
Average			**5.27**

978-1-4244-2934-9/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Korea Advanced Institute of Science and Technology. Downloaded on December 10, 2009 at 00:10 from IEEE Xplore. Restrictions apply.

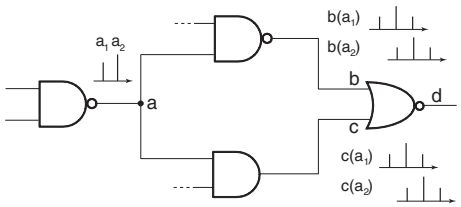

Fig. 2. Sampling-evaluation to handle correlation due to reconvergent fanout.

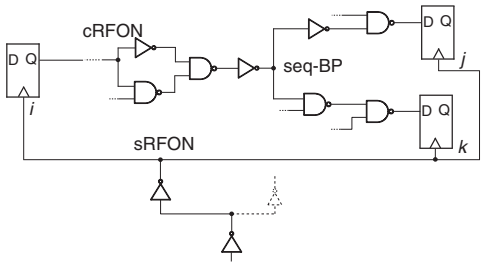

Fig. 3. Three sources of topological correlation in clocked sequential circuits.

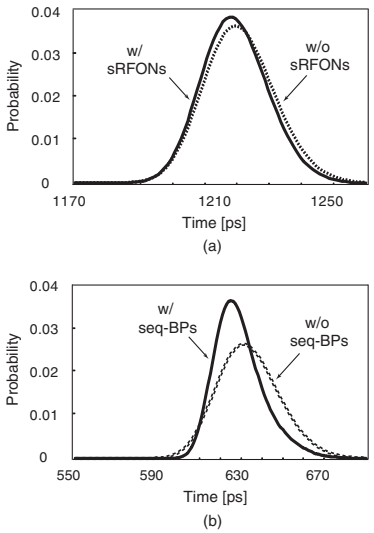

Fig. 4. PDF of maximum arrival time at primary outputs (a) with and without sRFONs in s298 benchmark, and (b) with and without seq-BPs in b13.

SSTA based on discrete PDF considering clock network correlation is proposed in this paper. Three sources of topological correlation are identified (Section II-C). Based on the observation that most correlations can be safely ignored without sacrificing accuracy, the key problem is to select the minimum number of correlations that are important for the accuracy of timing yield so that run time can be kept small, which we address in Section III. The prototype tool was implemented and compared to Monte Carlo simulation; speedup was 75× on average with less than 1% difference of timing yield.

II. Preliminaries

A. SSTA Based on Discrete PDF

The SSTA used for estimating timing yield is based on discrete PDF [2], even though analytical SSTA can be used as well. In this SSTA, the correlation due to reconvergent fanout node (RFON) is handled by an approach called sampling-evaluation. In Fig. 2, assume that RFON a has two events for its discrete PDF of AT, a_1 and a_2. Let a_1 generate a group of events $b(a_1)$ and a_2 generate $b(a_2)$ at node b; similarly for $c(a_1)$ and $c(a_2)$ at c. We combine $b(a_1)$ and $b(a_2)$ to derive AT PDF at b, and $c(a_1)$ and $c(a_2)$ for AT PDF at c. To derive AT PDF at d, however, event groups that originate from the same source event have to be considered; we take $b(a_1)$ and $c(a_1)$ and derive one event group at d, take $b(a_2)$ and $c(a_2)$ for another event group, and then combine two event groups. Note that if we simply derive a single event group for b (without explicitly deriving $b(a_1)$ and $b(a_2)$) and another for c, and use them to derive a PDF for c, we implicitly consider $b(a_1)$ and $c(a_2)$, and $b(a_2)$ and $c(a_1)$, which causes an error.

The sampling-evaluation is a main source of large run time, especially when RFONs are nested and the number of gates from RFON to the gate where fanouts converge is large.

B. Timing Yield

The timing yield for max-time constraints can be expressed by

$$Prob\left[\max_{\forall i \rightsquigarrow j}(T_{cq,i}+D_{i,j}+T_{su,j}+S_i-S_j) \;\le\; P\right], \qquad (1)$$

where P is clock cycle time; $T_{cq,i}$ is clock-to-Q delay of launching flip-flop i; $D_{i,j}$ is the maximum delay of combinational block between i and j, $i \rightsquigarrow j$; $T_{su,j}$ is setup guard time of capturing flip-flop j; and S_i is clock arrival time at i. Note that all the parameters except P are random variables, thus are associated with discrete PDFs; S_i and S_j are correlated if they share the common path in the clock network; MAX operation in (1) implicitly assumes correlations between combinational blocks if they share common nodes.

The timing yield for min-time constraints can be defined similarly. The correlation between maximum path delay and minimum path delay is reported to be small [9], which allows two timing yields to be computed independently and then be combined.

C. Sources of Topological Correlation

In clocked sequential circuits, we identify three sources of topological correlation as illustrated in Fig. 3. Combinational RFON (cRFON) is RFON within combinational block, which we discussed in Section II-A; it affects the accuracy of $D_{i,j}$ in (1). Sequential RFON (sRFON) is a node in the clock network, where clock paths to launching flip-flop i and capturing flip-flop j branch out, i.e. clock path from a clock source up to sRFON is shared between i and j; it affects the accuracy of $S_i - S_j$ in (1). Sequential branch-point (seq-BP) is a node within combinational block, where combinational paths to two capturing flip-flops j and k branch out, i.e. $i \rightsquigarrow j$ and $i \rightsquigarrow k$ share the same combinational path up to seq-BP; it affects the

978-1-4244-2934-9/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Korea Advanced Institute of Science and Technology. Downloaded on December 10, 2009 at 00:10 from IEEE Xplore. Restrictions apply.

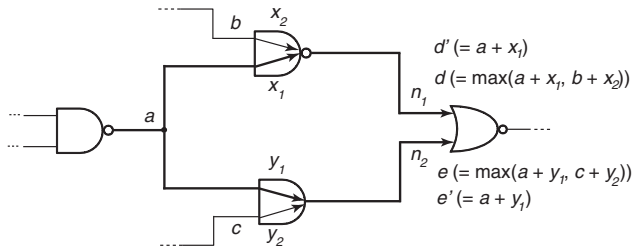

Fig. 5. Detecting effective cRFON.

accuracy of MAX operation in (1).

The effect of sRFONs on PDF of maximum arrival time at primary outputs is illustrated in Fig. 4(a) by comparing two PDFs, with and without correlations from sRFONs; the effect of seq-BPs are similarly shown in Fig. 4(b). This clearly shows the importance of sRFONs and seq-BPs as well as cRFONs in estimating timing yield.

III. Estimating Timing Yield in Clocked Sequential Circuits

A. Overall Algorithm

All three sources of topological correlation (cRFON, sRFON, and seq-BP) can be handled by sampling-evaluation (see Section II-A); we however need to select a handful of them to keep run time within control. For a given sequential circuit with its clock network (either placed or not), we first find all the cRFONs out of which we select those that are potentially important for the accuracy of timing yield, which we call *effective* cRFONs. We then select a set of effective flip-flops (flip-flops that are important for estimating timing yield), which implicitly determine effective sRFONs and effective seq-BPs. We perform SSTA while we handle selected correlation sources with sampling-evaluation, which finally leads us to compute timing yield (1).

B. Effective cRFONs

Consider cRFON shown in Fig. 5; each node is associated with a random variable for its AT; a random variable is assigned to each timing arc of NAND and AND gate to represent a delay. For two inputs of the NOR gate n_1 and n_2, we also derive approximate ATs (d' and e'), which exclusively consider the delay along the path from RFON to n_1 and n_2, respectively. Only when d and d' are similar (thus the path from RFON to n_1 is important in timing-wise) and e and e' are similar, we declare that cRFON is effective, thus is subject to sampling-evaluation.

The similarity between d and d' (between e and e' as well) should be defined in a statistical way, since both are random variables. Let their corresponding (discrete) PDFs be denoted by f_d and $f_{d'}$. We define their similarity by

$$S_{d,d'} = \sum_i \sqrt{f_d[i] \cdot f_{d'}[i]}, \tag{2}$$

where summation is done over all the values of AT i. Note that $0 \le S_{d,d'} \le 1$; it is 0 when there is no overlap between two PDFs and 1 when both are exactly the same. If $S_{d,d'} > \varepsilon_1$ and $S_{e,e'} > \varepsilon_1$ for some constant ε_1, cRFON in Fig. 5 is effective, otherwise its correlation is ignored.

Before we compare d and d' (and e and e'), we compare d and e. If $S_{d,e} < \varepsilon_1'$ for some other constant ε_1', i.e. if d and e are not close enough, their correlations are not important, thus cRFON is declared not effective, otherwise, we then check the cRFON via aforementioned process.

C. Effective Capturing Flip-Flops

Once we find all the effective cRFONs, we perform SSTA to calculate the max-time at capturing flip-flop j:

$$m_j = \max_{\forall i} \left(T_{cq,i} + D_{i,j} + T_{su,j} + S_i - S_j\right), \tag{3}$$

where MAX is taken over all the launching flip-flops that have paths to j. Out of all the capturing flip-flops, we select the one with the highest probability of its max-time being larger than clock cycle time, i.e. $Prob\,(m_j > P)$. Let m_j of that flip-flop be denoted by m^*. We then derive the similarity of m_j and m^* for each capturing flip-flop j, if it is larger than some threshold, we declare j to be effective, i.e. j is effective if $S_{m_j,m^*} > \varepsilon_2$.

Note that when calculating m_j, and thus m^*, from (3), we do not consider the correlation from sRFON, i.e. S_i and S_j are considered to be independent. The rationale behind this is that m_j thus computed is a conservative estimation in a sense that its PDF will be wider than it actually has to be because the correlations from sRFONs are ignored; this provides us fast calculation of (3).

D. Effective Launching Flip-Flops

To find effective launching flip-flops, we derive a slack of each launching flip-flop that can be reached (in backward topological order) from effective capturing flip-flops; those that cannot be reached from effective capturing flip-flops are declared to be not effective. The slack of launching flip-flop i is

$$s_i = r_i - a_i, \tag{4}$$

where r_i is its RAT and $a_i = S_i + T_{cq,i}$ is AT. The RAT is obtained by propagating the RAT r_j of each effective capturing flip-flop via SSTA, where r_j is set to its AT, i.e. $r_j = \max_{\forall i}\left(T_{cq,i} + D_{i,j} + T_{su,j} + S_i\right)$.

Those flip-flops with probability of negative slack being larger than some threshold are declared to be effective, i.e. i is effective if $Prob\,(s_i < 0) > \varepsilon_3$.

IV. Experimental Results

We performed experiments on a set of sequential ISCAS and ITC benchmarks to assess the accuracy and run time of the proposed statistical yield analysis algorithm, called SSTAc, which we implemented in SIS [10]. Each circuit was synthesized with SIS and mapped into a 45-nm gate library, which we built based on a predictive model [11]. The threshold voltage (V_t) was chosen as a representative source of process variation; it was assumed to have a normal distribution with 0.22 V as its mean (μ) and 20 mV as its standard deviation (σ).

978-1-4244-2934-9/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Korea Advanced Institute of Science and Technology. Downloaded on December 10, 2009 at 00:10 from IEEE Xplore. Restrictions apply.

TABLE II

CIRCUIT STATISTICS AND COMPARISON OF TIMING YIELD AND RUN TIME OF MONTE CARLO AND SSTAC

Circuits			# cRFONs		# sRFONs		# seq-BPs		# Eff.	Yield (%)			Run time (s)		
Name	# CLs	# F/Fs	Total	Eff.	Total	Eff.	Total	Eff.	F/Fs	MC	SSTAc	Diff.	MC	SSTAc	Speedup
b03	171	30	14	1	3	1	48	3	6	93.08	92.90	0.18	18.50	0.45	41.1×
b11	609	31	117	12	5	1	69	0	3	92.52	93.01	0.49	61.41	2.44	25.2×
b13	381	53	62	3	9	1	74	1	4	93.16	93.37	0.21	36.42	0.34	107.1×
s208	96	9	22	1	3	1	18	0	4	93.58	93.16	0.42	9.55	0.45	21.2×
s298	134	14	22	4	12	1	23	0	2	92.96	93.10	0.14	14.54	1.15	12.6×
s838	340	32	107	3	12	1	68	1	4	92.91	93.03	0.12	39.91	0.17	243.8×
s1423	713	74	171	6	30	1	154	1	3	94.70	93.09	1.61	69.52	5.45	12.8×
s5378	1386	163	317	6	25	1	151	1	3	93.61	93.51	0.10	168.01	0.92	182.6×
s13207	3182	490	757	40	72	1	561	0	2	94.31	93.55	0.76	453.03	158.77	2.9×
s38584	13158	1424	2357	21	144	0	2525	0	1	92.25	93.51	1.26	1977.97	19.87	99.6×
Average												**0.53**			**74.9**×

For a particular combination of input transition time and output load, each gate (timing arc, to be precise) is simulated with SPICE at seven different V_t-values ($\mu-3\sigma$, $\mu-2\sigma$, $\mu-\sigma$, μ, $\mu+\sigma$, $\mu+2\sigma$, and $\mu+3\sigma$), which yields a discrete delay PDF. The clock network was arbitrarily constructed in a way that all the effective flip-flops are connected to the same (inverted) buffer to maximize the effect of sRFON; the remaining flip-flops are equally distributed to buffers; the whole buffers are then connected to a clock source via another stage of buffers. The timing yield obtained by SSTAc was compared to Monte Carlo (MC) simulation of 10000 runs.

The first three columns of Table II show the name, the number of combinational gates, and the number of flip-flops of benchmark circuits. The numbers of total cRFONs and effective cRFONs are shown in columns 4-5; the threshold we used to detect effective cRFON was $\varepsilon_1 = 0.8$ (with $\varepsilon_1' = 0.4$), which was empirically obtained. Columns 6-9 show the numbers of total and effective sRFONs and seq-BPs, which were determined by the number of effective flip-flops shown in column 10 (we used $\varepsilon_2 = 0.9$ for effective capturing flip-flops and $\varepsilon_3 = 0.4$ for effective launching flip-flops). It can be readily seen that the number of correlation sources are greatly reduced. There is only one effective flip-flop for s38584; there is no effective capturing flip-flops since max-time is dominated by a primary output; out of launching flip-flops that can be reached from that primary output, one is declared as effective; this example is thus free from sRFON.

Timing yield from SSTAc and MC are compared in columns 11-13, which shows that SSTAc is very accurate with yield difference of 0.53% on average. The next three columns compare the run time; SSTAc achieves about 75× speedup over MC.

V. CONCLUSIONS

We have proposed a statistical static timing yield analysis algorithm, which can handle topological correlations in sequential circuit with its clock network. Systematic methods have been described to filter out combinational RFONs and launching and capturing flip-flops, so that all three correlations (cRFONs, sRFONs, and seq-BPs) can be handled without sacrificing accuracy within reasonable amount of run time. Timing yield obtained by SSTAc showed difference of 0.53% on average with 75× run time improvement compared with Monte Carlo simulation.

However, for certain benchmarks, e.g. s13207, the speedup is not significant, because of outstanding number of cRFONs. The possibility of further reducing the correlation sources while maintaining the same accuracy remains as a future work. Since sRFONs have non-negligible amount of effect on timing yield, different topology of clock network will lead to different yield, which is another topic for future work.

REFERENCES

[1] P. Zuchowski, P. Habitz, J. Hayes, and J. Oppold, "Process and environmental variation impacts on ASIC timing," in *Proc. Int. Conf. on Computer Aided Design*, Nov. 2004, pp. 336–342.

[2] J. Liou, K. Cheng, S. Kundu, and A. Krstic, "Fast statistical timing analysis by probabilistic event propagation," in *Proc. Design Automation Conf.*, June 2001, pp. 661–666.

[3] A. Agarwal, D. Blaauw, V. Zolotov, and S. Vrudhula, "Computation and refinement of statistical bounds on circuit delay," in *Proc. Design Automation Conf.*, June 2003, pp. 348–353.

[4] J. Jess, K. Kalafala, S. Naidu, R. Otten, and C. Visweswariah, "Statistical timing for parametric yield prediction of digital integrated circuits," in *Proc. Design Automation Conf.*, June 2003, pp. 932–937.

[5] H. Chang and S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," in *Proc. Int. Conf. on Computer Aided Design*, Nov. 2003, pp. 621–625.

[6] C. Visweswariah, K. Ravindran, K. Kalafala, G. Walker, and S. Narayan, "First-order incremental blockc-based statistical timing analysis," in *Proc. Design Automation Conf.*, June 2004, pp. 331–336.

[7] S. Raj, S. Vrudhula, and J. Wang, "A methodology to improve timing yield in the presence of process variations," in *Proc. Design Automation Conf.*, June 2004, pp. 448–453.

[8] D. Blaauw, K. Chopra, A. Srivastava, and L. Scheffer, "Statistical timing analysis: from basic principles to state of the art," *IEEE Trans. on Computer-Aided Design*, vol. 27, no. 4, pp. 589–607, Apr. 2008.

[9] M. Pan, C. Chu, and H. Zhou, "Timing yield estimation using statistical static timing analysis," in *Proc. Int. Symp. on Circuits and Systems*, May 2005, pp. 2461–2464.

[10] E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Sldanha, H. Savoj, P. R. Stephan, R. K. Brayton, and A. Sangjovanni Vincentelli, "SIS: a system for sequential circuit synthesis," May 1992, Tech. Rep. UCB/ERL M92/41.

[11] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45nm design exploration," in *Proc. Int. Symp. on Quality Electronic Design*, Mar. 2006, pp. 585–590.

978-1-4244-2934-9/09/$25.00 ©2009 IEEE

Authorized licensed use limited to: Korea Advanced Institute of Science and Technology. Downloaded on December 10, 2009 at 00:10 from IEEE Xplore. Restrictions apply.