P^{2}CLRAF: An Pre- and Post-silicon Cooperated Circuit Lifetime Reliability Analysis Framework

Song Jin [1,2], Yinhe Han [1,2], Huawei Li[1,2], Xiaowei Li [1,2]
[1]*Key Laboratory of Computer System and Architecture,*
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, P.R. China
[2]*Graduate University of Chinese Academy of Sciences, Beijing, P.R. China*
{*jinsong, yinhes, lihuawei, lxw*}*@ict.ac.cn*

***Abstract*—Statistical static timing analysis (SSTA) considering process variation and aging effects is usually used to analyze circuit lifetime reliability at design phase. A key challenge for statistical lifetime reliability analysis is that an accurate statistical timing model is needed to carefully model practical variation distribution as well as delay correlation. In this work, P^{2}CLRAF, a circuit lifetime reliability analysis framework is proposed. It calibrates pre-silicon SSTA result by learning the collected data from path delay testing at post-silicon timing validation phase. A neural network inside P^{2}CLRAF is trained to learn variation distribution and delay correlation based on the statistic of path delay testing. The learned information is then fed back to SSTA to further improve the accuracy of circuit lifetime reliability analysis. Experimental results demonstrate the effectiveness of the proposed analysis framework.**

***Keywords*-Lifetime Reliability, Process Variation, NBTI**

I. INTRODUCTION

With the aggressive scaling of feature size, fabrication-induced process variation (PV) causes circuit delay to manifest as statistical distribution and in turn posing significant impact on the circuit reliability [1-2]. On the other hand, aging effect, such as negative bias temperature instability (NBTI), degrades circuit performance continuously during the run-time, leads to the circuit still suffering from reliability problem during the service lifetime [3-5].

Process variation and NBTI effects have strong influence on each other [6-7]. Due to PV-induced uncertainty on device parameter, NBTI-induced threshold voltage degradation should also be modeled as a statistical process. Similarly, the statistic of circuit timing will change under NBTI effect as circuit operational time increasing. Therefore, exploiting statistical static timing analysis (SSTA) which considers the joint effect of PV and aging is an effective way on analyzing circuit lifetime reliability at design phase. In [6], Wang et al. firstly proposed to analyze circuit aging under considering PV effect and applied the analysis method to the single circuit path. In [7], Lu et al. proposed an analysis framework to evaluate circuit lifetime reliability under PV and aging effects at pre-silicon phase.

A key challenge in statistical circuit lifetime reliability analysis is lack of reliable statistical timing model (STM). The underlying STM used in SSTA should accurately model practical parameter variation as well as delay correlation. However, the magnitude of mismatch between simulated behavior at pre-silicon phase and actual behavior measured at post-silicon phase increases as the processes advancing. Therefore, It is essential to exploit the data of post-silicon timing measurement to calibrate the SSTA result at pre-silicon phase.

Some researches have been done on post-silicon timing validation. Liu et al. [8] proposed to divide the die into grids and put corresponding test structures on the grids to measure parameter variation. Benjamin et al. [9] proposed to refine the bound on delay correlation imposed by SSTA by learning the actual delay correlation from path delay testing at post-silicon phase. In [10], Wang et al. proposed to rank the circuit paths by support vector machine learning based on the post-silicon path delay testing. However, they didn't take the impact of aging effect on path delay variation and correlation during the circuit lifetime into account.

In this work, we propose P^{2}CLRAF, an pre- and post-silicon cooperated timing analysis framework to predict circuit lifetime reliability under considering the joint effect of PV and aging. At pre-silicon phase, a variation-aware gate-level aging degradation model is formulated to characterize lifetime reliability of the gate. By applying this gate-level model to the whole circuit, SSTA is performed to identify a set of statistically critical paths (SCPs) based on the conservative assumption of variation distribution.

At post-silicon phase, path delay testing is performed on a few amount of sampled chips targeting the SCPs. Then, a neural network is setup and trained given the statistic of path delay distribution. The training process gradually calibrates the assumed parameters in pre-silicon SSTA to more accurately reflect practical variation distribution and delay correlation. It terminates until the error is below the designated threshold. Finally, the learned parameters are fed back to SSTA in the circuit lifetime reliability analysis.

The rest of paper is organized as follows. Section II presents path identification process. Section III introduces post-silicon learning. Experimental results are presented in Section IV and finally, we conclude in Section V.

II. Path Identification at Pre-silicon Phase

This section introduces how to identify the statistically critical paths using SSTA which is based on the gate-level PV-aware aging-induced delay degradation model.

A. PV-aware Statistical NBTI Model

NBTI is a special aging effect on PMOS transistor. Wang et al. [5] proposed a NBTI model to predict the long-term threshold voltage degradation of PMOS:

$$\Delta V_{th\_nbti} = \left[\sqrt{K_v^2 \cdot T_{clk} \cdot \alpha} / \left(1 - \beta_t^{(1/2n)}\right)\right]^{2n} \quad (1)$$

where $\beta_t = 1 - \frac{2\varepsilon_1 \cdot t_e + \sqrt{\varepsilon_2 \cdot C \cdot (1-\alpha) \cdot T_{clk}}}{2t_{ox} + \sqrt{C} \cdot t}$

$K_v = \left(\frac{qt_{ox}}{\varepsilon_{ox}}\right)^3 \cdot K_1^2 \cdot C_{ox} \cdot (V_{gs} - V_{th}) \cdot \sqrt{C} \cdot exp\left(\frac{2E_{ox}}{E_{o1}}\right)$

In this model, α is the duty cycle, i.e., the percentage of time that PMOS is negative biased. For diffusion species are H_2, n is around 0.16. t is the operational time that the circuit experienced. V_{th} is the nominal threshold voltage.

Considering PV effect, V_{th} of PMOS can be expressed as:

$$V_{th} = V_{th\_nom} + \Delta V_{th\_sys} + \Delta V_{th\_ran} \quad (2)$$

where V_{th_nom} is the nominal threshold voltage. ΔV_{th_sys} and ΔV_{th_ran} are the changes of threshold voltage due to systematic (inter- and intra-die) and random variations.

Substituting V_{th} in Eq.2 into Eq.1, the PV-aware statistical NBTI model can be obtained. Moreover, by only concentrating on some important parameters which have strong impacts on NBTI degradation, the statistical model can be further simplified as:

$$\Delta V_{th\_nbti} = A \cdot (1 - \gamma \cdot (\Delta V_{th\_sys} + \Delta V_{th\_ran})) \cdot \alpha^n \cdot t^n \quad (3)$$

where A is a fitted parameter related to the technology and operational condition (such as V_{dd} and T). γ is also a fitted parameter which denotes the sensitivity of NBTI degradation to the PV effect. α and t are reserved due to their exponential impacts on NBTI-induced degradation [3].

In this work, the quad-tree model proposed in [11] is used to model inter-die as well as intra-die variations with spatial correlation. As shown in Fig.1, gate V_{th} variation can be modeled as the sum of random variables in one of the grids at different levels in quad-tree model. For example, the V_{th} variations of gates 1, 2 and 3 in Fig.1 can be expressed as:

$\Delta V_{th(1)} = \Delta V_{th(2,1)} + \Delta V_{th(1,1)} + \Delta V_{th(0,1)} + \Delta V_{th\_ran(1)}$
$\Delta V_{th(2)} = \Delta V_{th(2,4)} + \Delta V_{th(1,1)} + \Delta V_{th(0,1)} + \Delta V_{th\_ran(2)}$
$\Delta V_{th(3)} = \Delta V_{th(2,16)} + \Delta V_{th(1,4)} + \Delta V_{th(0,1)} + \Delta V_{th\_ran(3)}$

Applying the quad-tree partition to model PV-induced variation on V_{th}'s between different gates, for PMOS k in a gate, Eq.3 is reformulated as:

$$\Delta V_{th\_nbti(k)} = A \cdot \left(1 - \gamma \cdot \sum_i \Delta V_{th(i)}\right) \cdot \alpha_k^n \cdot t^n \quad (4)$$

where $\Delta V_{th(i)}$ corresponds to the one of the random variables in the model. The notation $\sum_i$ takes over all of the random variables in the model. For the variables which are not associated with PMOS k, the coefficient $A \cdot \gamma \cdot \alpha_k^n \cdot t^n$ is 0.

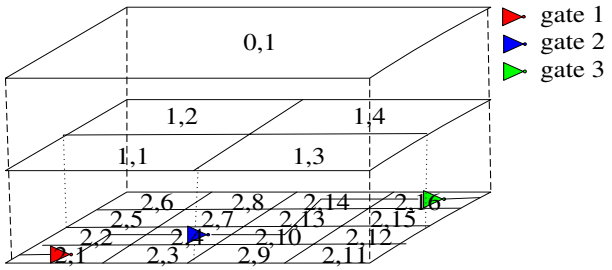

Figure 1. Quad-tree model.

B. Statistical Delay Degradation Model

Threshold voltage of PMOS k under the joint effect of PV and aging can be expressed as:

$$V_{th(k)} = V_{th\_nom(k)} + \Delta V_{th\_sys(k)} + \Delta V_{th\_ran(k)} + \Delta V_{th\_nbti(k)} \quad (5)$$

substituting $\Delta V_{th\_nbti(k)}$ in Eq.5 with Eq.4, we get:

$$V_{th(k)} = V_{th\_nom(k)} + A \cdot \alpha_k^n \cdot t^n + \left(1 - A \cdot \gamma \cdot \alpha_k^n \cdot t^n\right) \cdot \sum_i \Delta V_{th(i)} \quad (6)$$

here, $A \cdot \alpha_k^n \cdot t^n$ denotes V_{th} degradation induced by NBTI alone and is substituted by symbol V_{th_A} for simplicity in the rest of paper. Similarly, $(1 - A \cdot \gamma \cdot \alpha_k^n \cdot t^n) \cdot \sum_i \Delta V_{th(i)}$ denotes NBIT-induced V_{th} degradation under PV effect, and we substitute the coefficient $1 - A \cdot \gamma \cdot \alpha_k^n \cdot t^n$ with symbol $C_{A\&P}$ in the rest of paper.

The gate propagation delay can be approximately recognized as linear function of the threshold voltage [12]. Therefore, the propagation delay from input node k of the gate to the gate output can be expressed as:

$$D_k = D_{nom(k)} + B_k \cdot V_{th\_A} + \beta_k \cdot C_{A\&P} \cdot \sum_i \Delta V_{th(i)} \quad (7)$$

where $D_{nom(k)}$ denotes the nominal gate delay. B_k is the fitted coefficient which reflects increase of the gate delay due to NBTI-induced threshold voltage increase under nominal condition. β_k is also a fitted coefficient which denotes the impact of PV-induced threshold voltage variation on gate delay change without considering NBTI effect.

C. Fast Statistical Static Timing Analysis Method

This paper employs a similar statistical static timing analysis (SSTA) method as in [11]. In SSTA, gate delay and arrival time are both expressed using Eq.7. The arrival time A_i at gate input node i is:

$$A_i = A_{nom(i)} + \sum_j B_j \cdot V_{th\_A(j)} + \left(\sum_j \beta_j \cdot C_{A\&P(j)}\right) \cdot \sum_i \Delta V_{th(i)} \quad (8)$$

where $A_{nom(i)}$ denotes the nominal arrival time at input node i. The notation $\sum_j$ denotes the accumulative process of adding j gate delays onto the arrival time before the arrival time calculation reaches at node i.

Given the propagation delay D_i of the gate from input node i to the output node k:

$$D_i = D_{nom(i)} + B_i \cdot V_{th\_A(i)} + \beta_i \cdot C_{A\&P(i)} \cdot \sum_i \Delta V_{th(i)}$$

the arrival time A_k at k is:

$$A_k = A_{nom(k)} + \sum_l B_l \cdot V_{th\_A(l)} + \Big(\sum_l \beta_l \cdot C_{A\&P(l)}\Big) \cdot \sum_i \Delta V_{th(i)} \tag{9}$$

where $A_{nom(k)} = D_{nom(i)} + A_{nom(i)}$,

$\sum_l B_l \cdot V_{th\_A(l)} = B_i \cdot V_{th\_A(i)} + \sum_j B_j \cdot V_{th\_A(j)}$,

$\sum_l \beta_l \cdot C_{A\&P(l)} = \beta_i \cdot C_{A\&P(i)} + \sum_j \beta_j \cdot C_{A\&P(j)}$,

$l = j + 1$.

The maximum of two arrival times can be computed as follows. Given two arrival times A_1 and A_2:

$$A_1 = A_{nom(1)} + \sum_m B_m \cdot V_{th\_A(m)} + \Big(\sum_m \beta_m \cdot C_{A\&P(m)}\Big) \cdot \sum_i \Delta V_{th(i)}$$

$$A_2 = A_{nom(2)} + \sum_p B_p \cdot V_{th\_A(p)} + \Big(\sum_p \beta_p \cdot C_{A\&P(p)}\Big) \cdot \sum_i \Delta V_{th(i)}$$

where m and p are the number of gates visited during the computing processes of A_1 and A_2, respectively. A_3 , the maximum of these two arrival times is:

$$A_3 = A_{nom(3)} + \sum_q B_q \cdot V_{th\_A(q)} + \Big(\sum_q \beta_q \cdot C_{A\&P(q)}\Big) \cdot \sum_i \Delta V_{th(i)} \tag{10}$$

where $A_{nom(3)} = max\left(A_{nom(1)}, A_{nom(2)}\right)$,

$\sum_q B_q \cdot V_{th\_A(q)} = max\big(\sum_m B_m \cdot V_{th\_A(m)}, \sum_p B_p \cdot V_{th\_A(p)}\big)$,

$\sum_q \beta_q \cdot C_{A\&P(q)} = max\big(\sum_m \beta_m \cdot C_{A\&P(m)}, \sum_p \beta_p \cdot C_{A\&P(p)}\big)$

where $q \leq m + p$. A_3 is actually the upper bound of the maximum of A_1 and A_2 [11]. Using Eq.9 and Eq.10, statistical timing analysis can be performed to compute the statistical circuit maximum arrival time or the statistical maximum arrival time at any node.

D. Identifying Statistically Critical Path

To make statistically critical paths (SCPs) reflect parameter variation and delay correlation under PV and aging effects, a priority factor is used to identify the SCPs:

$$PF_i = D_i \cdot NG_i \tag{11}$$

where PF_i is the priority factor of path i. D_i is the delay of path i obtained from SSTA. NG_i is the number of grids that path i passing through given the quad-tree model.

Algorithm 1 illustrates how to exploit the priority factor to identify SCPs.

Algorithm 1{

1) Initialize SCP set $S_{SCPs} = \{\phi\}$ and preliminary path set $S_{PPs} = \{\phi\}$, apply quad-tree model to the circuit netlist;
2) Perform SSTA, for each path i in the circuit, obtain D_i, NG_i and then calculate PF_i ;
3) Sort the paths in descendant order based on their $PF's$ and put them into S_{PPs};
4) Fetch the first path from S_{PPs} and record the grids that the path passing through. Putt this path into S_{SCPs} and remove it from S_{PPs};
5) If all of the grids in quad-tree model are covered, algorithm terminates. Or else, go back to step 4.}

III. Post-silicon Learning

For SSTA at pre-silicon phase based on quad-tree model, it is essential to obtain practical proportions of intra-die variation allocated to the levels in the model. Given this reason, in this work we try to learn the related variation information from the measured data at post-silicon phase.

We setup a neural network model for post-silicon learning purpose. The neural network has two layers and each layer has its own transfer function. The variances of SCPs are the inputs for the network. The target outputs are the proportions of intra-die variation allocated to the levels in quad-tree model and the random variation.

For training the network, Monte Carlo simulation can be performed many times based on different assumptions on total variations, allocations of intra-die variations and random variation. Static timing analysis is then performed to obtain the sampled path delays. The assumed variation allocations in Monte Carlo simulations are used as the target outputs. The network training terminates until the designated training iteration time is achieved.

IV. Experiment

Experiments are performed on some ISCAS benchmark circuits. A 3-level quad-tree model is used to model the PV-induced threshold voltage variation. Each gate in the circuit netlist is randomly allocated a location on the 4×4 grid in the bottom level and then, the random variables associated to the gate along the hierarchy are determined as well. Parameters in Eq.4 and Eq.7 are fitted by HSPICE simulation under PTM 65nm [13] technology node.

A. Post-silicon Learning Evaluation

On post-silicon learning, we simulate 20 sampled chips after fabrication by 20 times Monte Carlo simulations and pick out 10 SCPs totally. Static timing analysis is used to simulate path delay testing at post-silicon phase. Delays of SCPs for all of Monte Carlo simulations are grouped into an input matrix, while the corresponding variation assumptions in Monte Carlo simulations are grouped into a target output matrix.

We setup a two-layer neural network for post-silicon learning purpose by writing a MATLAB script. It takes 10 inputs and has 4 outputs. Hidden layer and output layer have 20 and 3 neurons, respectively. The network training stops after the validation error increased for six iterations.

Fig.2 illustrates the regression analysis after the training for c880. As shown in Fig.2, the network outputs track target outputs very well for training, validation and test. The R-value is over 0.99 for the total response (R-value measures the correlation between network outputs and targets outputs. A R-value of 1 means a close relationship).

After network training, we generate 10 sets of new input and output samples to verify the effectiveness of network learning. The network test results are listed in Table I.

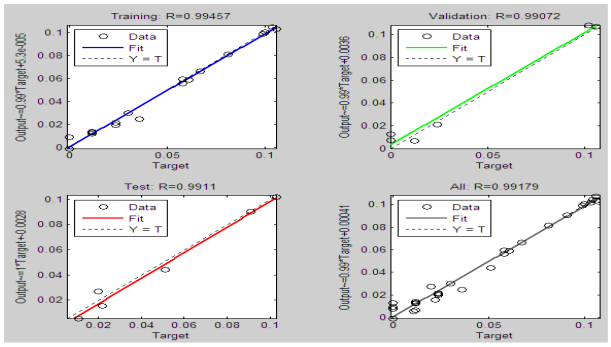

Figure 2. Regression analysis after learning.

Table I
NEURAL NETWORK TEST RESULT

circuit	I	MSR	R	circuit	I	MSR	R
c880	3	$7.2e^{-3}$	0.9688	c7552	7	$8.8e^{-4}$	0.9863
c1908	2	$5.3e^{-3}$	0.9779	s298	4	$2.7e^{-3}$	0.9778
c2670	2	$3.3e^{-4}$	0.9821	s820	2	$8.5e^{-3}$	0.9641
c3540	4	$6.8e^{-3}$	0.9911	s1196	3	$3.1e^{-4}$	0.9855
c5315	2	$7.7e^{-3}$	0.9656	s1238	6	$3.2e^{-3}$	0.9936
c6288	3	$6.3e^{-3}$	0.9882	s9234	2	$9.1e^{-4}$	0.9791

The columns 2 and 6 in Table I list the iteration time that the minimum MSR is achieved during the network test. The 3-4 and 7-8 columns list the MSR and R-value after network test. We can see that MSRs for all of the experimental circuits are very small (generally below 10^{-2}) while R-values are very close to 1. This demonstrates the effectiveness of the network learning.

B. Circuit Lifetime Reliability Analysis

Fig.3 illustrates the prediction on mean and variance of the circuit maximum arrival time assuming 5-year service lifetime by using P^{2}CLRAF. To simulate the workload that the circuit experienced, signal probabilities on primary inputs of the circuit are set to 0.5 and the work temperature is assumed at $375K$.

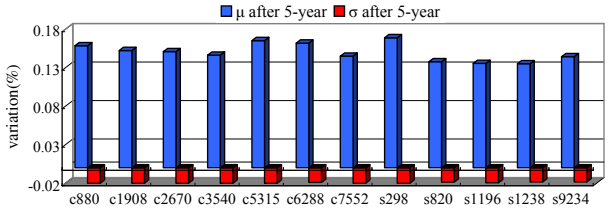

Figure 3. Circuit delay distribution under PV and aging effects.

As shown in Fig.3, with the time increasing, the mean of circuit delay increases while the variance decreasing. It also means that under the joint effect of PV and aging, more and more paths have larger probability to be critical path, which affects the circuit delay distribution and poses larger challenge on the circuit timing optimization.

V. CONCLUSION

This work proposes P^{2}CLRAF, an pre- and post-silicon cooperated circuit lifetime reliability analysis framework. It exploits neural network to learn the practical parameter variation and delay correlation given the collected data from path delay testing at post-silicon phase. The learned information is then fed back to the PV and aging effects aware SSTA to further calibrate the pre-silicon circuit lifetime reliability analysis result.

ACKNOWLEDGMENT

The work was supported in part by National Natural Science Foundation of China (NSFC) under grant No.(60633060, 60806014, 60831160526, 60776031), in part by National Basic Research Program of China (973) under grant No. 2005CB321604, 2005CB321605.

REFERENCES

[1] S. Borkar. et al. Parameter variation and impact on circuits and microarchitecture. Design Automation Conference, pp. 338-342, 2003.

[2] K. A. Bowman, S. G. Duvall, J. D. Meindl. Impact of die-to-die and within-die paramter fluctuations on the maximum clock frequency distribution for gigascale integration. IEEE Journal of Solid-State Circuits, Vol. 37, No. 2, pp. 183-190, Aug. 2002.

[3] W. Wang, Y. Cao, et al. The impact of NBTI on the performance of combinational and sequential circuits. Design Automation Conference, pp. 364-369, Jun. 2007.

[4] Y. Wang, H. Luo, K. He, R. Luo, H. Yang, Y. Xie. Temperature-aware NBTI modeling and the impact of input vector control on performance degradation. ACM/IEEE Design, Automation, and Test Europe, pp. 546-551, 2007.

[5] W. Wang, Z. Wei, S. Yang, Y. Cao. An efficient method to identify critical gates under circuit aging. International Conference on Comuter-Aided Design, pp. 735-740, 2007.

[6] W. Wang, Y. Cao, et al. Statistical prediction of circuit aging under process variations. IEEE Custom Integrated Circuits Conference, pp. 13-16, Sept. 2008.

[7] Lu. YingHai, et al. Statistical reliability analysis under process variation and aging effects. IEEE/ACM Design Automation Conference, pp. 514-519, Aug. 2009.

[8] Q. Z. Liu, S. S. Sapatnekar. Confidence scalable post-silicon statistical delay prediction under process variations. IEEE/ACM Design Automation Conference, pp. 497-502, June. 2007.

[9] B. N. Lee, L. C. Wang, M. S. Abadir. Refined statistical static timing analysis through learning spatial delay correlations. IEEE/ACM Design Automation Conference, pp. 149-154, 2006.

[10] L. C. Wang, P. Bastani, M. S. Abadir. Design-silicon timing correlation - a data mining perspective.

[11] A. Agarwal, D. Blaauw. Statistical timing analysis for intra-die process variations with spatial correlations. International Conference on Comuter-Aided Design, pp. 900-907, 2003.

[12] T. Sakurai, et al. Alpha-power law MOSFET model and its application to CMOS logics. IEEE Journal of Solid-State Circuits, Vol. 25, No. 2, pp. 584-594, 1990.

[13] W. Zhao and Y. Cao. New generation of predictive technology model for sub-45nm early design explorations. Available at $http : //www.eas.asu.edu/\_ptm$. TED, 53(11):2816-2823, Nov. 2006.