

Gate-Level Aged Timing Simulation Methodology for Hot-Carrier Reliability Assurance

Yoshiyuki Kawakami, [†]Jingkun Fang, Hirokazu Yonezawa, Nobufusa Iwanishi, [†]Lifeng Wu,

[†]Alvin I-Hsien Chen, ^{††}Norio Koike, [†]Ping Chen, [†]Chune-Sin Yeh, [†]Zhihong Liu

Matsushita Electric Industrial Co.,Ltd., Osaka, 570-8501, Japan

Tel: +81-6-6906-2424, Fax: +81-6-6906-1406, email: {kaw, yonezawa, iwanishi}@mrg.csdd.mei.co.jp

[†]BTA Technology, Inc., San Jose, CA95112, USA

Tel: +1-408-451-1210, Fax: +1-408-451-1211, email: {jkf, lfw, ac, pc, csy, zhl}@btat.com

^{††}Matsushita Electronics Corporation, Kyoto, 601, Japan

Tel: +81-75-662-8911, Fax: +81-75-681-0705, email: nkoike@krl.mec.mei.co.jp

Abstract - This paper presents a new aged timing simulation methodology that can be used for hot-carrier reliability assurance of VLSI. This methodology consists of a compact model and a unique algorithm. The ratio based model simplifies the aging I-V characteristics of MOSFET over time into the aged timing and the corresponding ratio at gate-level. A new algorithm is proposed including a gate primitive decomposition method and an aged slew rate propagation method. This algorithm provides good stress representation and can achieve comparable accuracy with the conventional transistor-level approach. The above methodology has been implemented in a new simulator. Experimental results demonstrate that the simulator based on this methodology realizes full-chip circuit capacity and can be applied to various reliability analyses including degradation-sensitive critical paths and clock skew.

I. INTRODUCTION

Recent advancements in manufacturing technology have been changing which physical phenomena are most dominant during VLSI design. The hot-carrier effect, in this regard, is becoming one of the major reliability concerns in the deep sub-micron era. The lifetime of CMOS circuit depends on the degree of degradation caused by this wearout and should be assured prior to the marketing of the product.

The circuit timing or speed, the most critical specification of a chip, is deteriorated. It is mainly caused by the transistor's characteristic shift over time with high electric field near drain edge generating high energy hot-carriers that are out of thermal equilibrium. Previously this phenomenon

was well managed by conservative manufacturing technologies. Accelerating miniaturization of the device, however, makes it difficult to assure HCI reliability as scaling of the device dimension surpasses scaling of the voltage. The time is coming when the hot-carrier effect can no longer be ignored in the speed-oriented VLSI design.

Many design-for-reliability activities have been reported especially in the field of microprocessors [1][2][3]. One well-known simulation approach is to use a transistor-level simulation that directly models the transistor's aging I-V curve behavior [4][10] given, for instance, by Eq.(1) [12]:

$$\frac{\Delta I_d}{I_{d0}} \propto \left(\int_0^T \frac{I_d(t)}{H} \cdot \left(\frac{I_{sub}(t)}{I_d(t)} \right)^b \cdot dt \right)^a \quad (1)$$

where $I_d(t)$ and $I_{sub}(t)$ are drain current and substrate current, respectively; ΔI_d is the drain current degradation; I_{d0} is the fresh drain current; a , b , and H are process dependent parameters which should be calibrated for different technologies; T is the circuit operation time. With the process dependent parameters a , b , and H properly calibrated, this equation is applicable to VDSM technologies [13] and is implemented in almost all major transistor-level simulations. This approach provides highly accurate but slow simulations and not practical for large circuit. On the other hand, a macromodeling has also been available although it is slightly less accurate than the transistor-level approach. For instance, *iRULE* [5] uses macromodels to approximate degradation of the logic gate. This approach transforms series

transistors in multi-input logic into an equivalent transistor for simplification. Although it can handle larger circuits than transistor-level ones, it is still difficult to consider the behavior of the multi-input logic and the complex logic correctly.

Designers are forced to redesign their circuit according to the result of the hot-carrier reliability simulation. However the applicable circuit size and functionality available from the previous works were limited to a small part of a chip. Only the understanding of full-chip behavior can help the designers decide whether to fix the circuit. In order to realize such a diagnosis, it is essential for them to have a full-chip aged timing simulation tool.

To address these issues, we have developed a new gate-level aged timing simulation methodology that can be used for hot-carrier reliability assurance of VLSI. The methodology is made up of a ratio based model [7][8] and a unique algorithm including two new methods. This model transforms the aging I-V characteristics of MOSFET into the aged timing and the corresponding ratio at gate-level. Due to compactness of the model, this higher abstraction of the circuit can reduce the amount of computation. Both a gate primitive decomposition method and an aged slew rate propagation method play an important role to maintain the comparable accuracy with the conventional transistor-level approach. The performance of the prototype simulator demonstrate that the proposed methodology can realize full-chip circuit capacity and can be applied to various reliability analyses including degradation-sensitive critical paths and clock skew.

Our former studies [7][8] focused mainly on the gate-level ratio based modeling. This paper, on the other hand, emphasizes the algorithmic aspect, especially two newly developed methods.

Sections II and III mention the basis of the gate-level aged timing simulation and the ratio-based modeling. Section IV shows the detailed explanation of the algorithm. Section V illustrates the effectiveness of the whole methodology with 0.25um chip results. Finally Section VI summarizes the paper.

II. GATE-LEVEL AGED TIMING SIMULATION

This section introduces a gate-level aged timing simulation. Fig. 1 shows a definition of the gate-level timing. Assuming logic gates connected to each other with wire RC parasitic

elements, the initial input-to-output delay T_{fresh} of a gate changes into the degraded delay T_{aged} after certain period of operation.

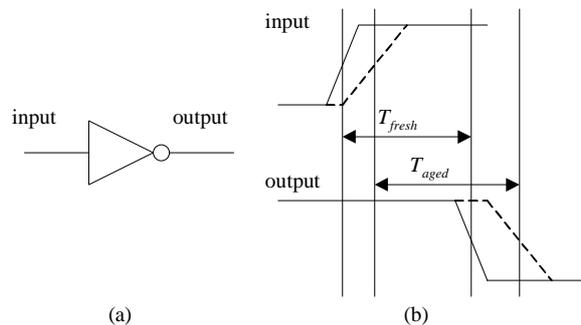


Fig.1. Gate-level timing definition. (a) and (b) shows an inverter as an example and a timing chart, respectively.

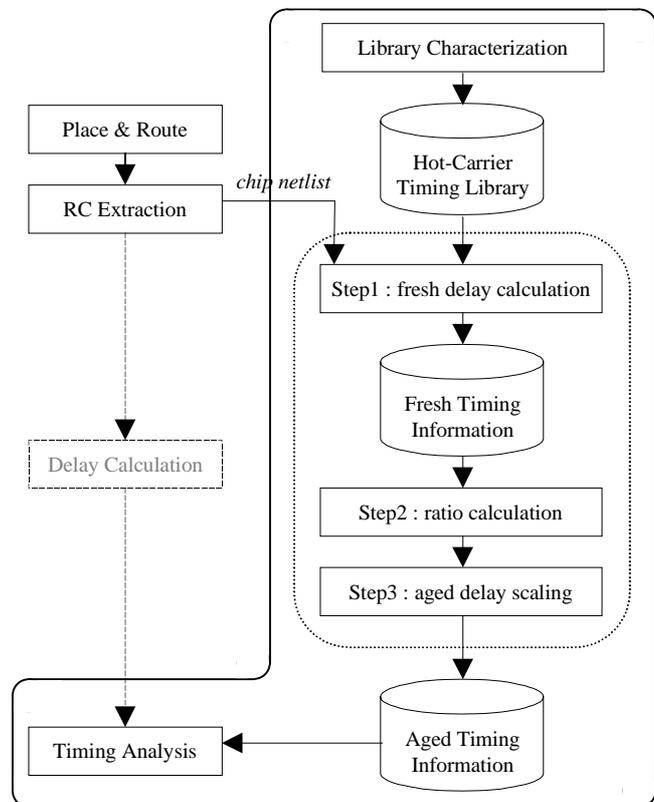


Fig.2. Proposed methodology and algorithm. Broken line and dotted line show the conventional fresh delay calculation flow and the new algorithm, respectively.

Fig. 2 illustrates proposed methodology and algorithm, where the hot-carrier timing library is automatically created during library characterization by running the transistor-level simulator in advance according to the model (See Section III). This library stores a ratio table, a gate timing data, and a drive ability data for each gate or primitive (See Section IV).

Aged timing information is generated through the proposed algorithm. Finally at gate-level a static or a dynamic timing analyzer or a logic simulator performs aged timing simulation by referring to the aged timing information.

III. RATIO BASED MODEL

In this section we briefly describe the ratio based model [7][8]. This gate-level model represents hot-carrier-induced timing change and consists of the following equations (2)-(4):

$$T_{aged} = \alpha \times T_{fresh} \quad (2)$$

$$\alpha = \sum_{i=1}^n \alpha_i - (n-1) \quad (3)$$

$$\alpha_i = f(T_{slew}, C_L, N_{sw}) \quad (4)$$

where T_{fresh} is a fresh input-to-output delay, T_{aged} is an aged input-to-output delay, α is a ratio, n is the number of series transistors in a gate, and α_i is an i -th input-to-output ratio. T_{slew} is an input slew rate, C_L is an output capacitance, and N_{sw} is a switching number. Eq.(2) is used for both gate and interconnection. Eq.(3) is a useful expansion to cover the multi-input gate which has series transistors.

The model calculates aged timing information multiplied with the predetermined ratio to the fresh timing information. The validity of the model itself has already been proven in [7][8] in comparison with the transistor-level simulator and Si chip. The ratio table in the hot-carrier timing library in Fig. 2 i.e. Eq.(4) is generated in the form of the 3D table for each transition type of each signal path of each gate or primitive.

IV. SIMULATION ALGORITHM

A. Overview

The new algorithm is developed as in Fig. 3. It consists of three major steps. Each step corresponds to each box inside the dotted line of Fig. 2. In Step 1, fresh timing information is calculated by using a conventional delay calculation technique e.g.[11]. In Step 2, all gates are decomposed into primitives. The primitive is a child gate of which the multistage gate is composed. To better reflect the switching behavior inside a gate, it is beneficial to handle a multistage gate at primitive-level. Then this step provides primitive delays and ratios utilizing two important methods. In Step 3, after recombining primitive data to be gate data, aged timing T_{aged} is obtained by scaling fresh timing T_{fresh} with above ratio

α in Eq.(5):

$$T_{aged} = \sum_{j=1}^m \left(\frac{t_{aged_j}}{t_{fresh_j}} \times t_{fresh_j} \right) = \sum_{j=1}^m (\alpha_j \times t_{fresh_j}) \quad (5)$$

where m , t_{fresh} and t_{aged} are the number of primitives in the target gate, fresh and aged input-to-output delays at primitive-level, respectively. We will next focus on Step 2 which is the core of this algorithm.

```

Procedure Algorithm for Gate-Level Aged Timing Simulation ( )

/*Step 1: Fresh delay calculation*/
for (all logic gates and interconnections) {
    Calculate  $T_{fresh}$  at gate-level;
}
/*Step 2: Ratio  $\alpha$  calculation*/
for (all logic gates) {
    /* Gate primitive decomposition method*/
    Decompose multi-input gates into primitives;
    Count switching activity of each primitive;
    Look up hot-carrier timing library and decide ratio  $\alpha$ ;
}
for (all primitives and interconnections) {
    /*Aged slew rate propagation method*/
    Calculate  $t_{fresh}$  at primitive level with slew rate propagation;
    Calculate  $t_{aged}$  at primitive level with slew rate propagation;
    Calculate  $\alpha$  by dividing  $t_{aged}$  by  $t_{fresh}$ ;
}
/*Step 3: Aged delay scaling*/
for (all logic gates and interconnections) {
    Recompose primitive-level data into gate-level data;
    Scale  $T_{aged}$  by multiplying  $T_{fresh}$  with ratio  $\alpha$  at gate-level;
}

end Procedure;

```

Fig.3. Algorithm inside the dotted line of Fig. 2.

B. Gate Primitive Decomposition Method

Some combinational logic gates have hierarchy. For instance, AND gate is made up of a NAND gate and an inverter as shown in Fig. 4. Thus the damage of NAND and the inverter should independently be obtained by monitoring each stress condition. When N_{sw} is counted at Y, it is necessary to know how many times the inverter has experienced X transitions. N_{sw} at Y depends on X transitions and cannot simply be determined by the switching number at inputs A and B. Therefore a new method has been introduced to break such a gate down into logic primitives, in this case NAND and inverter, to extract the switching numbers of the intra-gate node correctly. After the decomposition, the aged timing information is calculated at the logic primitive-level. Table 1 indicates the effectiveness of the decomposition. Two cases with and without decomposition are shown. In the latter case, the inverter experiences the same number of transitions as input A or B when A-to-Y or B-to-Y ratio is measured, because NAND and the inverter cannot be handled separately.

However in the former case, the inverter can experience the sum of transitions incurred to both A and B and result in the bigger ratios even under the same operational condition. Considering the amount of degradation, typically around 10% in the lifetime operation, the sub-percent improvement in Table I is significant.

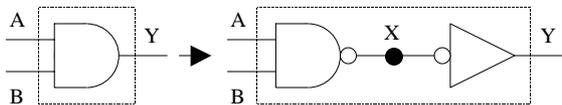


Fig.4. Decomposition of AND into NAND and inverter.

TABLE I
RESULTS WITH AND WITHOUT DECOMPOSITION

Signal path in AND gate	Ratio with decomposition	Ratio without decomposition
A to Y	1.0247	1.0239
B to Y	1.0254	1.0220

C. Aged Slew Rate Propagation Method

Hot-carrier damage is likely to be generated during the input signal transition period when the substrate current of MOSFET's meets its peak. Accordingly, accurate slew rate calculations are a key requirement for this algorithm to better represent the stress. Slew rate propagation occurs when the slew rate of the previous stage affects that of the next stage. Here we have adopted a two-pass propagation in which the waveform propagates two times during t_{fresh} and t_{aged} calculations. Moreover in this method the drive ability change over time is reflected to both gate and interconnect delays. From the primary input of the circuit to the output, first a fresh propagation is performed. Next an aged propagation is done using aged timing and aged drive ability.

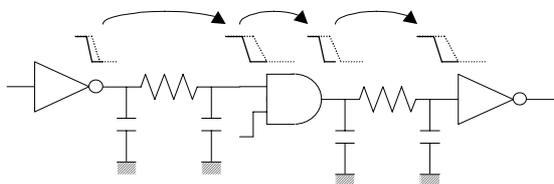


Fig.5. Aged slew rate propagation. Solid line and dotted line denote fresh and aged waveforms, respectively.

V. EXPERIMENTAL RESULTS

In this section we demonstrate the high performance of the methodology using the prototype simulator. Implementation of this simulator is based on Fig. 2.

A. Accuracy

Using small test circuits, we first confirmed the accuracy of the prototype simulator in comparison with the commercially- available transistor-level simulator BTABERT[9] which also implements the transistor degradation model based on Eq.(1). These circuits include the several wiring-rich signal paths which consist of inverter, buffer, NAND, NOR, AND, OR logic gates in a 0.25 μ m CMOS technology. 2.5[V] stress voltage, 200[MHz] frequency, and 10 year operation were assumed. The ratio of input-to-output gate delay was compared. Correlation in Fig. 6 shows good agreement within 1[%] accuracy to BTABERT. It is concluded that accurate aged timing simulation can be achieved by the proposed methodology. The error can be reduced by preparing denser 3D table in the hot -carrier timing library.

Next, we show the effectiveness of the aged slew rate propagation method. Without propagation in Fig. 7, the ratios are less than those in Fig. 6 and many points are out of 1[%] range. The reason is because the aged slew rate, the stress condition, and eventually the damage are underestimated. This is remarkable in the case of these wiring-rich circuits which have not yet discussed in [7][8].

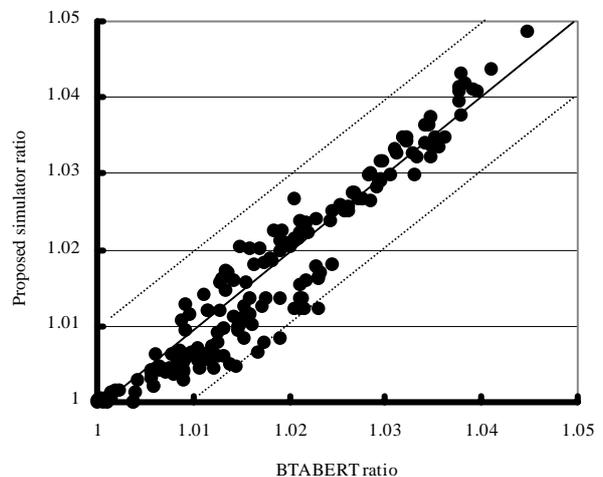


Fig.6. Ratio correlation with slew rate propagation. The number of points is 95.

B. Analysis for Chip Reliability Assurance

To demonstrate the performance, we have simulated a large logic chip with thousands of signal paths under the conditions summarized in Table 2, assuming 10 year dynamic operation.

TABLE II
SAMPLE LOGIC CHIP SPECIFICATION

Technology	0.25um CMOS
Voltage	2.5V
Clock Frequency	132MHz
Number of Logic Gates / Primitives	34271 / 37088
Variety of Logic Gates / Primitives	161 / 193
Number of Signal Paths	6100

Fig. 8 shows path delay ratio vs initial path delay. Ratios of most signal paths are found to be within 1.02 and some are over 1.05. It was also found that small initial delay was a result of fewer gate stages in a path. Therefore large input slope which contributes large degradation can be existed.

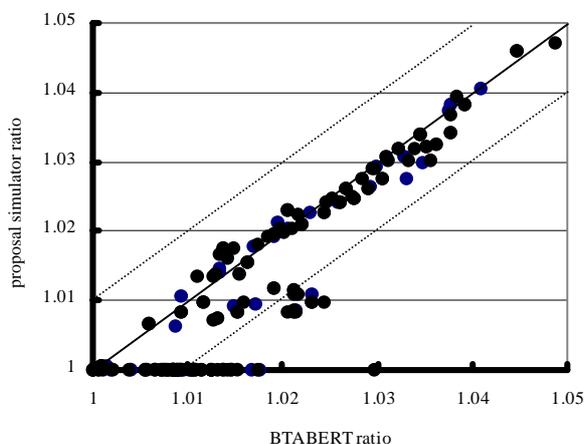


Fig.7. Ratio correlation without slew rate propagation. The number of points is 95.

The chip designer is normally interested in the wearout behavior of critical path or clock skew because it could be fatal to the future chip operation. As seen in Fig. 8, the designer can judge whether the design meets reliability requirement by setting certain criteria i.e. the critical ratio. In summary, the proposed methodology can characterize the time-dependent behavior of chip-size circuit quite well.

C. Discussions

The aging phenomenon occurs gradually over time. However we practically assume that the original fresh waveform or slew rate will last for a given stress period during the simulation. Although this approximates well according to our study, we can technically break the period into several intervals and execute iteration. For instance, the result after 1 year is simulated from the fresh result and the

result after 2 years is simulated from the former 1 year result.

Gate primitive decomposition method increases the computational burden for the characterization. However we found that the average number of primitives in a gate was less than 2 in the case of combinational logic. Therefore CPU time cannot be sacrificed too much.

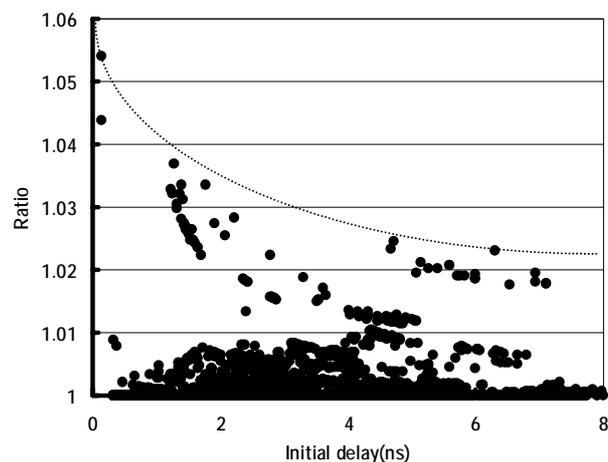


Fig.8. Path delay degradation. Dotted line is the upper bound.

Hot carrier degradation is a function of a switching number or a switching probability [6]. There are three general techniques applicable for our methodology to obtain the switching number: (1) logic simulation based technique, (2) probability analysis based technique, and (3) unified probability technique. Logic simulation usually outputs the dynamic switching information. Technique (1) must require a long input vector for the logic simulation. However the result depends on the quality of input vector similar to the power simulation. Further study is needed to search for the worst case vector. Technique (2) also needs a probability analysis of the target circuit. It is not easy to find the worst case probability as well. Our ongoing study suggests that the technique (3) in which all gates have one same probability is much more practical than other counterparts. As a dotted line in Fig. 8 qualitatively suggests, the upper ratio boundary exists being inversely proportional to the initial delay. The boundary is mainly resulted from the paths like a clock tree operating all the time at the highest frequency. By setting one highest frequency for all gates in the circuit, our methodology quickly outputs such a path ratio boundary. The designers can utilize this boundary information to decide a reliability guard-band of the whole chip or perform quality assurance.

VI. CONCLUSIONS

In this paper we have shown that the full-chip aged timing simulation for hot-carrier reliability assurance of VLSI can be achieved by the proposed methodology. This unique methodology contains the ratio based model and the algorithm. Furthermore a gate primitive decomposition method and an aged slew rate propagation method have been newly developed as part of the algorithm.

Higher gate-level abstraction of the circuit drastically reduces the computation so that the large circuit can be simulated without limitation of logic variety. The above algorithm provides good stress representation to maintain within 1[%] accuracy compared to the conventional transistor-level approach. Eventually the chip designers can utilize the significant benefits of this methodology for their full-chip guard-band diagnoses including critical path delay and clock skew.

Future study will be centered on the coupling noise effect on hot-carrier degradation and the quality assurance application.

ACKNOWLEDGMENT

The authors would like to thank Prof. Chenming Hu of University of California in Berkeley for his valuable suggestions. They also would like to thank Dr. A.Matsuzawa, K.Ohtani, Y.Okamoto, S.Tsuji, and K.Tatsuuma for their constant support.

REFERENCES

- [1] M.R.Choudhury and J.S.Miller, "A 300MHz CMOS microprocessor with multi-media technology," *IEEE ISSCC Digest of Technical Papers*, pp.170-171, 1997.
- [2] W.J.Grundmann, D.Dobberpuhl, R.L.Allmon, and N.L. Rethman, "Designing high performance CMOS micro-processors using full custom techniques," *IEEE/ACM Proc. Design Automation Conf.*, pp.722-727, 1997.
- [3] R.Khanna, A.Ben-Meir, L.DiGregorio, D.Draper, R.Krishna, R.Maley, A.Mehta, S.Oberman, L.Tsai, and T.Williams, "A 0.25mm x86 microprocessor with a 100MHz socket 7 interface," *IEEE ISSCC Digest of Technical Papers*, pp.242-243, 1998.
- [4] R.H.Tu, E.Rosenbaum, W.Y.Chan, C.C.Li, E.Minami, K. Quader, P.K.Ko, and C.Hu, "Berkeley reliability tools - BERT," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol.12, no.10, pp.1524-1534, 1993.
- [5] C.C.Teng, W.Sun, S.M.Kang, P.Fang, and J.Yue, "iRULE : Fast hot-carrier reliability diagnosis using macro-models," *IEEE Proc. Custom Integrated Circuits Conf.*, pp.421-424, 1994.
- [6] P.C.Li, G.I.Stamoulis, and I.N.Hajj, "A probabilistic timing approach to hot-carrier effect estimation," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol.13, no.10, pp.1223-1234, 1994.
- [7] H.Yonezawa, J.Fang, Y.Kawakami, N.Iwanishi, L.Wu, A.I-H.Chen, N.Koike, P.Chen, C-S.Yeh, and Z.Liu, " Ratio based hot-carrier degradation modeling for aged timing simulation of millions of transistors digital circuits," *IEEE Int. Electron Devices Meeting Technical Digest*, pp.93-96, 1998.
- [8] H.Yonezawa, J.Fang, Y.Kawakami, N.Iwanishi, L.Wu, A.I-H.Chen, N.Koike, P.Chen, C-S.Yeh, and Z.Liu, "Accurate gate-level aged timing modeling for full-chip hot-carrier reliability simulation," *IPSIJ DA Symposium*, pp.21-26, 1999.
- [9] BTABERT 3.0 User's Manual, *BTA Technology Inc.*, 1998.
- [10] Y.Leblicic and S.M.Kang, "Hot-carrier reliability of MOS VLSI circuits," *Kluwer Academic Publishers*, 1993.
- [11] J.-T.Kong and D.Overhauser, "Digital timing macro-modeling for VLSI design verification," *Kluwer Academic Publishers*, 1995.
- [12] C.Hu, S.Tam, F.Hsu, P.Ko, T.Chan, and K.Terrill, "Hot-electron-induced MOSFET degradation: model, monitor, and improvement," *IEEE Trans. Electron Devices*, Vol. 32, No. 2, pp.375-385, 1985.
- [13] W.Jiang, H.Le, S.Dao, S.Kim, B.Stine, J.Chung, Y.Wu, P. Bendix, S.Prasad, A.Kapoor, T.Kopley, T.Dunganm, I. Manna, P.Marcoux, L.Wu, A.Chen, and Z.Liu, "Key hot-carrier degradation model calibration and verification issues for accurate AC circuit-level reliability simulation," *Proceeding of IEEE International Reliability Physics Symposium*, pp.300-306, 1997.