Circuit-Based Reliability Characterization Methods in Advanced CMOS Technologies

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Xiaofei Wang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Chris H. Kim, Advisor

January 2014

Acknowledgements

First and foremost I want to thank my advisor, Professor Chris Kim. He encouraged me to learn from scratch when I just joined his lab. He has taught me, both consciously and unconsciously, how good engineering projects are done. I appreciate his constant support, patience and contribution of time to make my Ph. D. experience educational and rewarding and fulfilling. I am benefited by the high standards he has set and will continue to live up to them as I begin my career.

My thanks also go to Professors Yuichi Kubota and Jian-Ping Wang for being both my study advisor and life mentor during the first two years of my Ph. D. pursuit.

I would like to thank the members of my doctoral final defense committee: Professors Ramesh Harjani, Kiarash Bazargan, and Hubert Lim. I appreciate them taking time out of their busy schedules to critique my work. I am also grateful to Professors Sachin Sapatnekar and James Kakalios for serving on my preliminary oral defense committee.

I want to thank our collaborator Vijay Reddy in Texas Instrument for the insightful discussion and valuable feedbacks on my reliability projects.

I would like to thank all my colleagues in Prof. Kim's research group for making the lab a fun place to work. In particular, I am obliged to the senior group members Kichul Chun, Tony Kim, John Keane, Wei Zhang, Pulkit Jain and Dong Jiao for steeping my learning curves. I want to thank Seung-hwan Song, Ayan Paul, Bongjin Kim, Won Ho Choi, Jongyeon Kim, Weichao Xu, Qianying Tang, Saroj Satapathy, Somnath Kundu and

Chen Zhou for the numerous insightful and stimulating discussions, for the countless hours we were working together for the deadlines and for sharing the brunt of frustration and the joy of successes in research.

My time at UMN was made much balanced, fun and enriched by my friends who are either around the twin cities or far away. I will never forget the time I spent with my friends hiking and camping in the wild and our adventures in the national parks. I am grateful for my biking, skiing and rock climbing buddies and their enthusiasm on these side projects. And I feel lucky to have friends to share the best and the worst of the life during my graduate study and the friends who remember bringing me birthday cakes.

Lastly, I want to thank my family for all their love and encouragement. Thank you my parents for everything you have done to raise me up and support me in all my pursuits. I also want to thank my cousin sister for being my soul mate through the constant oversea phone calls and text messages.

This thesis is dedicated to my parents

for their endless love, support and encouragement.

Abstract

Integrated circuit reliability has become an increasingly important design consideration as the CMOS technology keeps aggressively scaling to its physical limit. The parametric shifts and circuit failures caused by reliability issues such as Bias Temperature Instability (BTI), Hot Carrier Injection (HCI), and Electromigration (EM) have become more prevalent as electrical fields and current density continue to increase in scaled devices. In addition, the rapid introduction of process improvements, such as high-k/metal gate stacks, has led to new reliability issues including positive BTI (PBTI) in n-type transistors.

Traditionally, designers deal with reliability problems by adding a conservative design guardband calculated by aging models based on the worst-case degradation scenario. However, there are a few issues associated with this method: 1. the power and performance overhead of guardbanding have started to increase in the newer technologies. 2. the aging models used by designers to are mostly based on the device probing data, which is not accurate to predict the circuit performance degradation. 3. the device probing has the drawbacks of expensive costs of probing equipment and limited timing resolution.

In order to resolve these issues, one of the key aspects is to develop accurate and efficient means to measure the effects of different aging mechanisms on circuit parameters accurately. For this purpose, several unique on-chip circuit-based sensing

systems have been proposed, which provide us with important advantages: namely, pico-second timing resolution for usage condition stress, micro-second measurement interruption to prevent unwanted recovery, and excellent immunity to voltage and temperature drifts. The proposed odometer designs utilize standard logic gates and a simple scan-based interface, making them suitable for integrating into an actual processor system. In this thesis, four dedicated on-chip circuit designs that we have implemented over two generation of process technology to characterize reliability issues on various types of circuits will be presented.

Table of Contents

List of figures

Chapter 1.

Introduction

The semiconductor industry is facing unprecedented challenges as the complementary metal oxide semiconductor (CMOS) technology approaches its fundamental limit. Process variation, leakage power and device reliability issues have emerged as serious problems that limit the performance benefits gained by traditional device scaling. In particular, circuit parametric shifts or possible failures caused by front-end-of-line device reliability issues including Bias Temperature Instability (BTI), Hot Carrier Injection (HCI), as well as the back-end-of-line reliability issue such as electromigration (EM) and Joule heating have become more severe with shrinking device sizes and voltage margins. The higher voltage stress and elevated temperatures in chips these days are causing the above issues to become an increasingly important design consideration in high performance systems. Moreover, process technology improvements such as metal gate/high-k (MG/HK) devices have introduced new degradation concerns including positive-BTI in n-type devices, in addition to the already complex aging behavior [1].

In order to account for the impending logic slow-down that will come with aging, the traditional way is to reduce the operating frequency or supply level by a certain amount, which is called guardband [2, 3]. The amount of guardbanding that designer has to put

into the circuit for guaranteed lifetime is estimated by aging models based on device level characterizations. However, there are two issues that rely on the device level models to estimate circuit performance. Firstly, the operating conditions on which a transistor is operated in a real chip is considerably different from individual devices, especially when power and performance control become even more complex in modern processors with the techniques such as DVFS (dynamic voltage and frequency scaling), turbo/near-threshold voltage mode operation and power gating. Secondly, with the same amount of device aging, the end effect of aging on circuit performance would largely depend on the particular circuit under analysis. It is more accurate and meaningful to directly measure out the performance degradation from a circuit and to develop aging models upon that, than to use the models derived from device level data.

Research, process and design groups are devoting significant resources and efforts to better understand the aging mechanisms, and to explore strategies to conduct precise aging prediction. Our critical aspect of that work involves developing accurate and efficient on-chip monitoring circuits to measure the effects of the different aging mechanisms on circuit parameters. This new paradigm provides a number of benefits over traditional device probing such as higher measurement resolution, shorter measurement time, reduced test structure area, shorter test time, and simpler test setups. The proposed sensor designs utilize standard logic gates and a simple scan based interface, making them suitable for integrating into an actual processor system. Based on the instant aging information from an integrated in-situ monitor, real-time compensations

and architecture level solutions can be triggered for any wear-out issues experienced by the system, which could fundamentally change the way we deal with aging issue in future processors.

The remainder of this session is organized as follows. Section 1.1 provides a brief introduction on the major transistor degradation mechanisms. Section 1.2 introduces relationship between device level aging and circuit performance degradation. Section 1.3 introduces ring oscillator based test structures to measure the frequency degradation in digital circuits. A projection of high-level application of the on-chip monitoring will be discussed in section 1.4, followed the summary of contributions of the projects that this thesis contains in section 1.5.

1.1 Brief Introduction to CMOS Device Reliability Mechanisms

In this section, we will briefly introduce the mechanisms of the two major CMOS device reliability mechanisms: bias temperature instability and hot carrier injection, as well as the leading back-end-of-line reliability mechanism: electromigration.

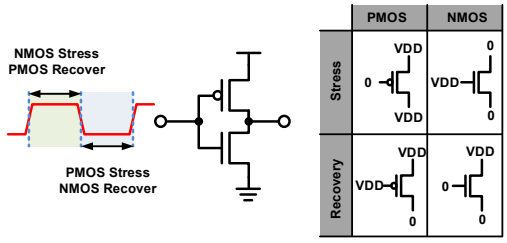


Figure 1.1: PBTI and NBTI stress condition for an inverter during standard operation.

Bias temperature instability (BTI) is one of the most critical device degradation mechanisms in both conventional poly-Si/SiON and high-k/metal gate (HK/MG) [4-7]. As the name indicates, BTI refers to a time-dependent instability accelerated with increasing bias and temperature. Specifically, it is characterized by a positive shift in the absolute value of the MOSFET threshold voltage (V_t), which occurs when a device is biased in the inversion mode, but with a small, or no, lateral electric field (i.e., $V_{ds} \approx 0$ V). When a stressed device is turned off, it immediately enters the "recovery" phase, where trapped holes are released, thereby reducing the absolute value of the V_t. The degradation and recovery process happens during the normal digital operation conditions, as shown in Fig. 1.1. The term negative bias temperature instability (NBTI) is used for PMOS, whereas for NMOS the degradation is called positive bias temperature instability (PBTI) since the corresponding gate bias is negative and positive, respectively. For conventional devices with poly-Si electrodes and Si based gate dielectrics, BTI is mostly observed in PMOS, and it is known that BTI in SiON gate dielectrics is more severe as compared to SiO_2. With the introduction of MG/HK, BTI in PMOS has become equally significant compared to NMOS.

Reaction-diffusion (RD) theory has long been assumed to be essentially correct that the V_t shift attributes to a hydrogen-diffusion controlled interface state creation [8, 9]. The V_t shift is generally due to that interface states generated the breaking of Si-H bonds at the gate dielectric interface hold holes in the inversion layer, which generates positively charged interface traps (for PMOS) (Fig. 1.2 (a))[10]. After the bias is

removed, the hydrogen diffused back and the defects are removed with the annealing of the dangling bonds, as shown in Fig. 1.2 (b) [11-13].

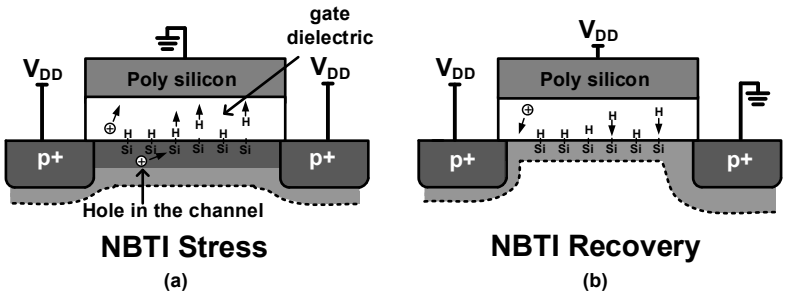


Figure 1.2: Transistor cross sections illustrating NBTI reaction-diffusion (RD) theory.

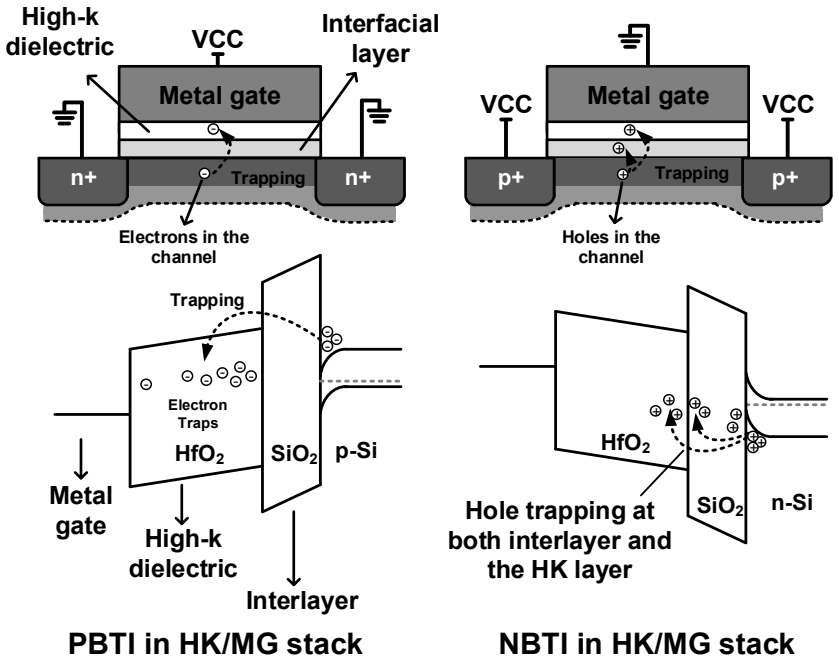


Figure 1.3: HK/MG PBTI and NBTI trapping mechanism comparison

Recent studies suggest that direct hole trapping to the existing defects in both the bulk and interface is a significant degradation contributor especially to small stress times, based on the inconsistence of the experimental observations with the RD theory [14]. And with the usage of the MG/HK stacks, the previous small electron trapping effect in SiO_2 and SiOH becomes significant in HK layer [1, 15]. Fig. 1.3 illustrates the trapping mechanism in the MG/HK devices for PMOS and NMOS in both cross section and band diagram. During PBTI (in NMOS), it is mainly observed that the negative charge trapping is located mostly in the HK layer or in the region between the HK and the interfacial oxide layer. For NBTI (in PMOS), the holes can be trapped at interface-states close to the interface and in the oxide bulk. The trapped charges close to the inversion layer degrade the channel carrier mobility due to the Coulomb scattering effect.

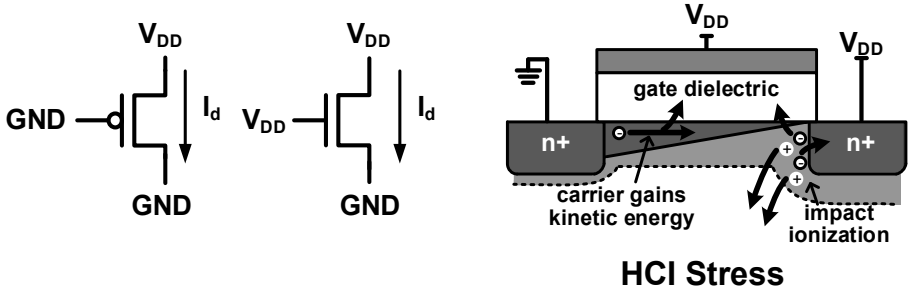


Figure 1.4: HCI stress condition and mechanism demonstrated in a transistor cross section view.

Hot carrier injection (HCI) degradation appears when a large drain-to-source voltage and gate-to-source voltage are applied (Fig. 1.4) [16-20]. Hot carriers (i.e., those

with high kinetic energy) accelerated toward the drain by the corresponding lateral electric field across the channel create traps at the silicon substrate/gate dielectric interface, as well as dielectric bulk traps, and hence degrade device characteristics such as V_t, I_{linear} and I_{on}. HCI has become less prominent in recent year with the down scaling of operating voltages.

It is more of a concern in analog circuit and in power devices where the supply levels are high and the channel current density and duration are large [21]. It draws less concern in normal digital circuits as the device operation is outside of the HCI worst case stress condition ($V_{gs}=V_{ds}=V_{DD}$). HCI is still a reliability concern for certain digital applications require long transition phases, high channel current and high duty factor [22].

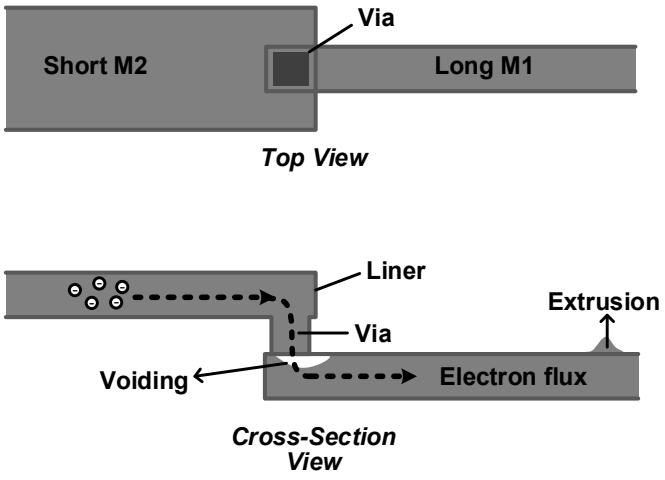


Figure 1.5: EM mechanism illustrated in an multilayer metallization for down-stream case.

Electromigration (EM) is the most concerned back-end-of-line reliability issue as the current density keep increasing in the recent technology nodes. It is the process of metal-ion transport due to high current density stress in metals as shown in Fig. 1.5 [23-29]. As the metal-ions migrate due to the mechanical interaction by electron flux, metal depletion or extrusions occurs. The extrusions can results in shorts between interconnects, while the voids at the metal depletion can cause increases in interconnect resistance or catastrophic disconnections. Note that the direction of the electron flux is from the wide and short M2 to narrow and long M1 (down-stream), which is easier to fail compared to the case that the electrons move from wide and short M1 to narrow and long M2 (up-stream). The reason for this is because the void is more likely to form under the via for the down-stream case which causes a larger resistance shift [28]. The mean time to failure (MTTF) is the parameter to evaluate EM reliability, which can normally be fitted by lognormal distribution [30, 31]. MTTF is related to the average current density and the absolute temperature of the interconnect by Black's equation [25].

1.2 Relationship between Device Level and Circuit Level Reliability

1.2.1 Interconnect RC Impact on Aging

Interconnects used in clock networks, signal buses, network-on-chips, memory wordlines/bitlines, and high-speed I/Os are critical components in modern ICs. CMOS devices in interconnect drivers experience a complex time-varying voltage stress which is a function of the interconnect load. As a result, performance degradation of interconnect

drivers due to reliability mechanisms such as bias temperature instability (BTI) and hot carrier injection (HCI) depends on the length and width of an interconnect wire. BTI is considered as the primary reliability concern in modern CMOS processes and occurs when a device is biased in strong inversion mode. Once a device is turned off, the degradation caused by previous stress periods starts to recover immediately. Despite the scaling of supply voltage in advanced technologies, HCI remains to be an important reliability concern especially towards the end of a product lifetime owing to the 4-5x larger time exponent [20, 32]. Moreover, interconnect drivers (for example, clock drivers) have a higher activity factor and larger average current driven compared with random logic gates, making HCI a greater concern.

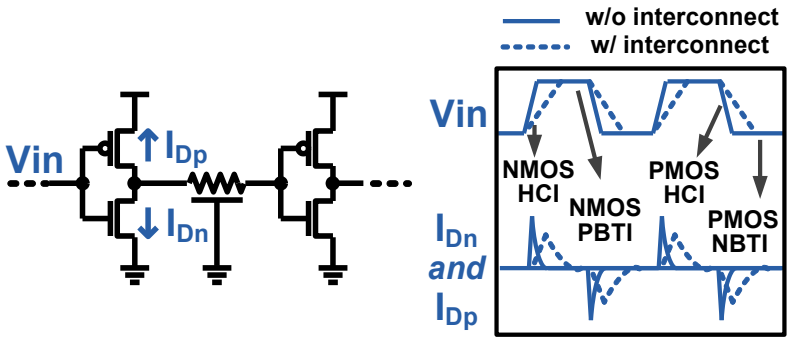


Figure 1.6: Impact of large interconnect RC on voltage and current behavior of a driver. With the interconnect RC, the slew rate is smaller, with longer current duration but smaller peak current.

BTI and HCI mechanisms have different sensitivities to the operation conditions which depend on the interconnect configuration. Moreover, sheet resistance and parasitic capacitance of long wires have not been scaling favorably in advanced processes which

could lead to interconnect dominated paths having drastically different aging behavior compared to logic dominated paths in the future. The large interconnect RC increases the amount of current driven by the transistor to complete each logic transition, while decrease the peak current due to the voltage dividing cause by interconnect resistance, as demonstrated in Fig. 1.6. These changes on current condition would in turn, impact the HCI aging in the interconnect drivers. Also, given the same clock cycle, the decrease slew rate would make the transistors spend less time in the on state, which reduced the BTI stress duty-cycle.

Although there have been previous works showing the impact of fanout load on transistor aging [33-36], almost no attention has been paid to the aging behavior in interconnect drivers for long RC wires. A deeper insight into interconnect driver aging will enable a more complete picture of system level aging and allow us to build interconnect circuits that are more tolerant to device aging.

1.2.2 Asymmetric BTI aging induced duty-cycle shift and its impact on circuit performance

Low power SRAMs, dynamic register files, and domino gates typically rely on both the rising and falling edges of the clock to generate internal timing signals. Unlike standard flip-flop or latch based pipelines where only the primary clock edge (e.g. rising edge) is utilized, the performance of the circuits mentioned above is directly affected by any change in the clock duty-cycle. Bias Temperature Instability (BTI) stress in the clock

signal path during idle or clock gated mode results in an aging-induced duty-cycle shift. Fig. 1.7 illustrates this situation in a typical clock buffer chain scenario. In an idle mode or clock gated mode, the input clock signal is not switching which results in a DC stress condition with NBTI and PBTI occurring in alternative gates. When the circuit is switched back to an active mode, the first clock signal (e.g. the first rising edge in Fig. 1.7) propagates through unstressed fresh devices while the second edge (i.e. the first falling edge in Fig. 1.7) traverses through the stressed devices. Consequently, the delay of the second edge becomes longer compared to that of the first edge due to BTI under DC stress resulting in a duty-cycle shift.

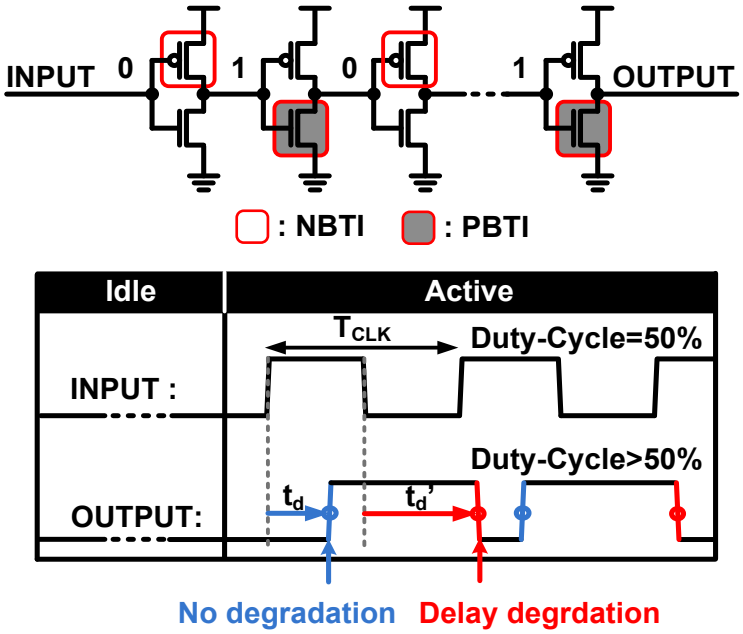


Figure 1.7: (upper) Alternating NBTI/PBTI stress in a buffer chain in idle mode. (lower) Asymmetric delay degradation of rising and falling edges results in duty-cycle shifts.

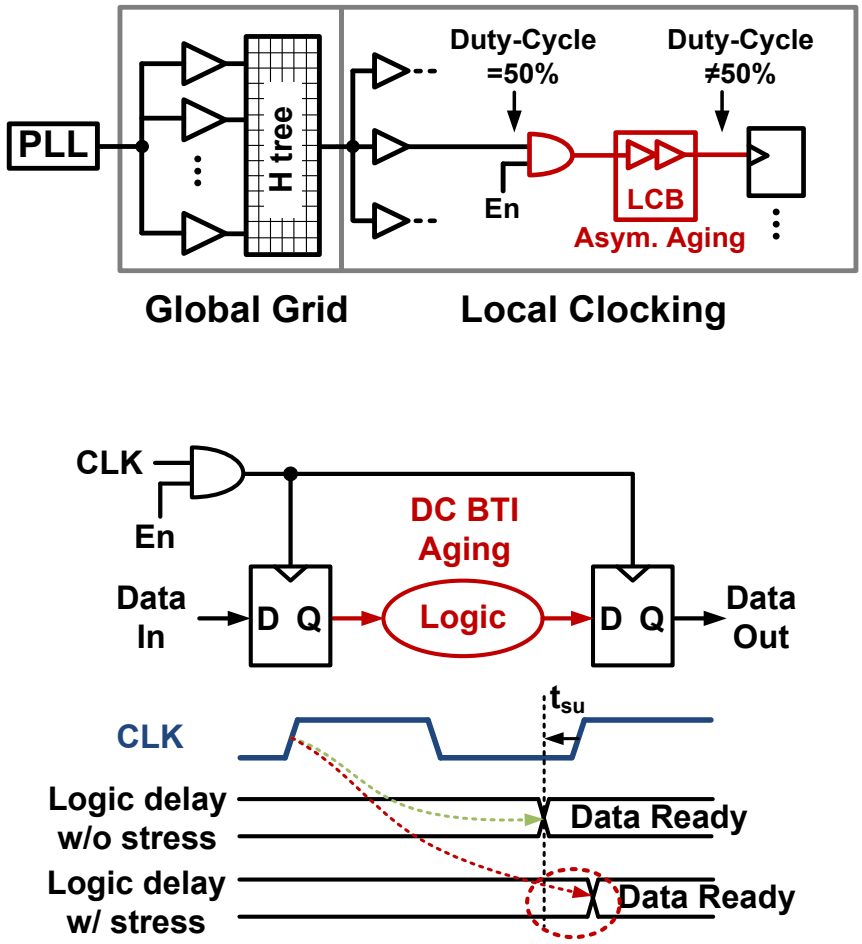


Figure 1.8: (upper) Asymmetric aging in clock distribution network. The local clock buffers after clock gater are under DC stress, resulting in duty-cycle modulation. (lower) Asymmetric aging in sequential circuit, the DC BTI stress causes larger clock to data delay which might violate the setup time.

The aforementioned asymmetric aging induced duty-cycle shift occurs in any circuit which contains a delay path experiencing DC stress condition. For clock distribution network, the clock signal generated by PLL is delivered to the local circuits through a

tree structure clock routing with equal propagation delay and duty-cycle, as illustrated in Fig. 1.8 (upper). The local clock paths can be put into idle mode by gating off the clock signals, where the DC stress condition modulates the clock duty-cycle at the output of the clock path in the next active mode. Since the clock gating is normally delayed locally, the additional delay caused by aging is small, which limits the magnitude of the duty-cycle modulation. Pipelining or sequential circuits under clock gating are also affected by the asymmetric aging, as the logic circuit between two DFFs can be sit in steady state for a long period, which puts all the transistors in DC BTI stress. The delay degradation in the first transition caused by the BTI during the long DC stress may create a setup timing violation.

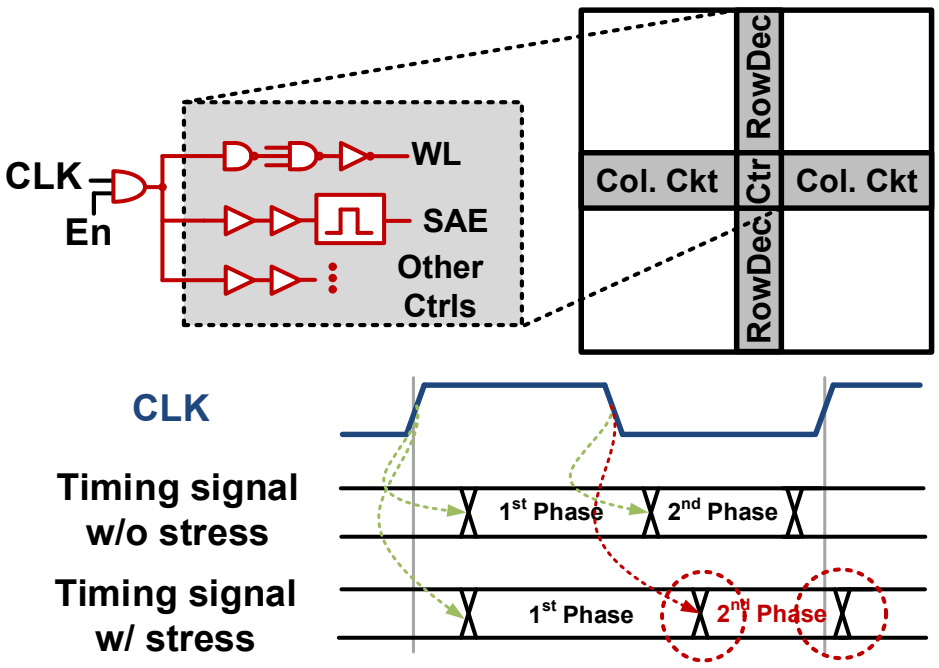


Figure 1.9: Duty-cycle shift SRAM timing signals under DC BTI stress.

With the clock gating technique used in memory blocks, the internal circuits are also exposed to the asymmetric BTI in the idle mode. This effect particularly impacts the SRAM with operation schemes that reply on both the rising and falling edges of clock signal, as shown in Fig. 1.9. As a single clock cycle is used to complete a full read or write operation, the internal controls are divided into two phases triggered by different clock edges. Take read operation for example, the precharge disable signal, wordline enable signal, data latch are related to the rising clock edge, while the column enable and sense amplifier enable are related to the second clock edge. Any shift in duty-cycle caused by the asymmetric BTI aging during the idle mode when the clock is gated off would directly affect the SRAM performance and even functionality.

1.3 On-chip Reliability Monitors and their high level application

Circuit failures due to device aging mechanisms such as BTI, HCI, TDDB, and EM have become increasingly problematic with shrinking device geometries and smaller voltage margins. Traditionally, designers deal with this problem by adding a conservative frequency guardband based on the worst-case degradation scenario. Guardbanding has been used extensively by the semiconductor industry for mitigating lifetime issues; however, the power and performance overhead associated with this conventional method is expected to increase in future process technologies. Recently, the circuit and architecture communities have been exploring an alternative design paradigm based on in-situ aging monitors that could fundamentally change the way we deal with lifetime

issues in future processors. Fig. 1.10 illustrates the overall concept in which a feedback loop triggered by a collection of aging sensors proactively compensates for any wear-out issues experienced by the system. Implementation of the compensation circuit itself (e.g. DVFS) has been relatively well studied so most of the recent effort has been focused on designing accurate and compact monitor circuits or developing architecture level mitigation strategies [37].

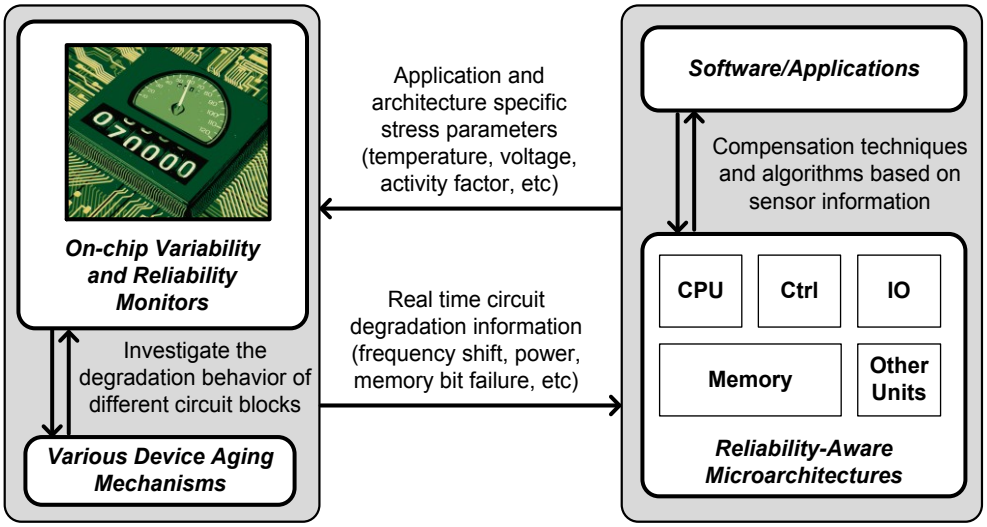


Figure 1.10: A cross-layer aging compensation framework based on in-situ reliability monitors.

Several aging measurement systems have been proposed in the past decade to demonstrate the effectiveness of using compact on-chip circuit for reliability monitoring. Karl et al. designed two separate compact circuits to measure NBTI and TDDB for real-time characterization [38]. Another monitor by Singh et al. captures the onset of TDDB using test chip data and an empirical formula [39]. Hoffman et al. measured BTI and

HCI from a critical path circuits using a single-ended ring oscillator based system [40]. Saneyoshi et al. and Chen et al. proposed monitors for measuring the BTI impact on logic delay [41, 42]. Kim et al. from IBM proposed a method to separate PBTI and NBTI impact on frequency degradation in a high-k metal gate (HKMG) process [43]. More recently, Lu et al. presented details of a ring oscillator based reliability sensor that was successfully deployed in IBM's z196 servers [44]. In that work, product aging data collected for an operation period of over 500 days was reported.

Design Considerations	*Examples of Practical Issues*	*Aging Sensor Implementation in IBM z196 Server [3]*
Type of Sensor	BTI, HCI, TDDB, RTN, transient errors, memory bit failures, etc.	Ring Oscillator based BTI monitor for long-term frequency degradation measurement
Temporal Granularity	Sensing period, threshold setting, dynamic range, etc.	Sampling period: once a week
Spatial Granularity	Per CPU/GPU/memory, per functional unit, per sub-block, etc.	Total: 5 sensors per chip; One sensor per core (x4 cores) plus one sensor in L2 cache
Stress and Measurement Condition	AC vs. DC, accelerated vs. usage condition, fast measurement	AC stress, usage condition, 0.5ms measurement time
Communication	Between data gathering sensor, across sensors, between sensors and processor	Sensors are integrated with IBM z196 pervasive infrastructure with firmware support
Interface and Protocol	Interrupt based, polling, event alarms, performance counter based, etc.	Interrupt based in-field frequency degradation measurement
Testing and Calibration	Similar to any other on-chip monitor circuit	Time 0 frequency shift unknown since first sample is taken after some stress

Figure 1.11: Design considerations and practical issues of on-chip aging sensors. Details of a sensor implementation in IBM z196 mainframe server are shown in the right most column.

Although the main focus of this thesis is on introducing the innovative features of each odometer design and demonstrating their circuit level capabilities, it's worth mentioning the various design considerations and practical issues designers have to consider when applying these sensors to a real processor system (Fig. 1.11). These issues

include, but are not limited to, the type of aging mechanism, the temporal and spatial granularity of the measurements, stress and measurement condition, firmware design, test methodology, and sensor calibration. Most of these design parameters can be readily determined once the system level requirements along with the area budget and interface protocol are known. For the reader's information, Fig. 1.11 shows the actual design parameters used for an aging monitor system recently deployed in IBM zEnterprise 196 mainframe systems aimed at in-field data collection. To the best of our knowledge, this was the first time an on-chip reliability monitor system was used in a commercial product and therefore it serves as a useful guideline for future odometer system designs. In this design, aging data was collected every week from 5 ring oscillator based sensors located across the die that was exposed to the same stress pattern as the main processor chip. One interesting challenge that the authors faced was that unlike in a lab environment where the stress can be precisely controlled, the initial fresh frequency difference between the reference and stressed circuits could not be measured until the system was due for its first check-up routine. This so-called "time zero problem" and other practical issues could pose interesting test and calibration challenges for this emerging class of on-chip monitors.

1.4 On-Chip Reliability Monitoring by Beat Frequency Detection System

For accurate and efficient characterization of the different reliability mechanisms, an on-chip monitoring circuit must be capable of measuring extremely small shifts in operating frequency within a very short measurement time. For a typical stress experiment, the stress voltage applied to the monitor circuit during long stress periods is briefly lowered using on-chip power gates to characterize the degradation at a nominal supply voltage. In order to capture the precise aging data before the BTI recovery takes place, frequency measurements must be completed in a few microseconds. Furthermore, a practical aging monitor should also be able fully-digital, support various stress modes, occupy a small silicon area, use a simple test interface for convenient data collection, and allow independent characterization of the different aging mechanisms.

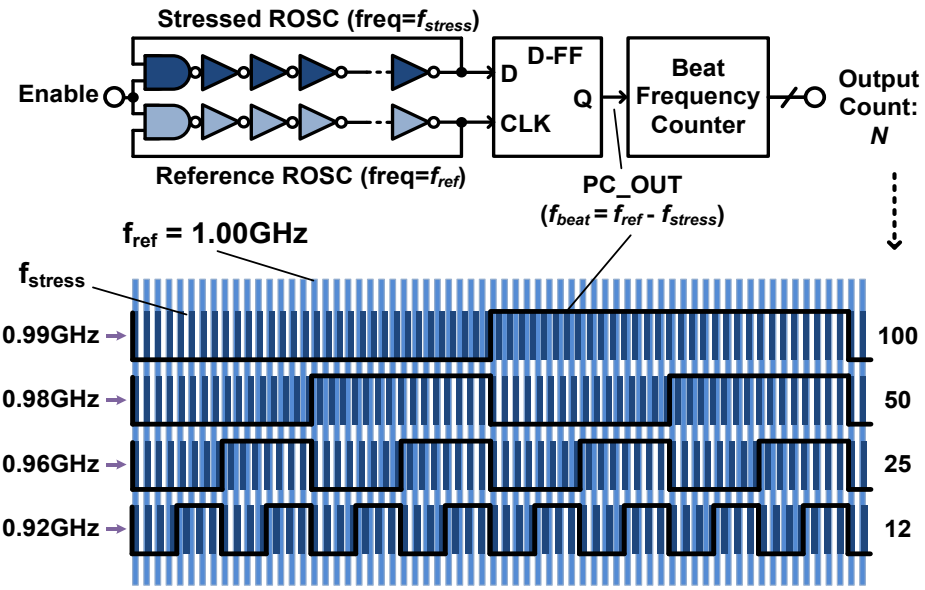


Figure 1.12: Concept behind the beat frequency detection system. Small frequency shifts induced by circuit aging are magnified by measuring the beat frequency (i.e. Δf) rather than the raw frequency of a single ring oscillator.

All these requirements can be met using a novel beat frequency detection concept illustrated in Fig. 1.12 [35, 36]. During the short measurement periods, a D-flip-flop uses a fresh reference Ring OSCillator (ROSC) to sample the output of an identically stressed ROSC. In this configuration, the output of the stressed ROSC is sampled at every rising edge of the reference ROSC output producing a signal that exhibits the beat frequency. The beat frequency is measured using a counter circuit clocked by the output of the reference ROSC. This beat frequency count is recorded after each stress period to calculate the actual shift in the stressed ROSC frequency. Consider the example shown in Fig. 1.12 where the frequency of the reference ROSC is 1GHz and the frequency of the stressed ROSC frequency degrades from 0.99GHz to 0.98GHz. In this case, the count value changes from 100 to 50 for a 0.01GHz (=1%) shift in frequency achieving an extremely high measurement sensitivity. Similarly, the frequency corresponding to a count change from 100 to 99 is just 0.0001GHz (=0.01%). Note that the measurement time, which is the product of the count and the ROSC period, is less than a microsecond in most of our designs which practically eliminates any unwanted BTI recovery. Since the reference and stressed ROSCs are identical structures placed next to each other in the same power domain, the proposed monitor circuit has a high immunity to voltage and temperature drifts.

1.5 Summary of Thesis Contributions

This session will discuss the benefits of the sensor designs that we have implemented to accurately monitor the impact of the transistor aging on the circuit performance degradation. Each circuit was dedicatedly designed and built targeting a specific circuit reliability issue. All the designs deployed the Silicon Odometer beat frequency detection system, which allows us to performance frequency degradation measurements with timing and resolution that is impossible for any conventional measurement system.

The first design studies the detailed aging behavior of interconnect paths, for the first time. The degradation in interconnect performance caused by BTI and HCI in the driver circuit exhibits a strong dependency on the interconnect RC parameters. This dependence was carefully characterized by this design. Based on the understanding verified by the experimental data, simple aging models and modeling method applicable to interconnect designs in advanced technologies were built.

The second on-chip monitor design focuses on the duty-cycle shift effect caused by the asymmetric BTI, and its impact on logic and memory circuits. In particular, detailed duty-cycle modulation in clock distribution network and SRAM read speed degradation are characterized based on the measurements from test chips.

The third design uses a ring oscillator based circuit for separately characterizing PBTI and NBTI induced frequency shifts are demonstrated in a high-k metal gate process. The proposed design, for the first time, supports AC stress with a realistic recovery condition.

Other benefits over the previous works include sub-microsecond measurement time, sub-picosecond resolution and a simple calibration procedure.

The last design is the first known on-chip EM characterization vehicle which supports GHz stress with realistic current waveform. The purposes of the design are to characterize the impact of EM effect on the circuit performance, and to study the AC and Pulsed DC EM lifetime dependence on stress frequency.

Chapter 2.

The Dependence of BTI and HCI Induced Frequency Degradation on Interconnect Length and Its Circuit Level Implications

2.1 Introduction

Interconnects used in clock networks, signal buses, network-on-chips, memory wordlines/bitlines, and high-speed I/Os are critical components in modern ICs. CMOS devices in interconnect drivers experience a complex time-varying voltage stress which is a function of the interconnect load. As a result, performance degradation of interconnect drivers due to reliability mechanisms such as bias temperature instability (BTI) and hot carrier injection (HCI) depends on the length and width of an interconnect wire. BTI is considered as the primary reliability concern in modern CMOS processes and occurs when a device is biased in strong inversion mode [3, 7, 9, 45]. Once a device is turned off, the degradation caused by previous stress periods starts to recover immediately. Despite the scaling of supply voltage in advanced technologies, HCI remains to be an important reliability concern especially towards the end of a product lifetime owing to the

4-5x larger time exponent [20, 21, 32]. Moreover, interconnect drivers (for example, clock drivers) have a higher activity factor and larger average current driven compared with random logic gates, making HCI a greater concern.

BTI and HCI mechanisms have different sensitivities to the operation conditions which depend on the interconnect configuration. Moreover, sheet resistance and parasitic capacitance of long wires have not been scaling favorably in advanced processes which could lead to interconnect dominated paths having drastically different aging behavior compared to logic dominated paths in the future. Although there have been previous works showing the impact of fanout load on transistor aging [33-36], almost no attention has been paid to the aging behavior in interconnect drivers for long RC wires. A deeper insight into interconnect driver aging will enable a more complete picture of system level aging and allow us to build interconnect circuits that are more tolerant to device aging.

For the first time, this work presents measurement results highlighting the dependence of BTI and HCI induced aging on wire length [22]. Our previous "all-in-one" silicon odometer framework [35, 46] was adopted to separate the BTI and HCI contributions with picosecond order resolution and microsecond order measurement interrupts. Measurement data from a 65nm test chip shows that BTI-induced degradation decreases monotonically for longer interconnect length while HCI exhibits a non-monotonic dependency on interconnect length. Based on a detailed circuit level analysis, simple aging models incorporating interconnect parameters are proposed for estimating BTI and HCI induced degradation in interconnect drivers. These models show good

agreement with measured data. Several real-world examples are provided that demonstrate the practicality of the proposed models in the context of interconnect driver design.

2.2 Interconnect Odometer Design

Test Block #	Wire L (μm)	# of ROSC stages
1	0	37
2	250	19
3	500	13
4	1000	7

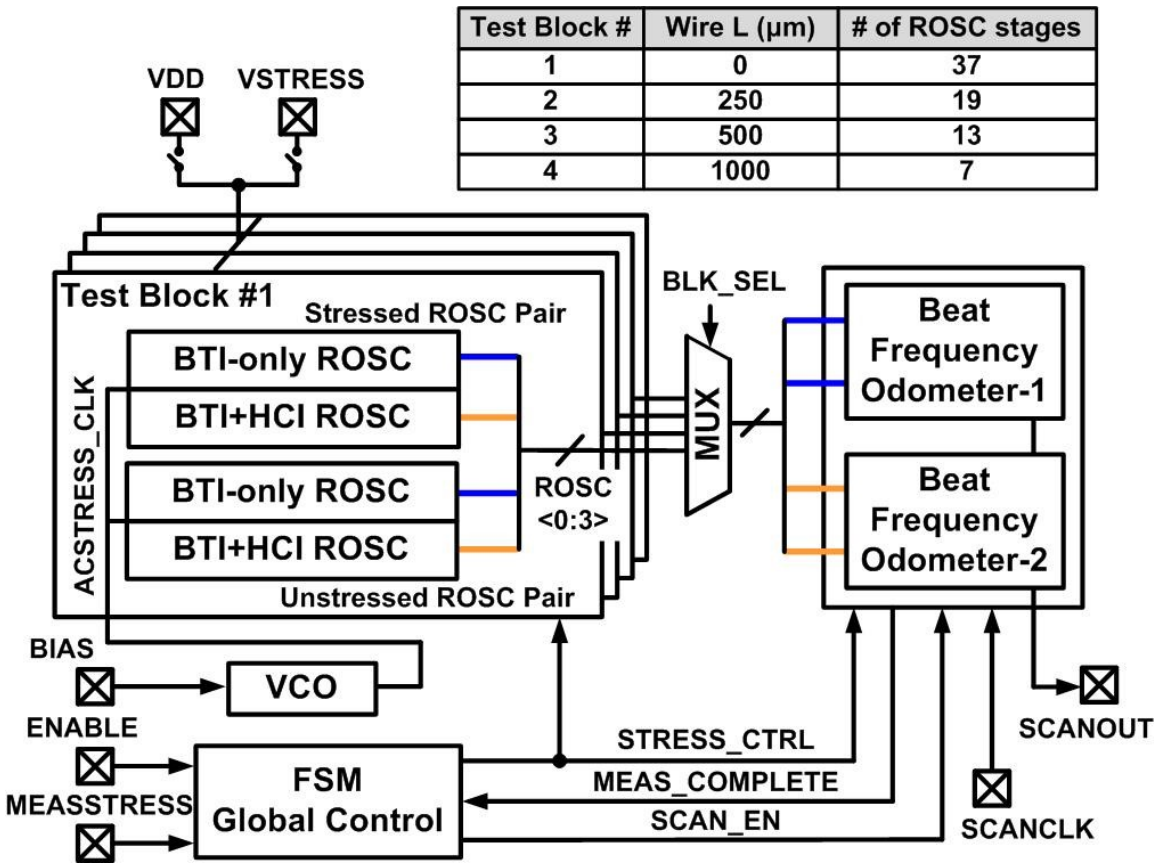


Figure 2.1: Interconnect odometer test chip diagram. Four ROSCs (stressed pair and unstressed pair) are used for each wire configuration to separately monitor BTI and HCI induced frequency shifts.

The top level block diagram of the interconnect odometer test chip is shown in Fig. 2.1. Four ROSC configurations with different interconnect lengths of 0μm, 250μm, 500μm, and 1000μm were implemented. Transistor dimensions of each ROSC stage are $(W/L)_{PMOS}$=6μm/0.06μm and $(W/L)_{NMOS}$=3μm/0.06μm regardless of the interconnect length. Although this configuration may not represent an interconnect circuit optimized for speed, it allows us to separate out the impact of interconnect length on frequency degradation. Among the four ROSCs, one undergoes BTI stress only, one undergoes both BTI and HCI stress, and the other two are remained unstressed serving as a frequency reference point. Each stressed oscillator is paired up with its unstressed counterpart, and fed into a beat-frequency detection system through multiplexers. On-chip power gates provide fast local stress voltage switching in the nanosecond order while a voltage controlled oscillator generates an AC stress frequency.

2.2.1 Beat-frequency detection technique

The beat frequency odometer system in Fig. 2.2 measures the percentage change in the stress ROSC frequency. We include a brief explanation of the beat frequency odometer system for convenience but further details can be found in [36]. The output of the reference ROSC is used as clock of the D flip-flop (DFF) to sample the stressed ROSC output. The initial frequency of the reference ROSC (f_{ref}) is set using trimming capacitors to be slightly higher than that of the stressed ROSC (f_{stress}). The DFF output toggles from low to high whenever the rising edge of the two ROSC outputs overlap. In other words, the output of the DFF exhibits the beat frequency f_{beat} defined as f_{ref} - f_{stress}.

A counter is implemented at the output of the DFF to record the number of reference ROSC periods corresponding to the beat period. The count is registered after each stress period, and the frequency shirt in the stress ROSC can be conveniently calculated using straightforward algebraic equations shown in Fig. 2.2.

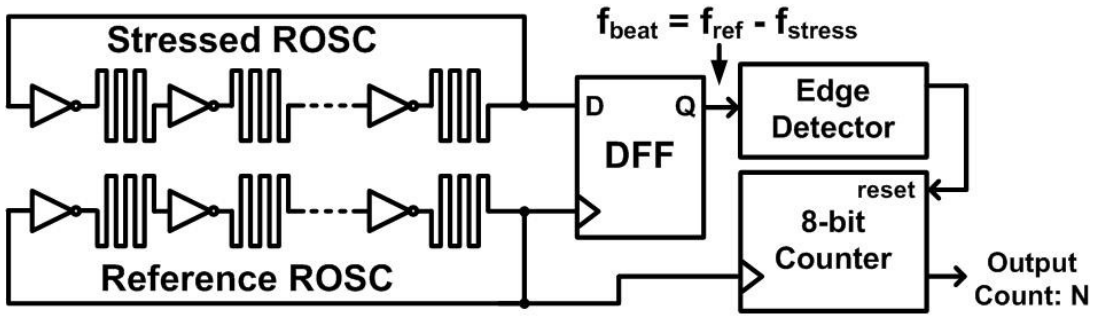


Figure 2.2: Beat frequency odometer system used in this work. N1 and N2 are the counts from the counter output, recorded before and after a certain stress period. Using the equations listed above, we can calculate the percentage frequency change with picosecond resolution and sub-microsecond measurement time.

The main highlight of the beat frequency odometer system is that it provides extremely high-resolution frequency shift measurements (>0.01%) with microsecond order measurement interruption which eliminates the unwanted BTI recovery effects. A detail comparison of various ROSC based frequency degradation measurement techniques can be found in our previous all-in-one odometer paper [35, 46].

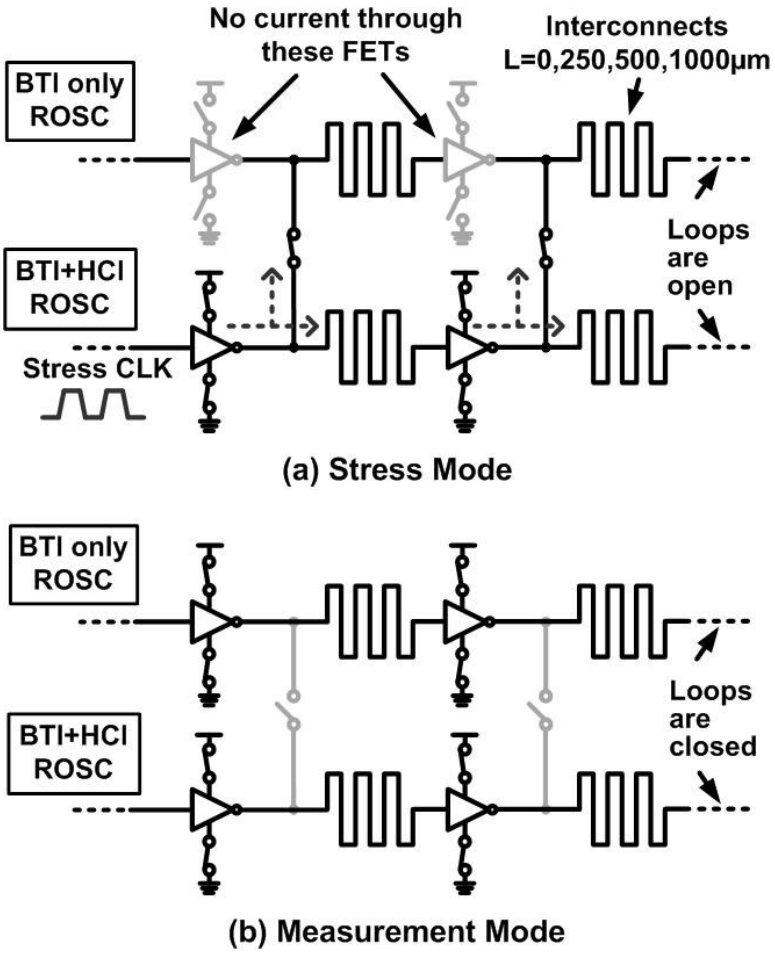


Figure 2.3: ROSC pair configuration in stress and measurement modes: the top ROSC exhibits the same amount of BTI as the bottom ROSC but without any HCI .

2.2.2 Separately monitoring BTI and HCI

The BTI and HCI contributions were separately measured by adopting the "all-in-one" odometer concept illustrated in Fig. 2.3 [35]. In stress mode, the top ROSC is gated off from the supply with the bottom ROSC driving the inputs and outputs of both ROSCs. Using this configuration, the transistors in the top ROSC experience the same BTI stress

condition as those in the bottom ROSC but with negligible HCI degradation. Note that electromigration and Joule heating effect in the wires was negligible due to the small average current and RMS current according to the 65nm process used for the test chip and therefore the frequency degradation is purely due to the transistor aging. In measurement mode, the frequency degradations of the two stressed ROSCs are measured using two silicon odometer beat frequency detection systems. The HCI-induced aging can be then obtained by subtracting out the BTI component (top ROSC in Fig. 2.3) from the combined BTI+HCI effect (bottom ROSC in Fig. 2.3).

2.2.3 Layout details

The layout of a ROSC pair with a per-stage interconnect length of 1000μm is shown in Fig. 2.4. A stressed ROSC and its fresh counterpart are placed right next to each other symmetrically to minimize any systematic variation due to layout mismatches, voltage gradients, and temperature differences. Minimum width M2 wires were laid out in a serpentine manner to fit the long interconnect in the given die area. In order to minimize any coupling effect occurring at the end of each metal segment, double shielded wires with minimum metal-to-metal spacing was used. We understand this may increase the capacitive loading of the signal interconnect compared to a typical interconnect circuit. However, our analysis shows that by parameterizing circuit parameters such as wire capacitance, wire resistance, and the driver's equivalent resistance values, we can build universal aging models applicable to a wide range of interconnect designs across difference technologies and operating conditions. Further details on this point can be

found in Section IV. The die photo and key features of the 65nm interconnect odometer test chip are given in Fig. 2.5.

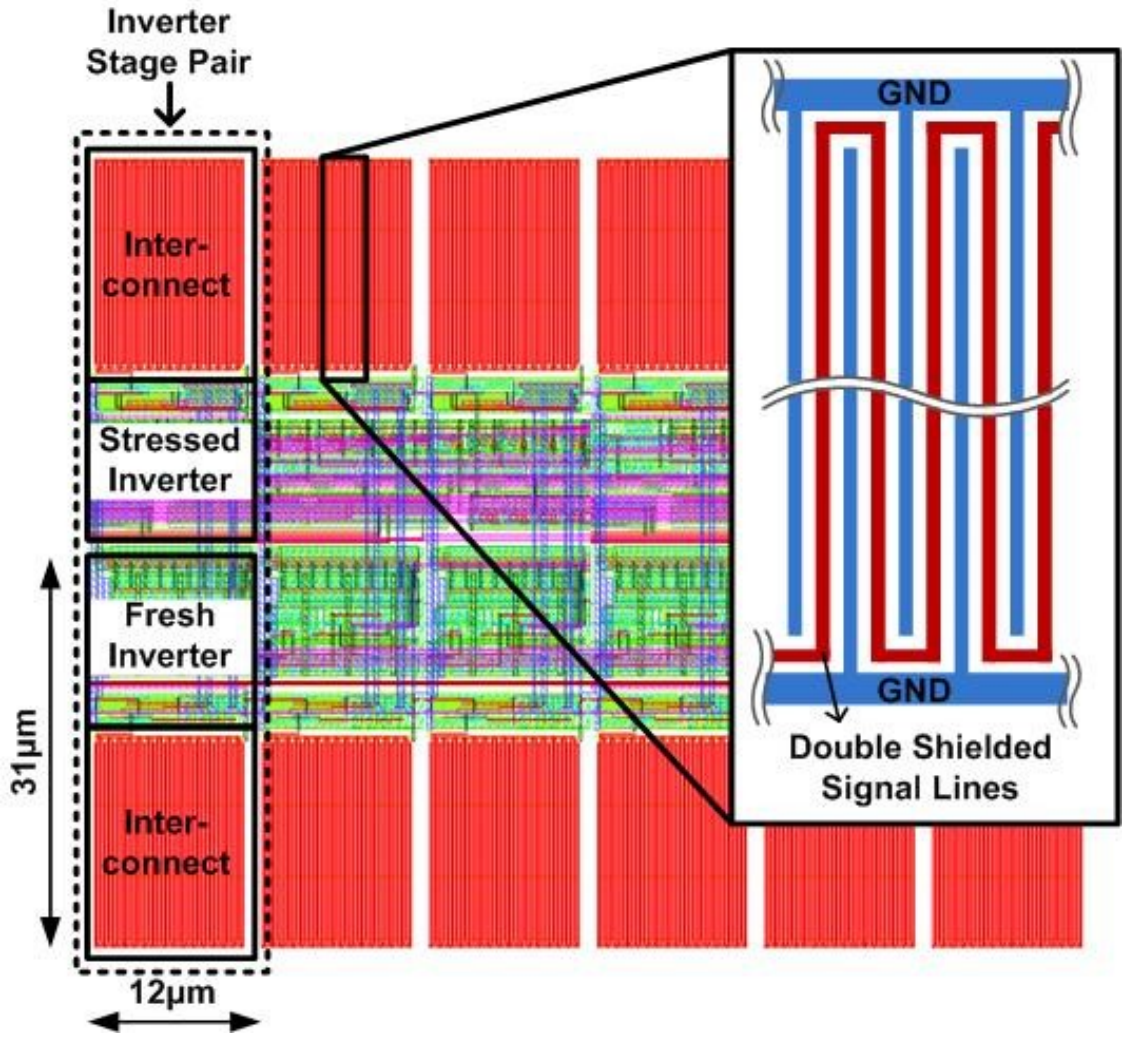
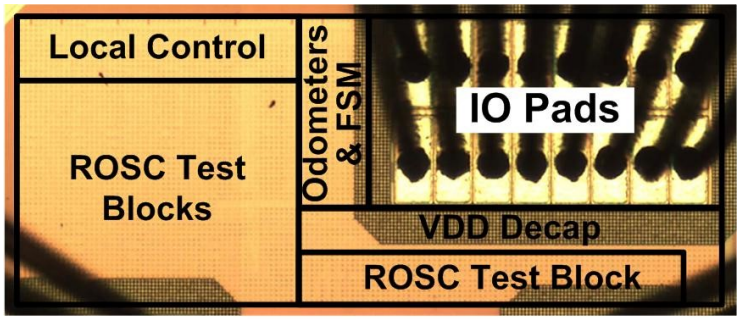


Figure 2.4: The ROSC pair layout is symmetric and contains double shielded signal wires.

2.3 Testchip Results

2.3.1 ROSC aging measurements



Process	65nm LP CMOS
Core / IO Supplies	1.2V / 2.5V
Stress Voltage	1.8V, 2.4V
Active Area	0.182mm^2
Interconnect Layer	M2, W=100nm, double shielded w/ 100nm spacing
Δf Resolution	> 0.016%
Meas. Interrupt	< 3μs

Figure 2.5: Die photo and feature summary of the 65nm interconnect odometer test chip.

We first present experimental data from the test chip in Fig. 2.6 showing the general behavior of HCI and BTI degradation from ROSCs without any long interconnects between the inverter stages. As expected, BTI shows a positive dependence on temperature, while HCI has a slightly negative dependence on temperatures which is mainly due to the reduced drain current resulting from increased phonon scattering (Fig. 2.6 (a)). BTI is at best weakly dependent on the frequency as verified in Fig. 2.6 (b),

while HCI degradation increases at higher frequencies due to the higher switching activity. Fig. 2.6 (c) shows that both aging mechanisms worsen at higher stress voltages with HCI displaying a stronger voltage dependence. Both BTI and HCI degradations show good agreement with simple power law models (i.e. t^n) using fitted power law exponents n which are denoted in the figure. BTI is the primary aging contributor at early stress times while HCI with its 4-6 times larger power law exponent surpasses BTI at longer stress times, which results in a cross over point between the two trend lines.

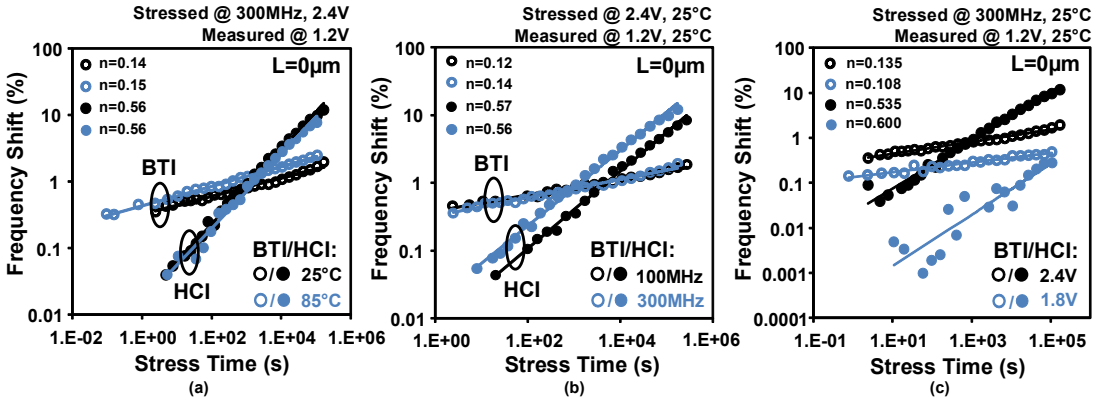


Figure 2.6: Measured BTI and HCI contribution under different (a) temperatures, (b) frequencies, and (c) stress voltages. The HCI component is obtained by subtracting out the BTI-only degradation from the BTI+HCI degradation.

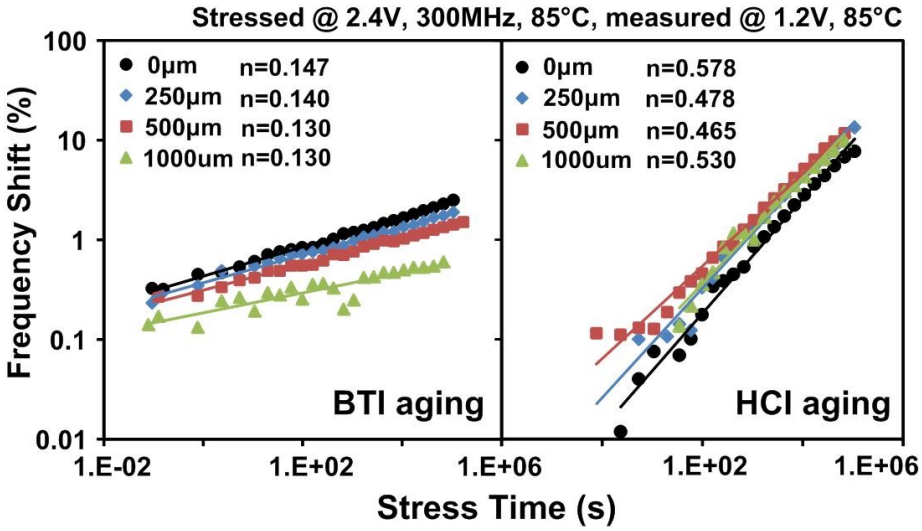


Figure 2.7: Measured frequency degradation induced by BTI and HCI for different interconnect lengths. BTI is the worst at L=0µm while HCI is the worst at L=500µm.

Having confirmed the general behavior of HCI and BTI through repeated tests, we carried out stress experiments on the ROSCs with long interconnects. Fig. 2.7 shows the BTI and HCI degradation versus stress time for different interconnect configurations. To our surprise, under an identical stress condition, the ROSC with no interconnect shows the worst BTI degradation, while the ROSC with an interconnect length of 500µm had the worst HCI degradation. The time it takes for HCI to overtake BTI (i.e. crossover time) decreases by almost three orders of magnitude as the interconnect length is increased from 0µm to 1000µm as plotted in Fig. 2.8 (b) under a stress condition of 2.4V, 300MHz at 85°C. In the following sections, we further examine the unexpected BTI and HCI trends shown in Fig. 2.7 which clearly indicates a difference in their wire length dependencies.

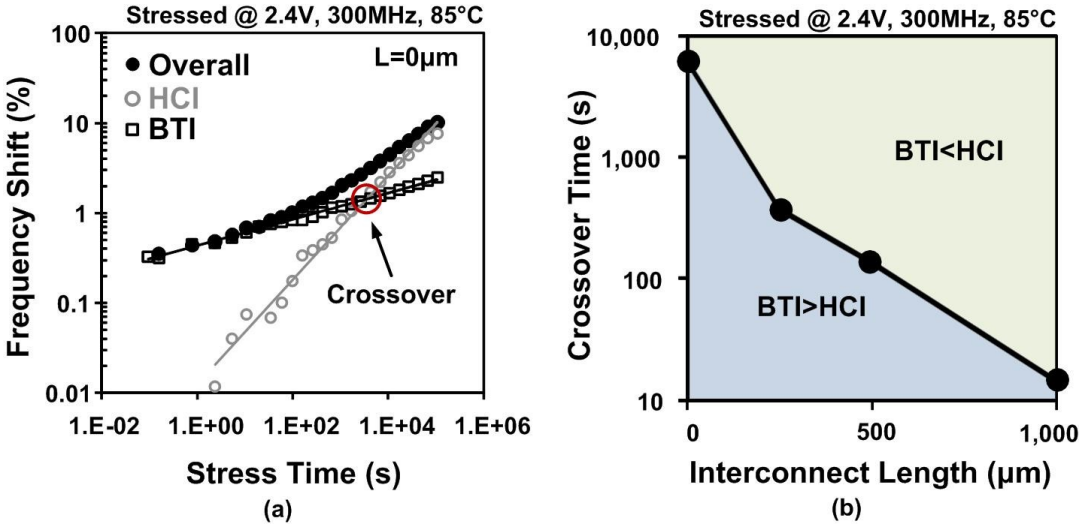


Figure 2.8: (a) The BTI and HCI components crossover as the BTI dominates initially while the larger slope of HCI makes it dominate at longer stress times. (b) The crossover time becomes shorter with a longer interconnect due to the larger HCI.

2.3.2 Interconnect length vs. BTI aging

BTI induced frequency shifts after 19 hours of stress at 2.4V are shown in Fig. 2.9 for different interconnect lengths. BTI aging decreases monotonically with longer interconnects for all three stress conditions. This can be explained by the longer signal transition time in longer wires which translates into a shorter time in which the transistors are actually exposed to a full BTI stress. Simulated waveforms in Fig. 2.10 show a longer slew rate for longer wires. The results summarized in the table confirms a 6.2X longer transition time ($t_T=t_R+t_F$) and a 9.6% reduction in the effective stress duty cycle (t_L/τ) as the wire length is increased from 0μm to 1000μm for a driver size of

$(W/L)_{PMOS}$=6µm/0.06µm and $(W/L)_{NMOS}$=3µm/0.06µm. Note that PBTI in NMOS is negligible in this 65nm process as it does not employ high-k metal-gate process. However, the general behavior will be exacerbated in the presence of PBTI as the duty cycle for the NMOS is also reduced for longer interconnects. Fig. 2.9 also reveals a weaker dependency on interconnect length for lower AC stress frequencies. This can be attributed to the smaller fraction of time spent for signal transition at lower input frequencies, which makes the duty cycle less sensitive to the interconnect length as shown in the simulations for different AC stress periods in Fig. 2.11.

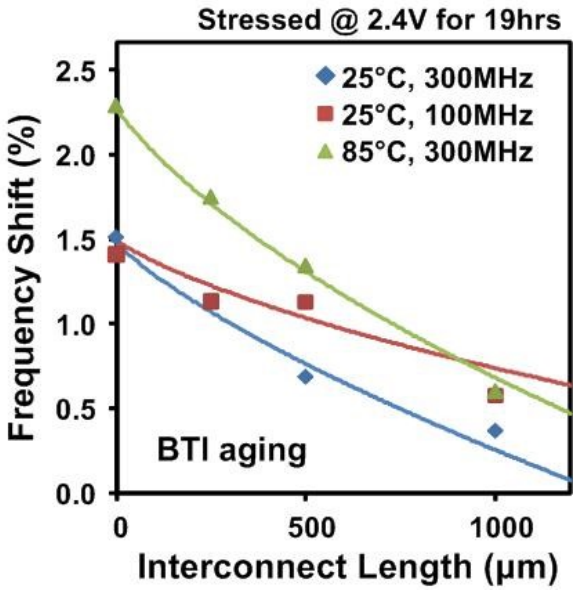
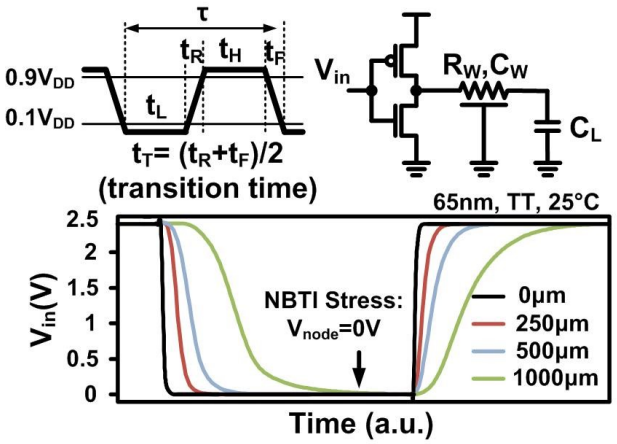


Figure 2.9: Measured data (markers) and modeling results (curves) for BTI induced frequency degradation.



Int. L (μm)	t_T (ps)	t_L/τ (effective duty cycle)	
		100MHz	300MHz
0	40.4	49.7%	49.3%
250	80	49.2%	47.0%
500	130	48.8%	45.2%
1000	250	47.6%	40.4%

Figure 2.10: Effective stress time (t_L) decreases in longer interconnects resulting in a smaller BTI degradation as shown in the measured data in Fig. 2.9.

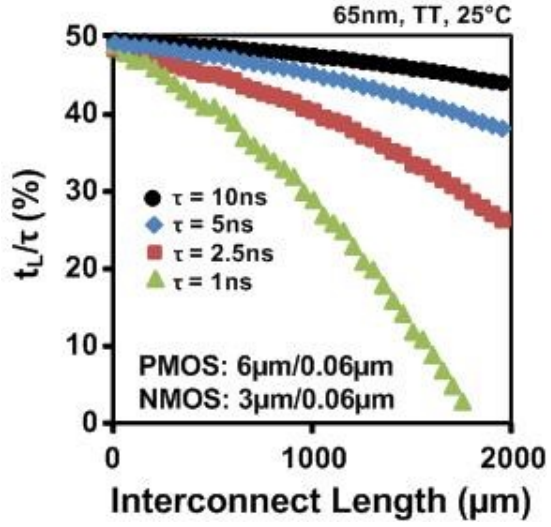


Figure 2.11: The time during which the device is exposed to a full BTI stress (i.e. t_L/τ) decreases with a longer interconnect due to the slower signal transitions.

2.3.3 Interconnect length vs. HCI aging

The effect of HCI in Fig. 2.12 shows a non-monotonic relationship with wire length. This is somewhat counter-intuitive but can be explained using the following two reasons.

1) A driver with a longer wire has a smaller peak current due to the voltage division effect between the wire resistance and the driver's equivalent resistance as shown in Fig. 2.13. In other words, the effective stress voltage V_{eff} that dictates the amount of HCI decreases with a longer interconnect. Additional simulation results in Fig. 2.14 confirm that the peak discharging current (I_{peak}) through the NMOS decreases with longer interconnect due to the aforementioned voltage division effect. For the simulation, we use a driver size of $(W/L)_{PMOS}$=6μm/0.06μm and $(W/L)_{NMOS}$= 3μm/0.06μm and a distributed RC wire model to obtain accurate results. The reduction of the peak current has a similar effect as having a lower effective stress voltage and therefore leads to a smaller frequency shift. Note that the peak current may actually increase for wire lengths from 0μm to 200μm due to the fast input slew rate that causes the NMOS to turn off before it enters the saturation mode.

2) A longer wire makes the current pulse wider due to the larger wire capacitance which has the effect of increased effective stress time (t_{eff}) compared to a shorter wire. A longer current pulse width shown in Fig. 2.13 is equivalent to an increased HCI stress time which leads to increased HCI degradation for longer interconnects [19].

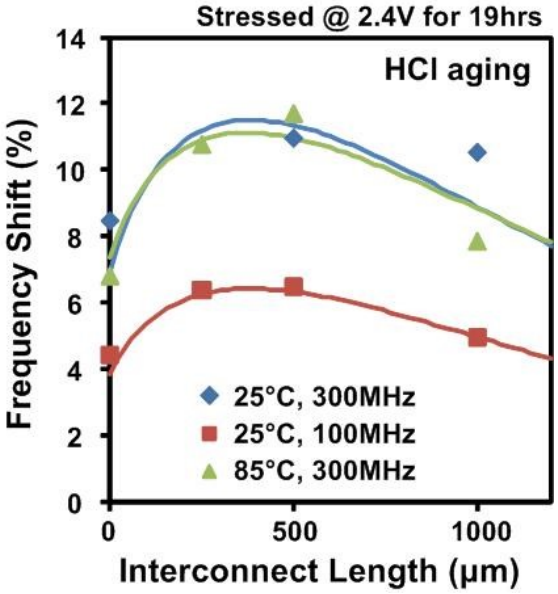
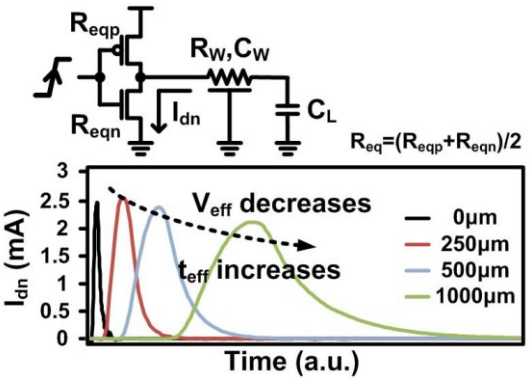


Figure 2.12: Measured data (markers) and modeling results (curves) for HCI induced frequency degradation.



Wire Len. (μm)	I_{peak}(mA)	t_{eff} (ps, eff. stress time)	V_{eff} (V, eff. V_{stress})
0	2.50	6.3	2.40
250	2.56	25	2.34
500	2.39	51	2.29
1000	2.12	127	2.19

Figure 2.13: In longer interconnects, the effective stress time increases while the effective stress voltage decreases resulting in the non-monotonic HCI trend in Fig. 2.12.

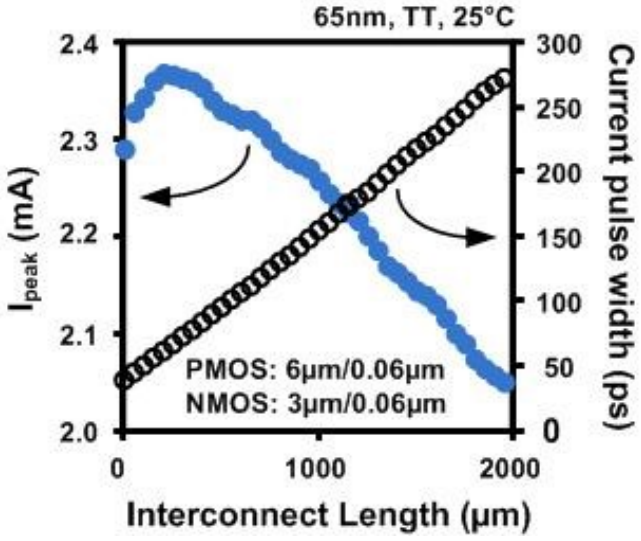


Figure 2.14: HCI parameters versus interconnect length. Peak current decreases while the pulse width increases for longer interconnects. The combined effect is a non-monotonic dependence of HCI induced frequency degradation on interconnect length.

The combined effects of 1) and 2), coupled with their different sensitivities on the frequency degradation, result in the non-monotonic relationship between HCI induced frequency shift and interconnect length.

2.4 Aging Models for Interconnect Driver

As we saw in the previous sections, frequency degradation due to BTI and HCI depends on the transition time and bias condition which are both strong functions of the interconnect load. However, none of the existing circuit aging models incorporate interconnect related parameters for estimating BTI and HCI induced degradation. In this section, we propose analytical models applicable to global interconnect drivers which

agree well with the experimental results in section III. The general approach for modeling the frequency degradation in global interconnects is described below.

- Step 1: The frequency degradation of an interconnect dominated path is less sensitive to the device aging compared to a logic dominated path due to the time invariant interconnect RC delay. We account for this difference by introducing a sensitivity factor α.

- Step 2: The amount of BTI and HCI aging depends on the effective stress time and effective stress voltage that vary with interconnect length. Existing BTI and HCI models with modified stress time and stress voltage are used to derive the final model.

Section IV.A describes the above step 1 where we will first analyze the sensitivity factor and derive its mathematical expression. In sections IV.B and IV.C, we perform step 2 where the final analytical BTI and HCI models for interconnect drivers are derived.

2.4.1 Sensitivity factor

Delay degradation of an interconnect dominated-path is less sensitive to the transistor parametric shift compared to its logic-dominated counterpart. This can be seen from the HSPICE results in Fig. 2.15 where a device V_t shift of 30% translates into a 12% frequency degradation for a ROSC with a 1000μm interconnect per stage, whereas the degradation for a ROSC with no interconnects is 15%. This effect can be easily captured in our model by introducing a sensitivity factor α defined as the ratio between the %

frequency degradation of an interconnect dominated path and that of a logic dominated path for the same amount of device aging:

$$\left(\frac{\Delta f}{f}\right)_{interconnect} = \alpha \left(\frac{\Delta f}{f}\right)_{logic} \tag{2.1}$$

For a given interconnect resistance (R_W), interconnect capacitance (C_W), load capacitance (C_L), equivalent driver resistance before (R_{eq}) and after stress ($\Delta R_{eq}=R_{eq}'-R_{eq}$), and the inverter stage number N_{stage}, the ROSC period can be calculated as:

$$\tau = 2N_{stage}\left[R_{eq}(C_W + C_L) + R_W\left(\frac{C_W}{2} + C_L\right)\right] \tag{2.2}$$

The % frequency degradation can then be expressed as:

$$\left(\frac{\Delta f}{f}\right)_{interconnect} = \frac{\Delta R_{eq}(C_W+C_L)}{R'_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)} \tag{2.3}$$

Note that R_W and C_W of a wire can be simply calculated from the sheet resistance and metal capacitance parameters. The percentage frequency degradation for a logic only path can be written as:

$$\left(\frac{\Delta f}{f}\right)_{logic} = \frac{\Delta R_{eq} C_L}{R'_{eq} C_L} = \frac{\Delta R_{eq}}{R'_{eq}} \tag{2.4}$$

Using equations (3) and (4), the expression of α can be derived as:

$$\alpha = \left(\frac{\Delta f}{f}\right)_{interconnect} / \left(\frac{\Delta f}{f}\right)_{logic}$$

$$= \frac{\Delta R_{eq}(C_W+C_L)}{R'_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)} \frac{R'_{eq}}{\Delta R_{eq}} = \frac{R'_{eq}(C_W+C_L)}{R'_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)}$$

$$= \frac{R'_{eq}(C_W+C_L)}{R'_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)} \tag{2.5}$$

The above sensitivity factor will be applied to the BTI and HCI models proposed in sections B and C.

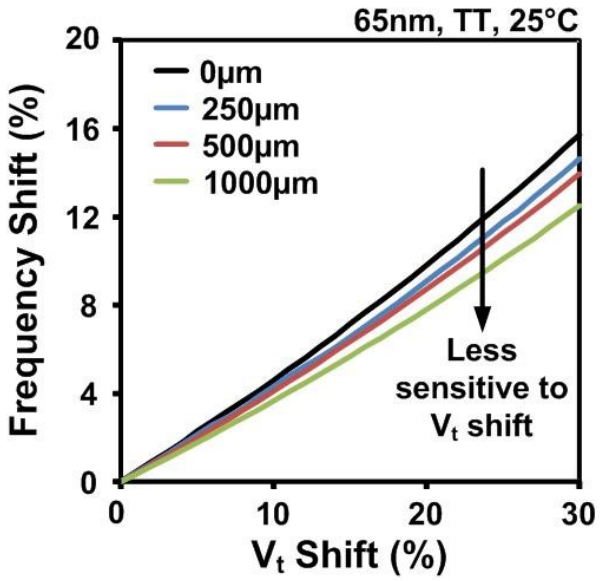


Figure 2.15: Frequency shift vs. V_t shift for different interconnect lengths. Frequency shift of an interconnect-dominated path is less sensitive to V_t shifts compared to a logic-dominated path due to the constant wire RC delay. This effect is captured using the sensitivity parameter α in section IV-A for an accurate aging estimation.

2.4.2 BTI aging model for interconnect drivers

While the detailed physics for BTI are still under debate, the following two models are generally considered for BTI induced degradation: the reaction-diffusion (R-D) model which follows a power law time dependence ($\propto t^n$), and the trapping-detrapping model

which follows a logarithmic time dependence ($\propto \log(Bt+1)$). Our experimental data shows good agreement with the power law dependence under different stress conditions and hence we report the time exponent values in this work.

BTI aging is a function of device on time. Hence, in the context of an interconnect driver, it can be expressed using the cycle time parameter t_L/τ, where t_L and τ are defined in Fig. 2.10. Employing the models in [6, 47] with dynamic stress modeling methodology in [48] adopted, the deviation of BTI from the ideal 50% duty cycle case can be expressed using $(50\%-t_L/\tau)f$, where f is determined empirically. The overall BTI induced frequency shift can be expressed as:

$$\left(\frac{\Delta f}{f}\right)_{BTI} = \left(\frac{\Delta f}{f}\right)_{@50\%}\left[1 - B\left(50\% - \frac{t_L}{\tau}\right)^f\right]$$

$$= A(V_{str} - V_t)^{-0.5} e^{(\gamma V_{str})} e^{\left(-\frac{E_a}{kT}\right)} t^n \left[1 - B\left(\frac{2t_T}{\tau}\right)^f\right] \quad (2.6)$$

Here, γ is the voltage acceleration factor; E_a is the temperature activation energy. Both parameter values are reaction characteristics related coefficients obtained from silicon results under different stress voltage and temperature conditions. V_{str} is the stress voltage, V_t is threshold voltage, t is the BTI stress time of a logic only path, n is the BTI time exponent, t_T is the transition time, and τ is the period of AC stress cycle. A, B and f are empirical parameters found to be 3.3×10^{-4}, 5.0 and 0.70 in the 65nm technology used for this work. The transition time t_T is interconnect RC dependent, which can be simply denoted as by using the text book Elmore model:

$$t_T = R_{eq}(C_W + C_L) + R_W\left(\frac{C_W}{2} + C_L\right) \tag{2.7}$$

Using the sensitivity factor from section A, the overall BTI frequency degradation for long interconnects can be derived as:

$$\left(\frac{\Delta f}{f}\right)_{BTI_{interconnect}} = \alpha\left(\frac{\Delta f}{f}\right)_{BTI}$$

$$= \frac{R'_{eq}(C_W+C_L)}{R'_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)}(V_{str}-V_t)^{-0.5}e^{(\gamma V_{str})}e^{\left(-\frac{E_a}{kT}\right)}t^n$$

$$\cdot\left[1 - B\left(\frac{2R_{eq}(C_W+C_L)+2R_W\left(\frac{C_W}{2}+C_L\right)}{\tau}\right)^f\right] \tag{2.8}$$

2.4.3 HCI aging model for interconnect drivers

From the general HCI models [16, 49], the degradation of frequency can be approximated as follows assuming a simple Arrhenius dependency:

$$\left(\frac{\Delta f}{f}\right)_{HCI} = \frac{C}{V_{str}-V_t}e^{\left(\frac{E_b}{kT}\right)}e^{\left(-\frac{D}{V_{eff}}\right)}t_{eff}^m \tag{2.9}$$

where C, D, and m are empirical parameters, t_{eff} is the effective HCI stress time which is directly related to the signal transition time, and V_{eff} is the effective drain to source voltage during stress. The experimental and simulation results in section III show that both the effective stress voltage and effective stress time depend on the interconnect RC load. As noted earlier, a long interconnect causes the output signal voltage to be divided between the interconnect resistance RW and the transistor drain-to-source (V_{ds})

while charging and discharging. So the effective HCI stress voltage can be approximated as:

$$V_{eff} = \frac{R_{eq}}{R_{eq}+R_W} V_{str} \tag{2.10}$$

where R_{eq} is the equivalent driver resistance, R_W is the interconnect resistance, and V_{str} is the HCI stress voltage in a path without interconnect.

Under the assumption that the HCI stress time is proportional to the transition time, the effective stress time considering interconnect impact can be expressed as:

$$t_{eff} = \frac{R_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)}{R_{eq}C_L} t \tag{2.11}$$

Here, t is the time a device in a logic dominated path is under HCI stress. Finally, the HCI induced frequency degradation can be derived using the sensitivity factor introduced in section A:

$$\left(\frac{\Delta f}{f}\right)_{HCI_{interconnect}} = \alpha \left(\frac{\Delta f}{f}\right)_{HCI}$$

$$= \frac{R'_{eq}(C_W+C_L)}{R'_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)} \cdot \frac{C}{V_{str}-V_t} e^{\left(\frac{E_b}{kT}\right)} e^{\left(-\frac{D}{V_{eff}}\right)}$$

$$\cdot \left[\frac{R_{eq}(C_W+C_L)+R_W\left(\frac{C_W}{2}+C_L\right)}{R_{eq}C_L} t\right]^m \tag{2.12}$$

The results from (8) and (12) are overlaid onto the measured data in Figs. 2.9 and 2.12 showing good agreement with actual hardware data. This verifies that the equations

based on the various interconnect parameters are capable of modeling the monotonic behavior of BTI and non-monotonic behavior of HCI for a wide range of interconnect lengths. Note that both models capture the voltage and temperature dependencies using the voltage acceleration factor γ, D and activation energy E_a, E_b respectively. The models with parameters used are summarized in Fig. 2.16.

BTI Induced Freq. Degradation:

$$(\Delta f/f)_{BTI} = \alpha A(V_{str}-V_t)^{-0.5} \cdot exp(\gamma V_{str}) exp(-E_a/kT) t^n [1-B(2t_T/\tau)]^f$$

Interconnect related parameters			
t_T	$R_{eq}(C_W+C_L)+R_W(C_W/2+C_L)$		
α	$\frac{R_{eq}(C_L+C_W)}{R_{eq}(C_W+C_L)+R_W(C_W/2+C_L)}$		
Aging mechanism parameters			
$\gamma(V^{-1})$	2.13	E_a(eV)	0.070
Fitting parameters			
A	3.3×10^{-4}	B	5.0
n	0.14	f	0.70

HCI Induced Freq. Degradation:

$$(\Delta f/f)_{HCI} = C(V_{str}-V_t)^{-1} exp(E_b/kT) \cdot exp(-D/V_{eff}) t_{eff}^{\ m}$$

Interconnect related parameters			
V_{eff}	$\frac{R_{eq}}{R_{eq}+R_W} V_{str}$		
t_{eff}	$\frac{R_{eq}(C_W+C_L)+R_W(C_W/2+C_L)}{R_{eq}C_L} t$		
Aging mechanism parameters			
D(V)	48	E_b(eV)	0.035
Fitting parameters			
C	1.23×10^{-4}	m	0.50

Figure 2.16: Summary table of proposed BTI and HCI model for interconnect drivers with parameter values.

2.5 Aging Models for Interconnect Driver

In this section, we present a closed loop modeling methodology that utilizes the aforementioned circuit aging models for accurate prediction of interconnect performance degradation. The impact of interconnect geometry on the overall performance

degradation was also examined. Finally, we show a case study of a global interconnect path design which uses the proposed closed loop design methodology to determine the optimal repeater count and sizing considering BTI and HCI effects.

2.5.1 Closed loop aging calculation for interconnect paths

For logic dominated paths, the delay degradation can be simulated by simply plugging in parameters such as stress voltage, stress time, activity factor, and switching frequency into a device reliability model. However, this "open loop" approach will not yield accurate results for interconnect dominated paths as the degradation in the driver circuit itself leads to a change in the stress condition which in turn affects the amount of HCI and BTI aging in the driver. In other words, degradation of an interconnect driver and the stress condition are interdependent. For example, in case of BTI, the transition time t_T (Fig. 2.10) determined by the equivalent driver impedance (R_{eq}) increases with stress time as a result of aging. This increase in t_T reduces the amount of BTI in the driver giving rise to a compensatory effect. Similarly, in case of HCI, the effective stress voltage (V_{eff}) and the effective stress time (t_{eff}) are functions of R_{eq}, so the actual stress condition changes depending on the transistor aging.

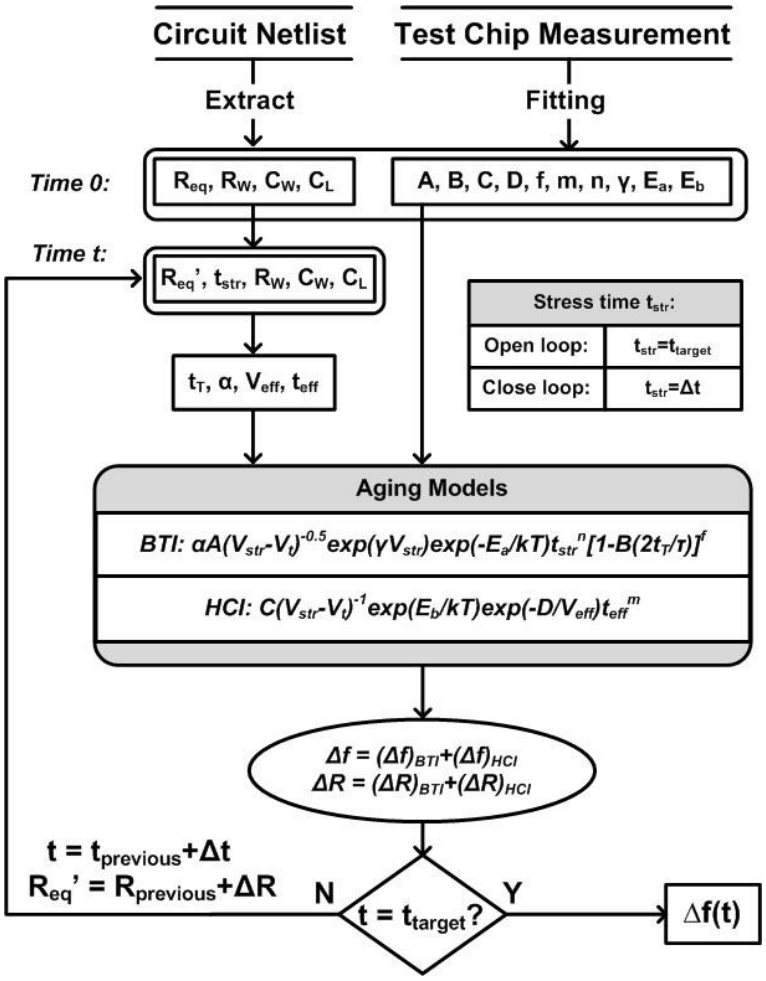


Figure 2.17: Closed loop calculation of interconnect driver aging incorporating the interconnect RC impact. Transistor parameters are updated every time step based on the aging and stress conditions in the previous stress time interval Δt.

To account for the interplay between the device aging and the stress condition in interconnect drivers, we propose a closed loop calculation method that feeds the parametric shift information back into the original aging model for accurate estimation of interconnect driver aging. The general flow of the closed loop calculation method is

shown in Fig. 2.17. First, the fitting constant in the BTI and HCI aging models such as A, B, C, D, f, m, n, γ, E_a and E_b are acquired from the curve fitting on testchip measurements. Then for the circuit of interest, the fresh electric parameters such as R_{eq}, R_W, C_W, and C_L are extracted from the layout design. These parameters are then used to calculate the stress related parameters such as t_T, α, V_{eff} and t_{eff}. To estimate the frequency degradation at time $t_{previous}+\Delta t$, the circuit parameters (i.e. R_{eq}, R_W, C_W, and C_L) at the previous time $t_{previous}$ and the additional stress period (Δt) are applied to the aging models. Additionally, the interconnect driver resistance R_{eq} is updated for the next time interval. The same procedure repeats until the total stress time $t_{str} = t_0+N\cdot\Delta t$ (N is the number of iteration) reaches the target stress time t_{target}. Note that the end results are the superposition of both BTI and HCI aging, as the overall shift value is the sum of the frequency shifts predicted by the models. A smaller time step used for the iteration will give more accurate aging results at the expense of a longer computation time.

The comparison between the open loop and closed loop methods for estimating the overall aging induced frequency shift $\Delta f\% = (\Delta f\%)_{BTI} + (\Delta f\%)_{HCI}$ are shown in Fig. 2.18. As mentioned before, there is a trade-off between time step and accuracy. That is, the prediction is more accurate with a smaller time step to update the parameters. We chose a time step of 10 minutes for the particular stress condition in Fig. 2.18 because the frequency shift difference between 10 minutes and 6 seconds is only 1.8×10^{-5}%, which is negligible comparing to the total frequency shift which is in the range of 3~8%. For different stress conditions and stress times, the minimum time step can be determined

based on the maximum error that can be tolerated. The results show that the open loop method underestimates the frequency shift by up to 4.3%. For extremely short (<100μm) or long (>2000μm) interconnects, the two calculation methods give almost the same results. This is because the transistor degradation has negligible influence on the stress condition for short interconnects while the delay itself is a weaker function of the drive resistance R_{eq} for long interconnects.

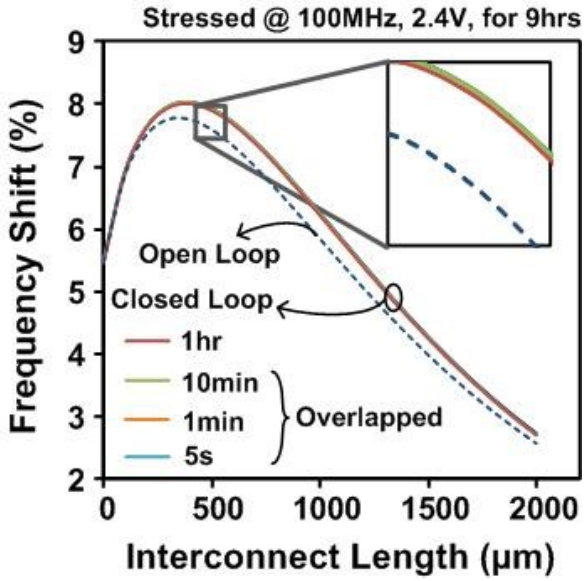


Figure 2.18: Comparison between open loop (=direct calculation) and closed loop (=parameters updated each time interval) aging results as a function of interconnect length. The impact of the close loop time step on frequency shift results is negligible when the interval is smaller than 10min for this stress condition.

Note that we used the open loop calculation to fit our experimental data in the previous section only because it was a sufficient and simple way for explaining the monotonic decrease of BTI and the non-monotonic behavior of HCI. A close loop model

however should be used for more accurate results in the general case. Although the discrepancy between the two methods may seem rather small in the 65nm technology used for this work, not accounting for these effects may lead to unforeseen interconnect timing failures in future technology nodes where aging-induced degradation is expected to worsen.

2.5.2 Interconnect geometry dependence

The proposed aging calculation method can be useful for studying the dependence of BTI and HCI on various interconnect geometries such as wire width, spacing between wires, and metal layer. Since the interconnect resistance R_W and capacitance C_W can be expressed using these interconnect geometrical parameters, we can apply the proposed methodology to a wide range of interconnect designs (e.g. clock network, global bus, memory, I/O) across different process technologies. To demonstrate the practicality of our approach, we provide an example wherein the impact of interconnect length and width on delay degradation is analyzed. The overall procedure is as follows. First, the relationships between the interconnect RC parameters (R_W and C_W in Fig. 2.17) and the wire length and width are enumerated. For example, the total wire resistance R_W is proportional to the wire length but inversely proportional to the wire width. The total wire capacitance C_W, on the other hand, must be decomposed into $C_{W\_HORIZONTAL}$ and $C_{W\_VERTICAL}$ as the two components have different dependencies on the wire geometry. Here, $C_{W\_HORIZONTAL}$ is defined as the capacitance seen by the wire's side plates while $C_{W\_VERTICAL}$ is the capacitance seen by the wire's bottom and top plates. $C_{W\_HORIZONTAL}$ is

proportional to the length but independent of wire width, while $C_{W\_VERTICAL}$ is proportional to both the wire's width and length. Based on these simple first order relationships, we can estimate the frequency degradation versus both interconnect length and width using the close loop modeling approach to obtain the results in Fig. 2.19. The minimum interconnect width for different metal layers of this technology are also listed on the same figure for reader's reference.

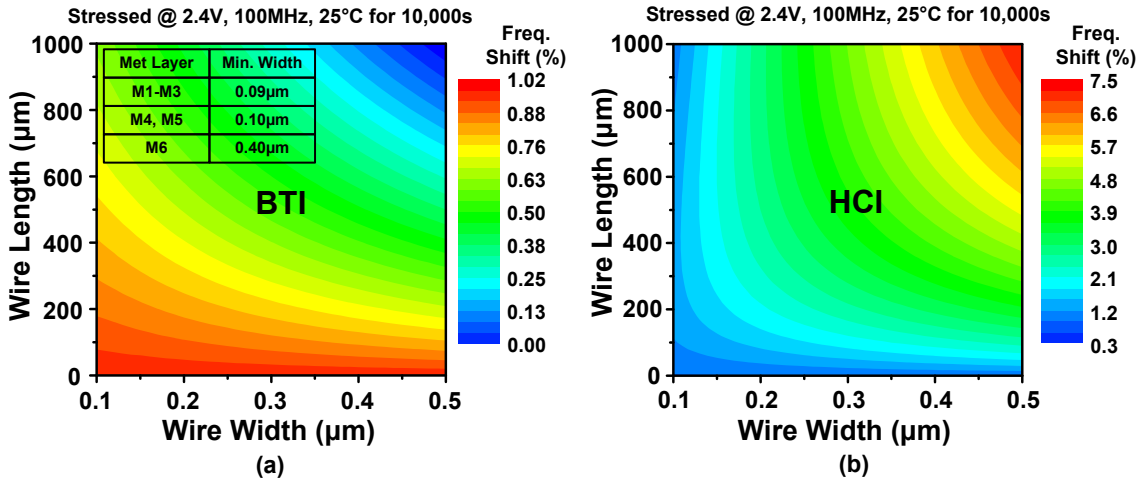


Figure 2.19: Dependence of BTI (a) and HCI (b) induced frequency shift on interconnect width and length.

Since the length dependency was already discussed extensively in section III, here we briefly comment on the width dependency of BTI and HCI degradation. Fig. 2.19 (a) indicates that a buffer driving a wider wire results in a lesser amount of BTI degradation. This is due to the increased signal delay which causes the effective BTI stress time (t_L/τ in Fig. 2.10) to go down. Contrary to the BTI results, HCI induced frequency degradation

increases with a wider wire due to the higher effective stress voltage (V_{eff} in Fig. 2.13) as shown in Fig. 2.19 (b). Note that the aging analysis provided in this section must be considered in conjunction with metrics such as performance and power consumption which are perhaps the primary design considerations.

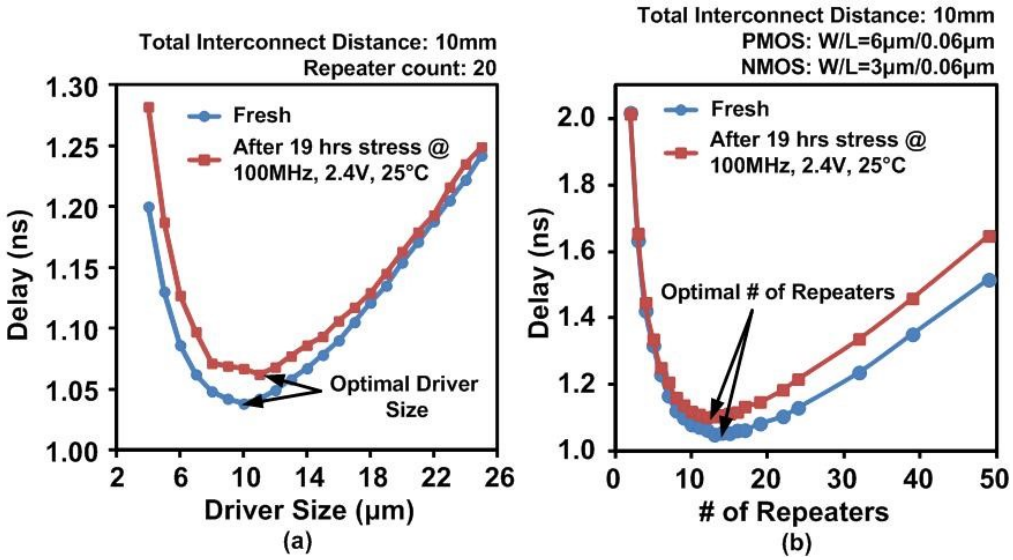


Figure 2.20: (a) Interconnect delay of a 10mm path with 20 repeaters as a function of driver size, before and after stress. (b) Interconnect delay of a 10mm path as a function of the number of repeaters, before and after stress. The sizing of each repeater was kept the same for easier comparison.

2.5.3 Interconnect path design with aging consideration

To show the application of the proposed aging models on real circuits, this section describes a methodology for estimating the optimal number of repeaters and the sizing of each repeater. To reduce the overall delay of a global interconnect path, signal wires are

typically broken into shorter segments separated by repeaters. The total propagation delay of the signal path can be expressed as:

$$t_d = N\left\{R_{eq}\left[\frac{l}{N}C_W + C_L(1 + t_{int})\right] + R_W\left(\frac{l}{N}\frac{C_w}{2} + C_L\right)\right\} \quad (2.13)$$

Here, t_{int} is the delay of an unloaded buffer, l is the total wire length, and N is the number of repeaters inserted. For a fixed number of repeaters, the transistor delay (first term in (13)) to become a larger portion of the overall propagation delay for an interconnect path with smaller drivers. This means that an interconnect path with numerous small drivers are more susceptible to aging effects compared to that with few large drivers. This behavior is verified through the simulations results in Fig. 2.20 (a) which shows larger delay degradation for smaller driver sizes. From the figure we can clearly see that a driver size of W_P/W_N=10µm/5µm achieves the minimum delay for a fresh circuit while W_P/W_N=12µm/6µm is required for minimum delay in the presence of device aging. We also varied the number of repeaters while fixing the driver size to study its impact on overall frequency degradation. The results given in Fig. 2.20 (b) show worse delay degradation for an interconnect path with a larger number of repeaters. This can be attributed to the transistor delay dominating the overall path delay as the number of repeaters is increased. Our results based on the stress condition denoted in Fig. 2.20 (b) also show that the optimal number of repeaters is 12 when aging is considered versus 14 when aging is not considered.

2.6 Conclusions

This study explores, for the first time, the detailed aging behavior of interconnect paths. The degradation in interconnect performance caused by BTI and HCI in the driver circuit exhibits a strong dependency on the interconnect RC parameters. This dependence must be thoroughly understood in order to build accurate aging models applicable to interconnect designs in advanced technologies. In this work, we first demonstrated a 65nm interconnect odometer test vehicle capable of accurately measuring the frequency degradation of signal paths with different interconnect lengths. Our previous all-in-one odometer concept was adopted to separately measure BTI and HCI aging. The measurement interrupt was kept below 3μs to avoid any unwanted BTI recovery. Experimental results show that the frequency degradation caused by BTI decreases with increasing interconnect length, while the HCI induced degradation peaks at around 500μm. This difference in the interconnect dependencies were explained using circuit simulations that account for the effective stress time and stress voltage during signal transitions. Simple circuit aging models were developed and compared against the measured data. We next proposed a closed loop modeling methodology which precisely captures the interplay between interconnect RC parameters and the change in stress condition. Finally, a case study of the closed loop model was shown where the delay shift for a practical interconnect path was minimized using optimal number of buffers.

Chapter 3.

Duty-Cycle Shift under Asymmetric BTI Aging: A Simple Characterization Method Based on Ring Oscillator and its Application to SRAM Timing

3.1 Introduction

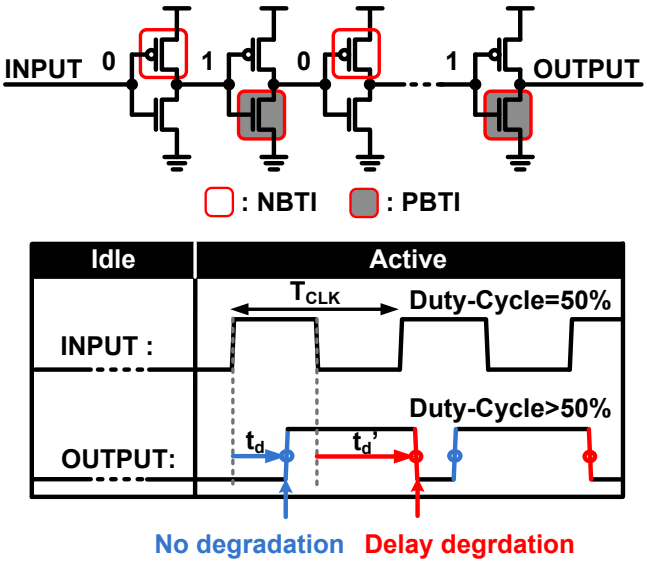


Figure 3.1: (upper) Alternating NBTI/PBTI stress in a buffer chain in idle mode. (lower) Asymmetric delay degradation of rising and falling edges results in duty-cycle shifts.

Low power SRAMs, dynamic register files, and domino gates typically rely on both the rising and falling edges of the clock to generate internal timing signals. Unlike standard flip-flop or latch based pipelines where only the primary clock edge (e.g. rising edge) is utilized, the performance of the circuits mentioned above is directly affected by any change in the clock duty-cycle. Bias Temperature Instability (BTI) stress in the clock signal path during idle or clock gated mode results in an aging-induced duty-cycle shift. Fig. 3.1 illustrates this situation in a typical clock buffer chain scenario. In an idle mode or clock gated mode, the input clock signal is not switching which results in a DC stress condition with NBTI and PBTI occurring in alternative gates. When the circuit is switched back to an active mode, the first clock signal (e.g. the first rising edge in Fig. 3.1) propagates through unstressed fresh devices while the second edge (i.e. the first falling edge in Fig. 3.1) traverses through the stressed devices. Consequently, the delay of the second edge becomes longer compared to that of the first edge due to BTI under DC stress resulting in a duty-cycle shift. Simulation results based on a 540ps delay path driven by a 1GHz clock in Fig. 3.2 show that the delay of the first edge is almost constant while the delay of the second edge is degraded by 110ps for a 20% V_t shift causing the duty-cycle to change from 50% before stress to 61% after stress.

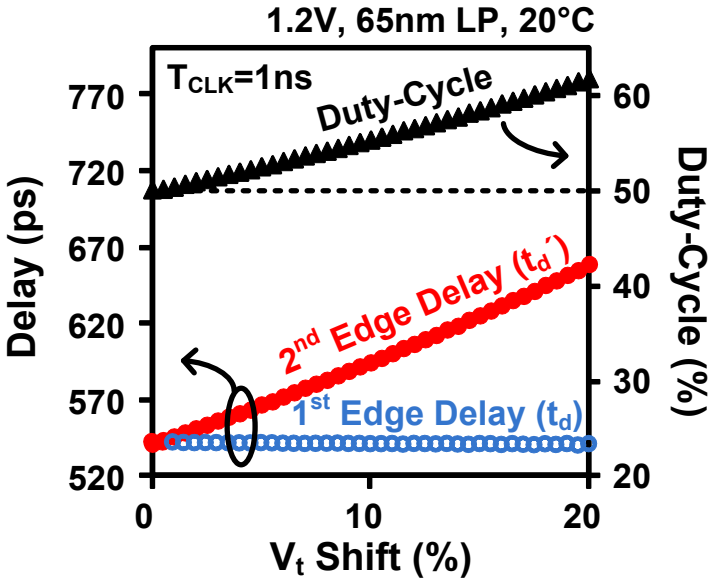


Figure 3.2: Simulated delay and duty-cycle shifts of a 540ns signal delay (=t_d) path driven by a 1GHz clock signal (=1/T_{CLK}). The duty-cycle shift is a function of the initial duty-cycle, t_d, T_{CLK}, and degradation of the 2nd signal edge delay.

The asymmetric BTI stress in SRAM signal and clock path affects the duty-cycle of various internal timing signals. This in turn negatively affects the read speed of low power SRAMs which typically rely on both the rising and falling edges of the input clock for generating timing signals [50]. When the input clock is gated off during long idle periods, DC BTI stress occurs in alternating devices along the clock path, as shown in Fig. 3.3. When the circuit switches from idle to active mode, as a result of the asymmetric BTI, we can see that phase '0' for address decoding, wordline driving, and bitline discharging becomes longer, while phase '1' for sense amplifier enable, bitline precharging, and data latching becomes shorter. This has two implications: (1) the clock-

to-dataout delay increases due to the delayed sense amplifier enable signal, and (2) the shorter precharge cycle leads to an incomplete bitline pre charge level which in turn increases the sensing time. This can be seen in the simulation results in Fig. 3.4.

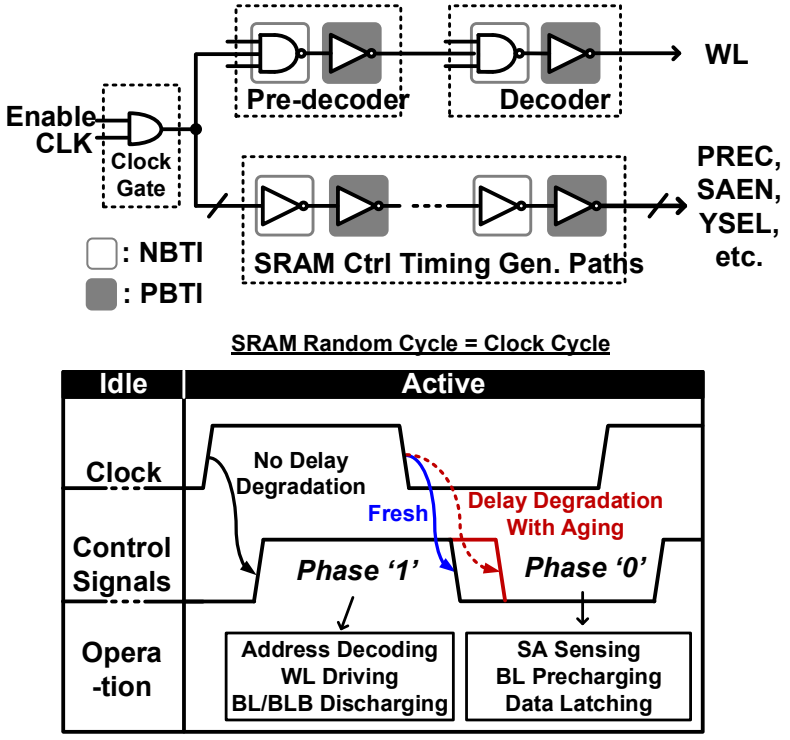
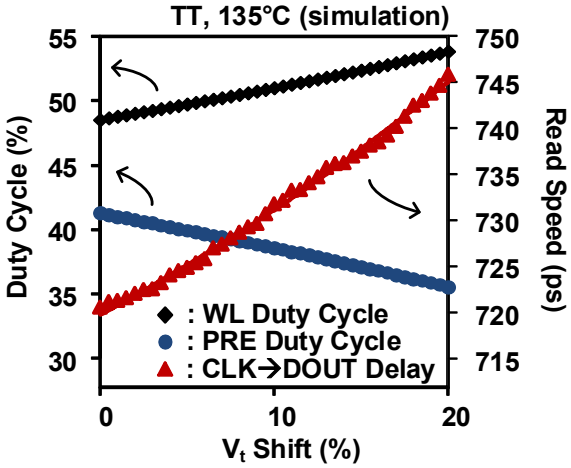


Figure 3.3: Asymmetric BTI occurs when a circuit path switches from idle to active mode. Circuits such as low power SRAM are particularly prone to asymmetric BTI aging as they utilize the second edge of the clock.

Even though this effect has drawn the attention of designers, none of the previous aging sensors [22, 35, 36, 41, 42, 48, 51-53] were able to verify it experimentally. In this work, we present a simple and practical duty-cycle characterization method based on the "silicon odometer" beat frequency detection (BFD) framework [35, 36]. Based on beat

frequency concept, we propose the first known on-chip reliability monitor to accurately characterize the impact on asymmetric BTI on SRAM read speed. By measuring the frequency difference between two SRAM arrays where the output signals are looped back to the input clock to form an oscillating circuit, we can achieve a picosecond order measurement precision with a measurement interrupt less than a microsecond.



Impact of Asymmetric BTI	Impact on Read Performance
Precharge phase ↓	Incomplete precharge
WL Enable Phase ↑	Increase WL→BL delay margin
CLK→SAEN Delay ↑	Read speed degration

Figure 3.4: SRAM read timing under asymmetric BTI aging. Duty cycle changes with stress.

3.2 Utilizing the Silicon Odometer Framework for Duty-Cycle Calculation

Our silicon odometer beat frequency detector which is capable of measuring stress-induced percentage change in the period of a Ring OSCillator (ROSC), can readily be used to estimate percentage change in duty-cycle of a clock driven by a chain of inverters under stress. In this section, we provide the mathematical derivation which shows how

the measurements from the odometer circuit can be used to calculate the duty cycle degradation.

Consider a ROSC with m inverter stages. During stress mode, the ROSC loop is open (NAND gate not shown in Fig. 3.5) and all inverters are exposed to a DC BTI stress. Since the NMOS and PMOS devices along the ROSC signal path are alternately stressed, the shift of the ROSC period can be expressed as:

$$\frac{\Delta T_{ROSC}}{T_{ROSC}} \approx \frac{\frac{m}{2}(\Delta t_{inv,PU}+\Delta t_{inv,PD})}{m(t_{inv,PU}+t_{inv,PD})} = \frac{1}{2}\cdot\frac{\Delta t_{inv,PU}+\Delta t_{inv,PD}}{t_{inv,PU}+t_{inv,PD}} \tag{3.1}$$

where the degradation in the inverter pull-up and pull-down delays are $\Delta t_{inv,PU}$ and $\Delta t_{inv,PD}$, respectively. Now consider an inverter chain with n number of stages that has undergone the same stress amount. The shift in the duty-cycle (D.C.) of the output signal becomes:

$$\Delta D.C. = \frac{t_d'-t_d}{T_{CLK}} = \frac{\Delta t_d}{t_d}\cdot\frac{t_d}{T_{CLK}} = \frac{\frac{n}{2}(\Delta t_{inv,PU}+\Delta t_{inv,PD})}{\frac{n}{2}(t_{inv,PU}+t_{inv,PD})}\cdot\frac{t_d}{T_{CLK}} \tag{3.2}$$

Here, t_d' and t_d are the total propagation delays of the n stage inverter chain before and after stress, respectively (see Fig. 3.1), and T_{CLK} is the period of the input clock signal to the inverter chain. Since t_d is equal to half the period of an unstressed ROSC with the same number of stages as the inverter chain, we can rewrite $(t_d'-t_d)/t_d$ as $\Delta T_{ROSC}/(T_{ROSC}/2)$ as assuming the fresh pull-up and pull-down delays are the same. Note that this quantity is independent of the number of stages m. This means that the duty-cycle degradation of the output clock can be expressed as:

$$\Delta D.C. = 2 \cdot \frac{\Delta T_{ROSC}}{T_{ROSC}} \cdot \frac{t_d}{T_{CLK}} \quad (3.3)$$

If we assume that the initial duty-cycle is 50%, the duty-cycle after stress can be described as:

$$D.C.' = 50\% + \frac{\Delta t_d}{t_d} \cdot \frac{t_d}{T_{CLK}} = 50\% + 2 \cdot \frac{\Delta T_{ROSC}}{T_{ROSC}} \cdot \frac{t_d}{T_{CLK}} \quad (3.4)$$

Therefore, by using the measured data $\Delta T_{ROSC}/T_{ROSC}$ from the odometer circuit and design specific parameters t_d and T_{CLK} obtained from circuit simulations, we can accurately calculate the duty-cycle shift of an arbitrary signal path.

Similar to the inverter chain example described above, we can compute the duty-cycle shift of a random logic path consisting of arbitrary gates (e.g. NAND, NOR, INV). The propagation delay of a random logic path can be expressed as:

$$t_d = \sum_i t_{i,PU} + \sum_j t_{j,PD}$$

where i and j denote the stages with a pull-up and pull-down transition, respectively. Next, we assume that the amount of delay degradation is a linear function of the threshold voltage shift and that the ratio between PBTI and NBTI is α. That is,

$$\Delta t_{inv,PD} = \beta \Delta V_{t,NBTI} \text{ , } \Delta V_{t,PBTI} = \alpha \Delta V_{t,NBTI}$$

Since the delay degradation depends on the type of gate [54] as well as the fanout (FO) [35], we introduce a sensitivity parameter γ to map the delay degradation of an arbitrary gate and arbitrary FO to that of an inverter with a known fanout of one. The degradation of the path delay can be now written as:

$$\Delta t_d = \sum_i \gamma_i \Delta t_{inv,Pull-up} + \sum_j \gamma_j \Delta t_{inv,Pull-down}$$

$$= \sum_i \gamma_i \alpha\beta \Delta V_{t,NBTI} + \sum_j \gamma_j \beta \Delta V_{t,NBTI} \quad (3.5)$$

Finally, the duty-cycle after stress can be expressed as:

$$D.C.' = 50\% + \left(\sum_i \gamma_i \alpha\beta + \sum_j \gamma_j \beta\right) \frac{\Delta T_{ROSC}}{T_{ROSC}} \cdot \frac{t_d}{T_{CLK}} \quad (3.6)$$

which can be easily calculated using the measured ROSC data (ΔT_{ROSC}, T_{ROSC}) and the various design specific parameters.

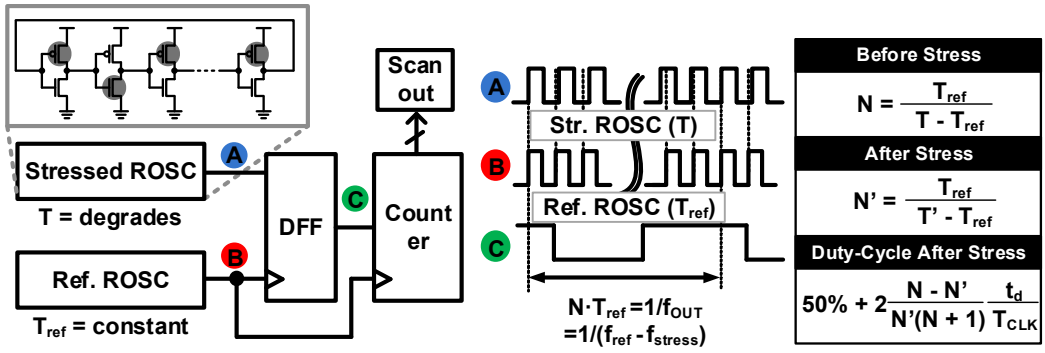


Figure 3.5: (left) Block diagram of silicon odometer beat frequency detection circuit. (right) Duty-cycle calculation formula based on the beat count before (N) and after (N′) the stress period.

The block diagram of the silicon odometer beat-frequency detection system is shown in Fig. 3.5 [35, 36]. A D-flip-flop is used to sense the frequency difference (=beat frequency) between a stressed ROSC and an identical fresh reference ROSC. A counter records the beat frequency by counting the number of reference ROSC period during one

period of beat output. The counts are then scanned out at different stress times for duty-cycle calculation.

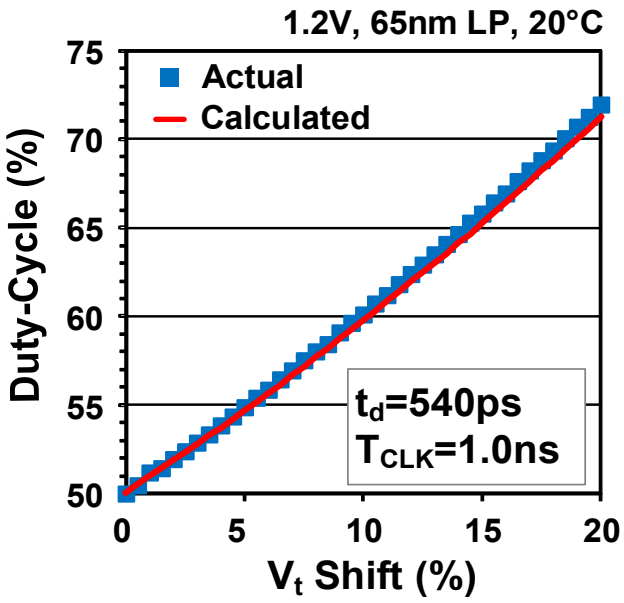


Figure 3.6: Proposed ROSC based duty-cycle estimation shows good agreement with actual duty-cycle shift.

Simulation results in Fig. 3.6 show an excellent match between the duty-cycle calculated based on silicon odometer data and the actual value. Duty-cycle shifts based on 65nm odometer test chips [53] under different stress conditions are plotted in Fig. 3.7. The amount of duty-cycle shift increases inversely with T_{CLK} and linearly with t_d as shown in Fig. 3.8 which was also predicted by the equation (3.6).

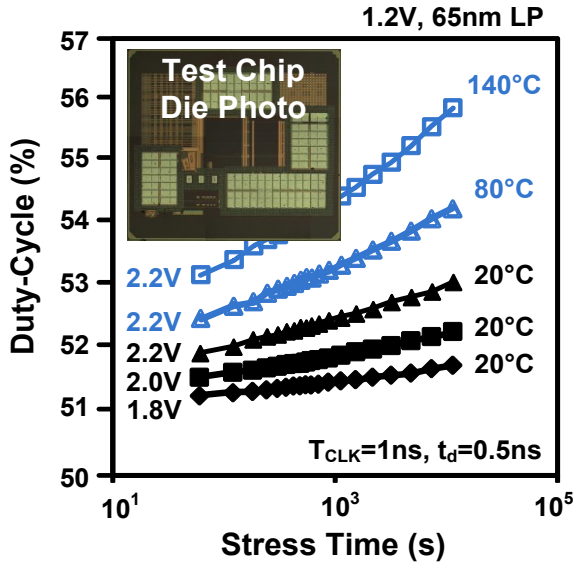


Figure 3.7: Duty-cycle shift based on 65nm odometer chip data under different stress voltages and temperature. Note that the y-scale of this log-log plot looks linear due to the limited dynamic range.

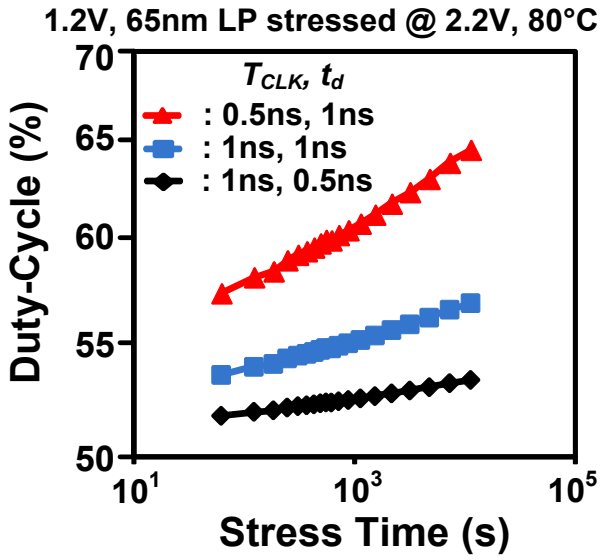


Figure 3.8: Duty-cycle vs. stress time for different clock frequencies and path delays.

3.3 SRAM Read Performance Degradation under Asymmetric NBTI and PBTI Stress

3.3.1 SRAM reliability test vehicle

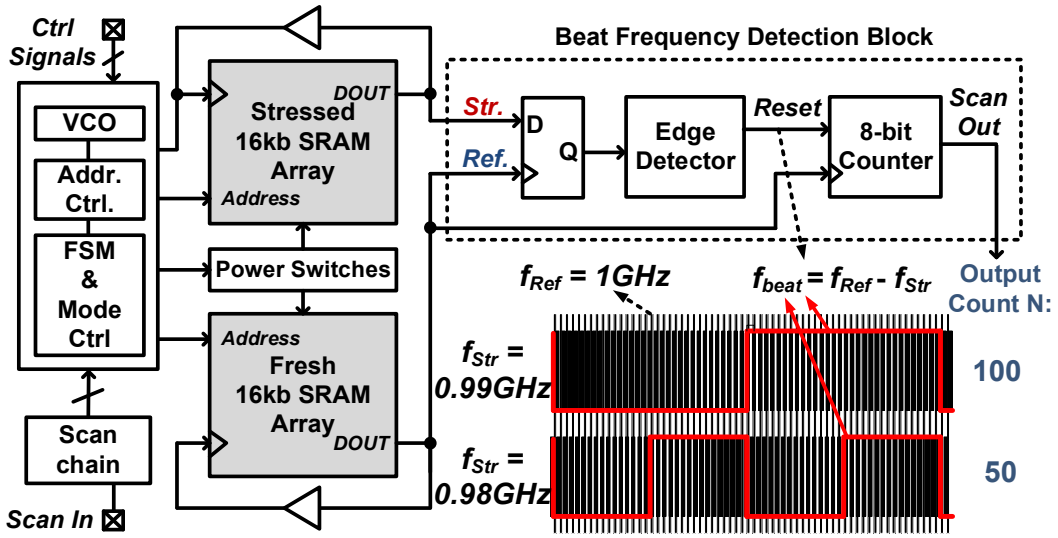


Figure 3.9: Top level schematic of the proposed SRAM test structure. The difference in read speed between a fresh and stressed SRAM array is measured using the beat frequency detection scheme which can achieve a frequency shift measurement resolution of 0.01% while minimizing the measurement time to µs. The output of the SRAM array is looped back to generate a oscillating frequency corresponding to the critical path delay.

The top level diagram of the proposed SRAM reliability test macro is shown in Fig. 3.9, which consists of two identical SRAM blocks, a beat frequency detection (BFD) system, global control circuits (scanchain, FSM, address control, etc.) and power gate switches.

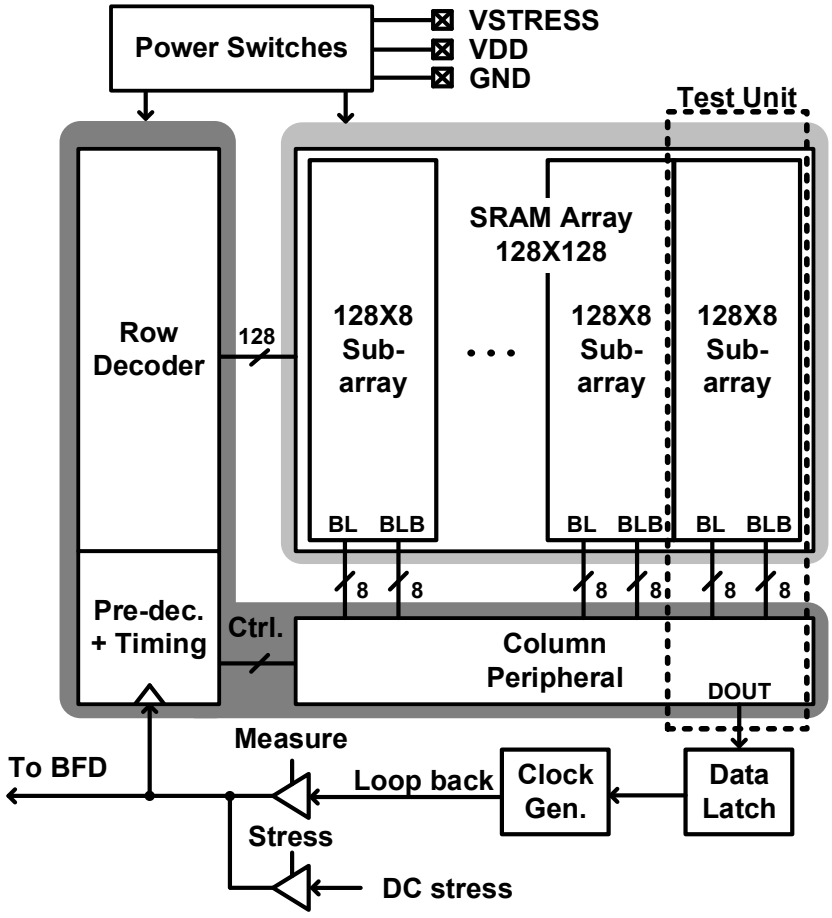


Figure 3.10: Schematic of the 128x128 SRAM test macro fabricated in a high-k metal-gate process. DOUT is looped back to trigger the clock input to generate an oscillating output during measurement mode.

During stress mode, a high supply voltage is applied to only one of the SRAM blocks by on-chip power switches, while the other SRAM is kept fresh by completely shutting off the power supply. In measurement mode, both SRAMs operate under a nominal VDD with their respective dataout signals looped back to generate two oscillating signals. The frequency difference between the output signals of the stressed and reference SRAM

arrays is then captured by the BFD block. The output count N in Fig. 3.9 represents the number of reference periods in one beat frequency period. The BTI induced frequency shift of the stress SRAM is amplified to achieve high precision (e.g. 1% frequency shift translated into a 50% count change for an initial frequency difference of 1%), while the measurement time is kept in the microsecond order to prevent unwanted BTI recovery. The differential sensing nature of the proposed circuit effectively rejects any common mode noise due to temperature or voltage drifts.

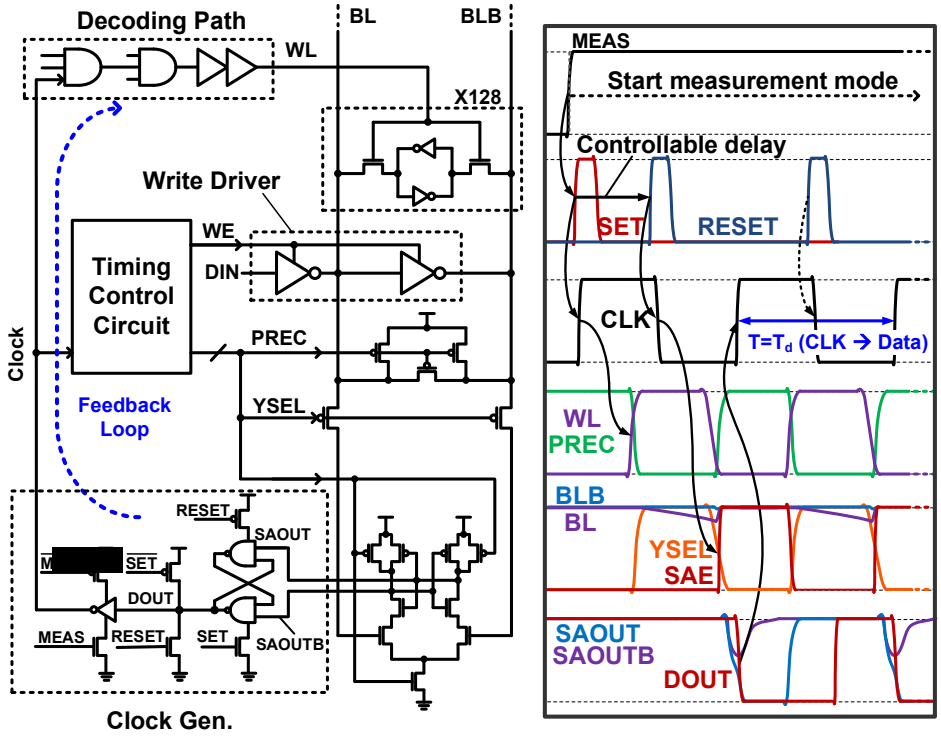
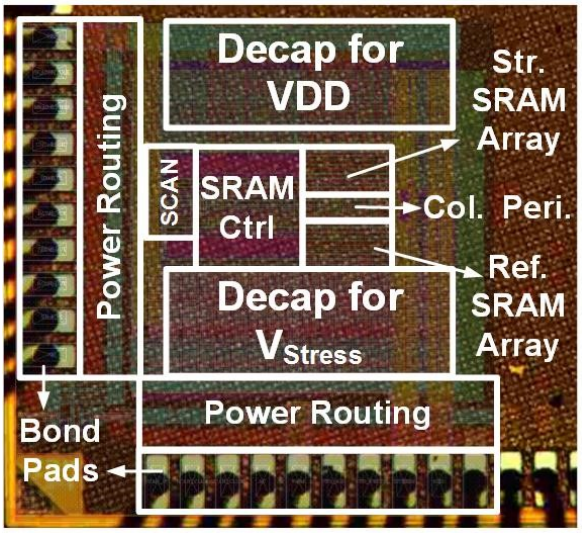


Figure 3.11: Detailed schematic and timing diagram of the SRAM read path with loop back configuration. After the first read cycle is complete, DOUT is fed back to trigger the next rising edge of clock. A reset signal is used to preset the clock for the next cycle.

Fig. 3.10 shows the schematic of the 16k bit SRAM macro comprising peripheral circuits and sixteen 128x8 sub-arrays. The clock is gated off to induce asymmetric BTI aging during stress mode. The detailed measurement sequence is described in the read path schematic and the internal control waveforms in Fig. 3.11. At the beginning of the measurement mode, the rising edge of the MEAS signal triggers the first SET signal, which generates a rising edge at the SRAM input clock. The clock is then switched back to 0 by an NMOS device controlled by the self-timed RESET signal. The tunable delay between SET and RESET pulses determines the pulse width of the clock. After the first read cycle is complete, the DOUT signal is used to trigger the next clock rising edge. As this process repeats, an oscillating signal is generated with a period corresponding to the SRAM critical path delay.



Process	HKMG SOI CMOS
VDD / IO Supplies	0.9V / 1.8V
Each SRAM Array Size	128X128
SRAM Cell Device Type	Floating-body device
Peripheral Device Type	Body-contact device
Area	500x455μm^2
Meas. Δf Resolution	0.01%~0.1%
Meas. Interrupt	~ 1μs

Figure 3.12: Die microphotograph and test chip feature summary.

3.3.2 Testchip measurement results

The proposed SRAM reliability macro was implemented in a high-k metal gate process. The die photo and feature summary table are given in Fig. 3.12. The entire 1k SRAM cells in a sub-array are initialized to '0' prior to applying stress, and subsequently intermittent measurements are taken from each cell between the long stress periods. The purpose of initializing the data to 0 is to ensure a failing edge at the DOUT signal to trigger the oscillation as can be seen from the waveform in Fig. 3.11. For a single measurement operation, the 8 row address bits and the 3 column address bits are controlled by an on-chip counter in order to scan through the whole array automatically. The actually measurement interrupt for each selected cell is around 1μs, then the system is switched to stress mode immediately, when the count information is slowly scanned out through an external data acquisition interface controlled by Labview^{TM}. The distributions of the fresh and stressed SRAM read frequency (f_{read}) under a 1.8V stress are plotted in Fig. 3.13. With the comparison of 0s, 27s and 1000s stress time, the average (μ) of the f_{read} distribution decreases with stress time while its standard deviation (σ) increases.

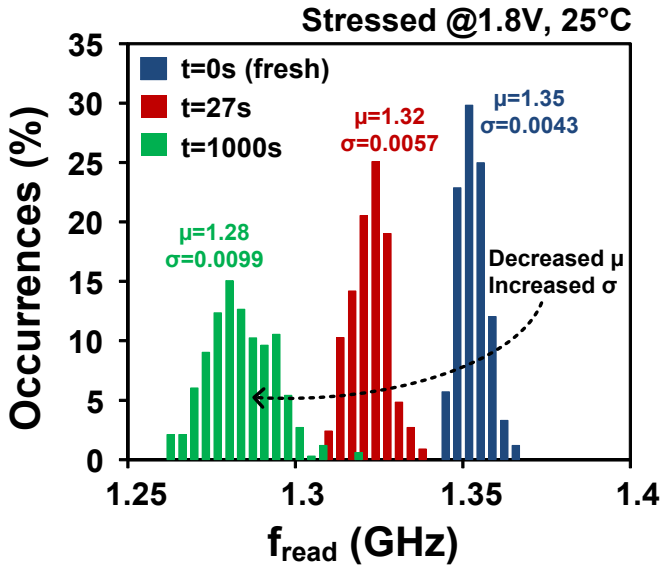


Figure 3.13: SRAM read frequency (fread) distribution at different stress times.

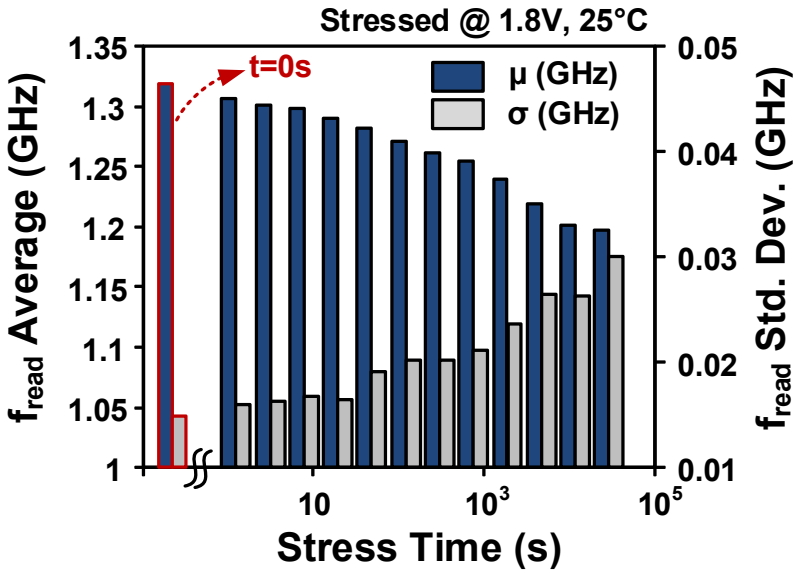


Figure 3.14: Average (μ) and standard deviation (σ) of fread versus stress time.

The μ and σ of f_{read} versus stress time is shown in Fig. 3.14. After being stressed under 1.8V, 25° for 40,000 seconds, μ of f_{read} drops from 1.33GHz to 1.2GHz, while σ of f_{read} increases from 0.015GHz to 0.030GHz. The BTI induced variation at the given stress condition and time is comparable to the process variation at time zero.

The distribution of f_{read} after a 1500s stress period under different stress voltages are compared in Fig. 3.15, showing that σ of f_{read} increases with higher stress voltage. It is expected as the larger number of defects induced by higher voltage BTI stress would increase the V_t fluctuation.

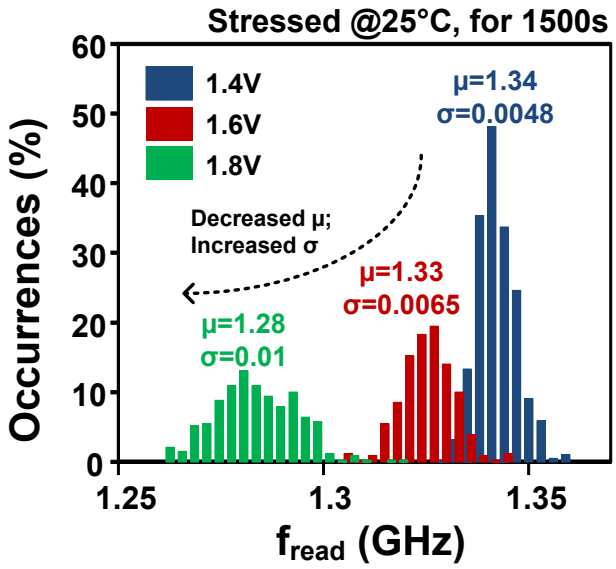


Figure 3.15: fread distribution after 1500 seconds of stress under different voltages.

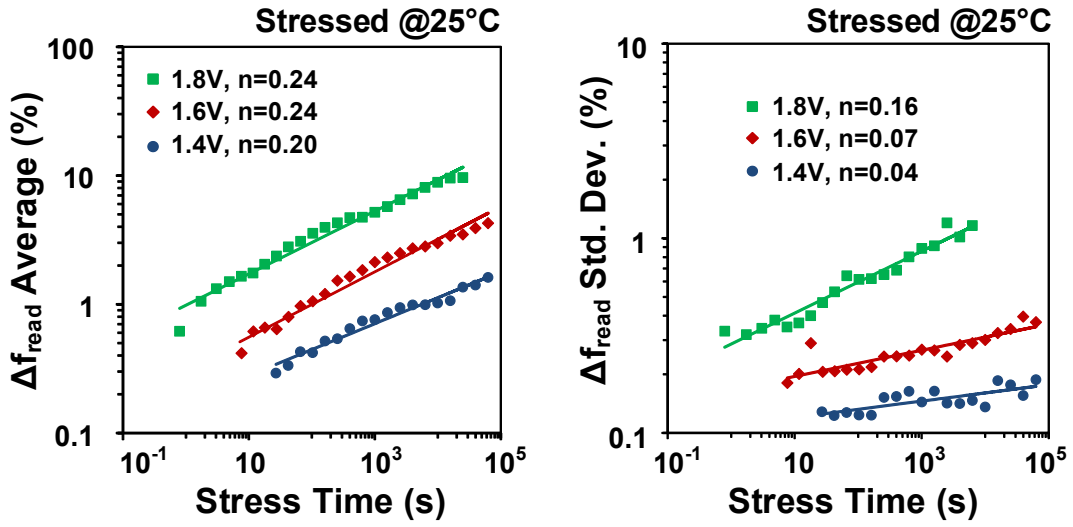


Figure 3.16: Both the average (μ) and standard deviation (σ) of Δf_{read} follow a power law relationship with stress time. The time exponent of σ(Δf_{read}) is smaller than that of μ(Δf_{read}).

The μ and σ of the read frequency degradation (Δf_{read}) versus stress time are shown in Fig. 3.16, where both μ and σ follow a power law dependence (t^n). Note that the degradation is caused by an overall effect of both PBTI and NBTI alone the various delay paths. A clean power law behavior for the frequency shift might indicate both PBTI and NBTI have the similar dependence over stress time, or one of the two effects is dominating. The time exponent (n) of σ is less than half that of μ, which agrees with modeling results based on discrete random charge fluctuation [55] as well as the previous measurement data from a ring oscillator array [53].

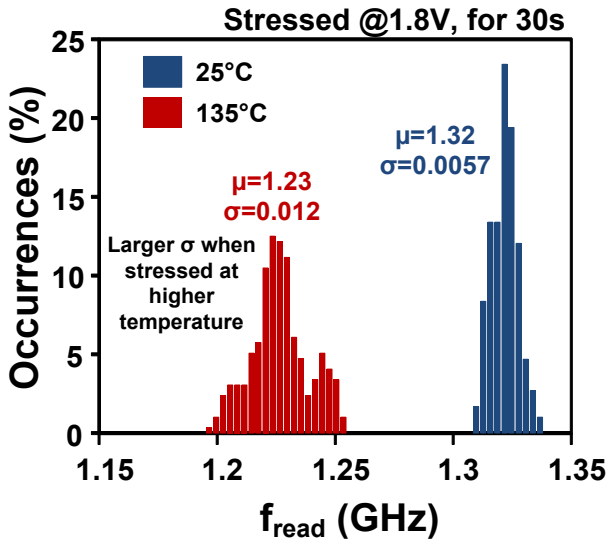


Figure 3.17: fread distribution after 30 seconds of stress under different temperatures.

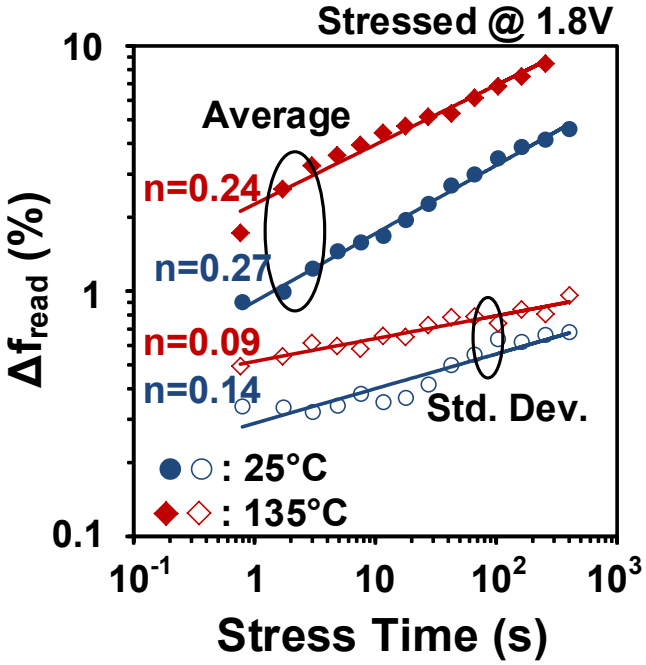


Figure 3.18: μ and σ of Δfread under different temperatures.

BTI at higher temperature induces more random defects, which in turn increases the σ of f_{read} as shown in Fig. 3.17. The magnitude of both μ and σ of Δf_{read} at 135°C are more than twice of those at 25°C, as observed in Fig. 3.18. The n value of Δf_{read} has a wide distribution which is weakly dependent on voltage and temperature as shown in Fig. 3.19. The magnitude of n value is larger than the previous observation, and no trap saturation induced slope reduction is observed at the given high temperature and high voltage stress conditions.

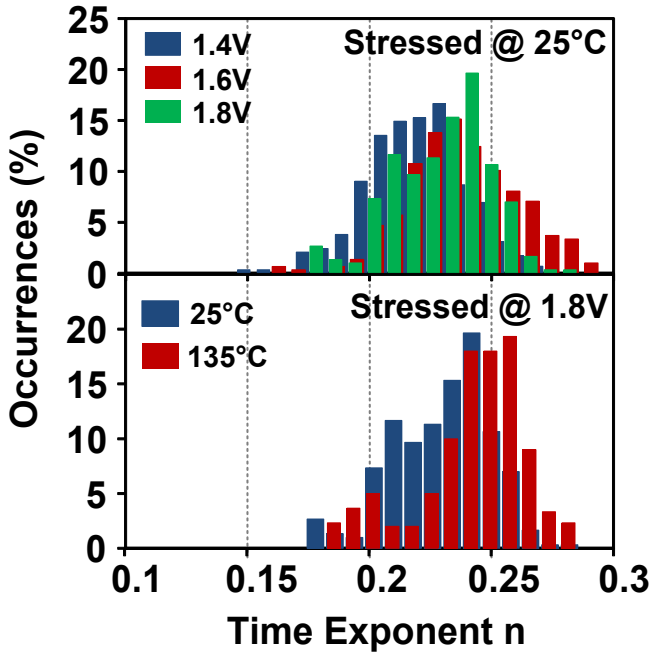


Figure 3.19: The distribution of time exponent n under different stress voltages and temperatures.

At last, the impact of asymmetric BTI on read failure was tested. The read errors at time zero are caused by sensing failure due to the very short wordline duty cycle. The

short duty cycle is achieved by tuning the clock pulse width which control the '0' phase demonstrated in Fig. 3.3, while keeping the '1' phase unchanged. The asymmetric BTI relaxes the wordline pulse width which lowers the read failure rate with time, but the average read frequency is reduced, as shown in Fig. 3.20.

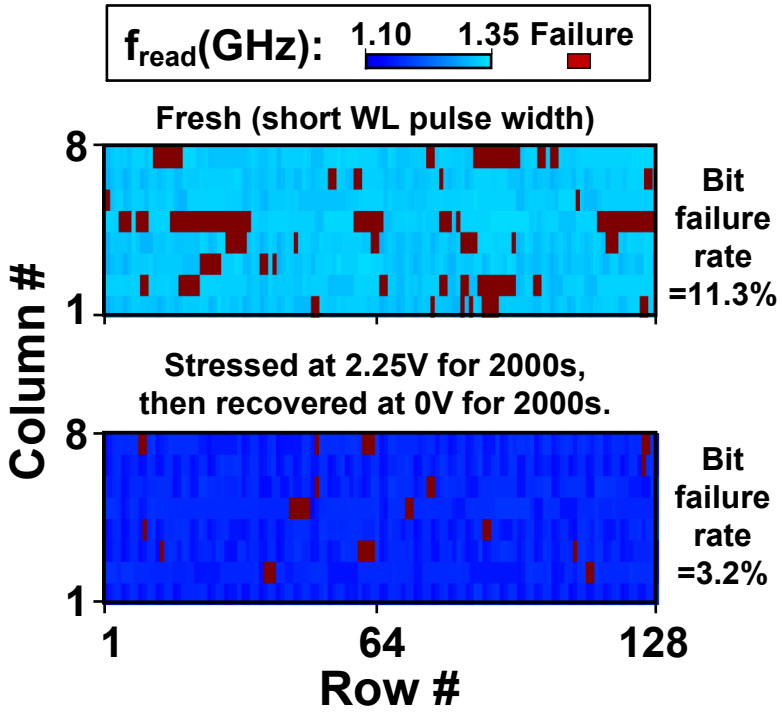


Figure 3.20: Spatial distribution of fread before and after DC stress.

3.4 Conclusions

DC BTI induced duty-cycle shift affects the performance of circuits relying on both of the clock rising and falling edges, particularly in various low-power logic and memory with clock gating techniques. This duty-cycle shift is caused by the BTI induced aging

alternately occurring to pull-up and pull-down networks on the consecutive stages along a logic path during clock gating mode, therefore the rising and falling edges are undergone different delays. In this work, we first present a simple and practical method to accurately measure the duty-cycle shift of a delay path based on the ring oscillator based beat-frequency detection framework. A SRAM reliability detection vehicle was then proposed to detect the asymmetric aging impact on the read performance in an actual SRAM.

Up to 6% of duty-cycle shift is observed from a delay path stressed at 2.2V, 140°C for 3hrs. Based on the statistical test results from SRAM reliability macro, we observed that the variation of read frequency increases with the stress intensity. Both of the average and the standard deviation of read frequency shift follow power law dependence with the stress time. The asymmetric aging increases the wordline duty-cycle, which reduce the sensing error caused by the short wordline duty-cycle.

Chapter 4.

Fast Characterization of PBTI and NBTI Induced Frequency Shifts under a Realistic Recovery Bias Using a Ring Oscillator Based Circuit

4.1 Introduction

Positive Bias Temperature Instability (PBTI) has become equally significant compared to Negative Bias Temperature Instability (NBTI) in high-k metal gate technologies. Due to the difference in their physical origin, the magnitude and behavior of PBTI and NBTI can be drastically different. Hence, there has been a growing need to develop test structures that are capable of separately characterizing the circuit level impacts from these two aging mechanisms. Ring oscillator (ROSC) is widely used to characterizing the device level impact on logic delay degradation due to its simplicity. However, the frequency degradation of conventional ROSC structure depends on both PBTI and NBTI.

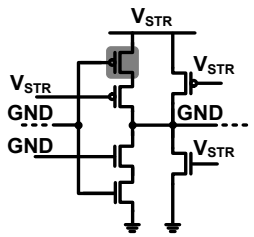
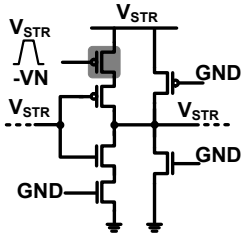
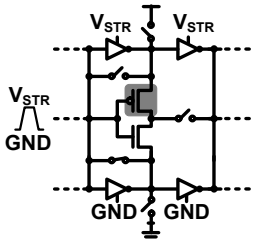
	[1]	[2]	This work
ROSC Stage Diagram (NBTI Stress Mode Shown)			
Stress Capability	DC No AC data	DC Unrealistic AC (i.e. V_{ds}=0)	DC Realistic AC (i.e. V_{ds}=~V_{STR})
Meas. Scheme	Simple counter	Simple counter	Beat frequency scheme with phase alignment
*Meas. Time for 0.01% Resolution	10,000 ROSC periods	10,000 ROSC periods	100 ROSC periods (<1μs)
Use Condition Test Data	No (poor resolution)	No (poor resolution)	Yes (high resolution)
Result Credibility	Stressed devices do not have switching input during transition	Stressed devices do not have switching input during transition	Delay caused by additional unstressed switches can be simply calibrated out

Figure 4.1: Comparison of ROSC based circuits for separately monitoring NBTI and PBTI induced frequency shifts.

To separately characterize each of the effect, a modified ROSC design proposed in which uses a keeper circuit to isolate the stress in each stage (Fig. 4.1) [56]. In this implementation, AC stress can be applied by toggling the VDD and ground signals but this limits the stress frequency to a few MHz. This limitation was addressed in subsequent design which applies the AC stress to the header or footer [43]. However, there are three major drawbacks in both previous works. First and foremost, neither design can achieve the correct recovery bias in Fig. 4.2 where the source-to-gate voltage is zero and the source-to-drain voltage is high. Providing a recovery bias that is closer to reality is important since a larger V_{ds} has proven to enhance recovery [57]. Secondly, a simple counter based scheme results in significant unwanted recovery due to the long

measurement times. Lastly, the additional switches for providing stress bias to the Devices Under Test (DUT) affect the circuit delay corrupting the aging measurement data. In this work, we address these issues by proposing a PBTI/NBTI odometer which supports a realistic recovery bias, sub-µs measurement time, sub-ps resolution, and eliminates the impact of the additional switches on the stress data.

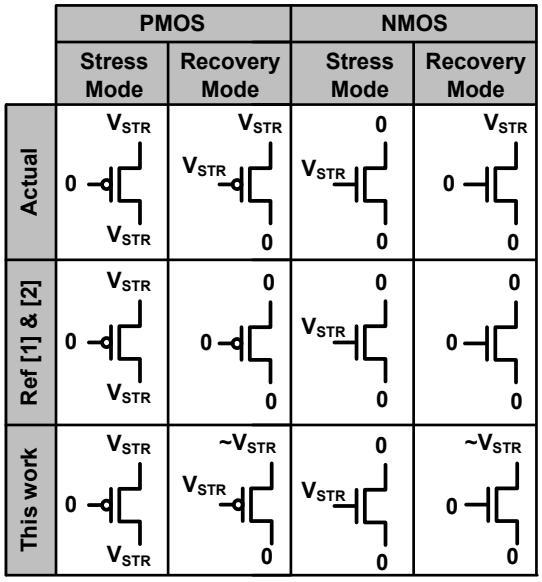


Figure 4.2: PMOS and NMOS bias conditions in stress and recovery modes. The previous designs apply zero bias between the CMOS terminal while the realistic condition is the absolute source-to-drain voltage is high.

4.2 ROSC Based Monitor with Beat Frequency Detection System

In order to separately stress the NMOS and PMOS transistors, we utilize two sets of tri-state drivers and switches between the gate and source of PMOS and NMOS in each

inverter stage as shown in Fig. 4.3. For NBTI stress, a V_{gs} bias is applied to all PMOS devices while the gate and source voltages of NMOS devices are shorted. By doing so, all NMOS devices are kept fresh while the PMOS devices undergo AC stress. Note that we allow a V_t drop to occur in the V_{ds} bias during recovery mode since otherwise the circuit structure had to become overly complicated. In measurement mode, degradation in the additional drivers and switches will not affect the main path delay as those devices are switched off. At the same time, the header and footer for each inverter stage are switched on and the feedback loop is closed to make the ROSC oscillate. The additional delay caused by the switches and pass gates does not depend on the device aging and can be calibrated out using the to-be-introduced methodology shown in Fig. 4.18. Two identical ROSCs are implemented side-by-side, with only one ROSC is stressed, but the other one is kept fresh by using on-chip power gates during stress mode. During measurement mode, the output frequency from both ROSC are feed into the detection block, where the beat frequency detection scheme [35, 36] is adopted for fast characterization of BTI. As shown in the block diagram (Fig. 4.4), a flip-flop measures the frequency difference between the stressed and reference ROSCs. The final output count corresponds to the number of reference ROSC periods constituting a single beat signal period. The beat frequency detection scheme is particularly sensitive when the two ROSC frequencies are closer to each other. For an initial frequency difference of 1%, a 1% frequency shift results in a count change from 100 to 50, as illustrated in Fig. 4.4. Our design shows that

the measurement interrupt can be reduced to sub-µs while achieving a frequency measurement resolution of 0.01%.

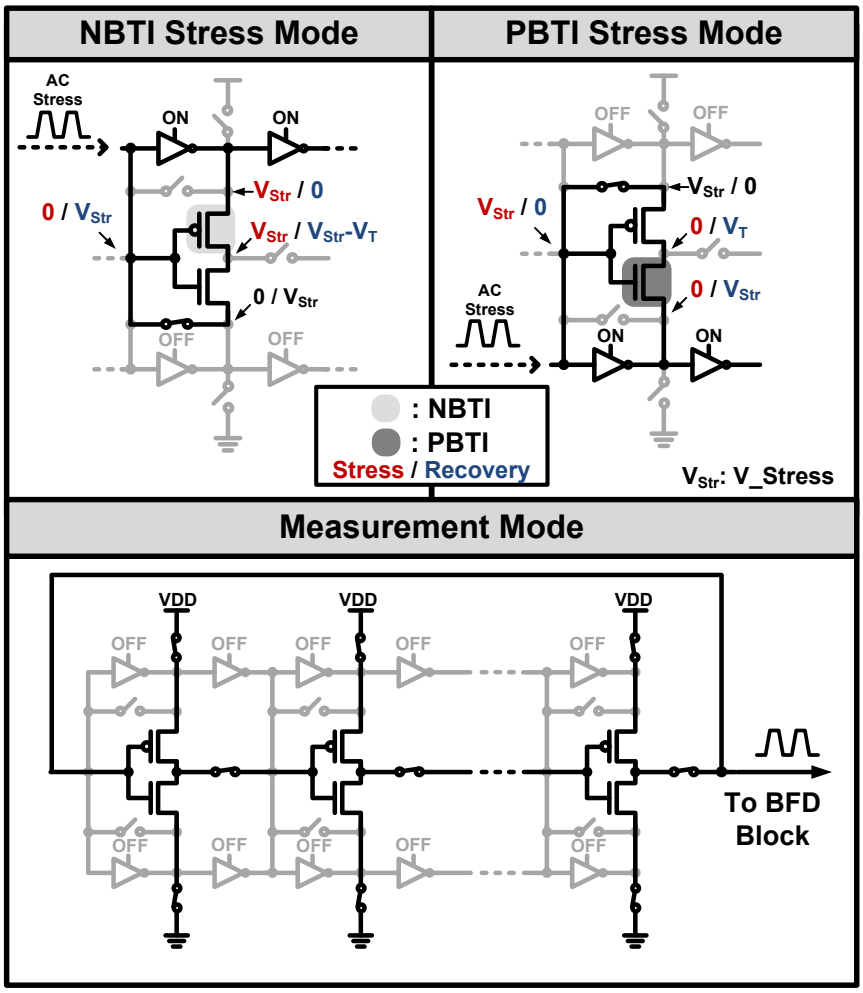


Figure 4.3: Operation modes of the proposed PBTI/NBTI odometer. Additional drivers and pass gates are used to apply realistic AC stress biases. Their delays can be calibrated out as shown in Fig. 4.18. Degradation in the supporting drivers does not affect the data since those circuits are switched off during measurement mode.

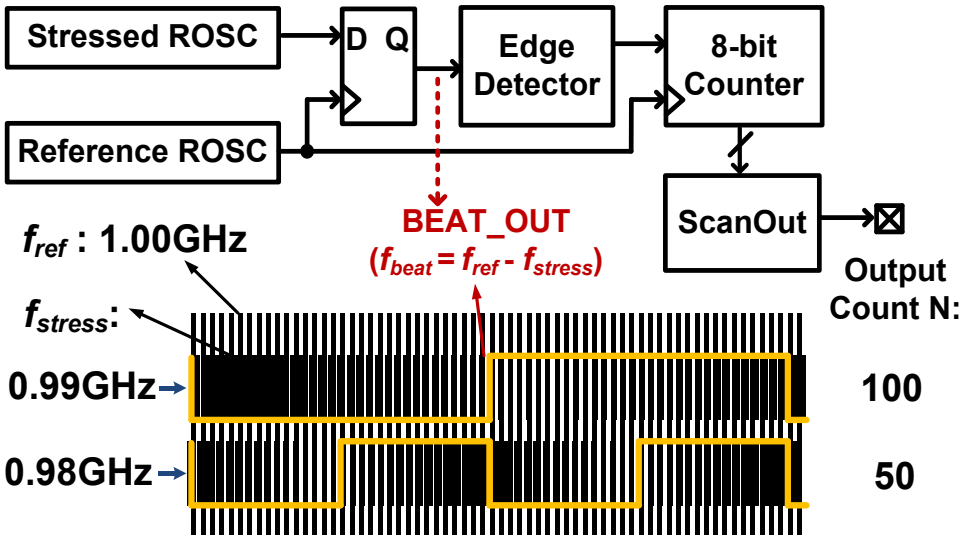


Figure 4.4: Principle of beat frequency detection technique adopted in this work. By measuring the beat frequency rather than the raw frequency, the measurement resolution can be greatly enhanced (e.g.<1ps) while minimizing the measurement time (e.g.< 1μs).

Since the phase difference between the stressed and reference ROSCs is unknown at the moment when they are switched to measurement mode, the first beat frequency result could be incorrect. So the first count has to be dropped and the measurement time is wasted. To further reduce the measurement interrupt, phase alignment technique was used at the initialization operation of the ROSCs. As shown in Fig. 4.5, a simple timing scheme was deployed at the stress-to-measurement mode transition. The STR_EN signal puts both of the stressed and the reference ROSC input to '0', regardless the previous state. After a short period delay to make sure all the internal signals are stable for each ROSC stage, the tri-state inverter at both ROSC input is turned off, and the loop is closed

at the same time, controlled by RO_EN and LOOP signal, respectively. The frequency output from reference ROSC are tapped out from the third stage for an intentional phase lag compared to the stressed ROSC where the signal is tapped out at the first stage. The rising edge of the faster reference ROSC signal can quickly catch up with the stressed ROSC rising edge, and when the two rising edge overlaps, a correct trigger signal is generated to start the beat counter. By doing so, the first count out of the beat frequency detection block is guaranteed to be correct, which reduced the measurement interrupt to 400ns.

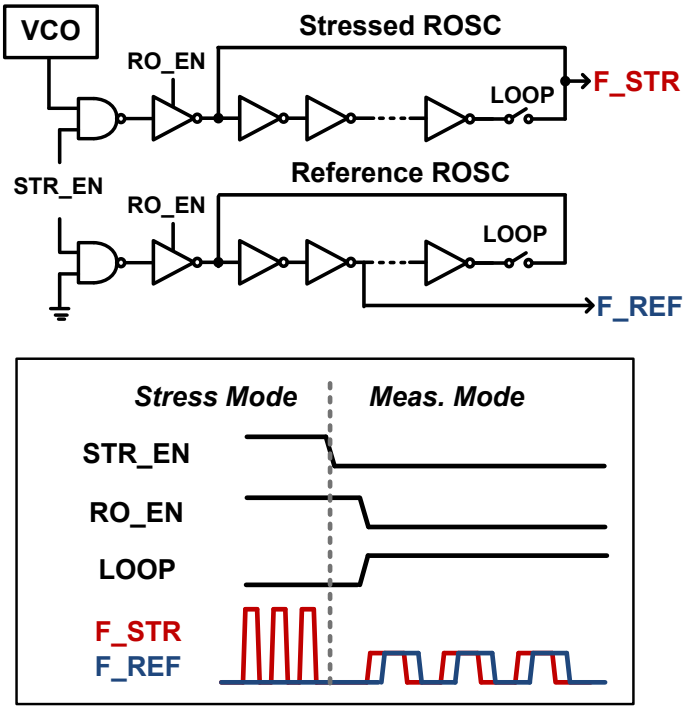
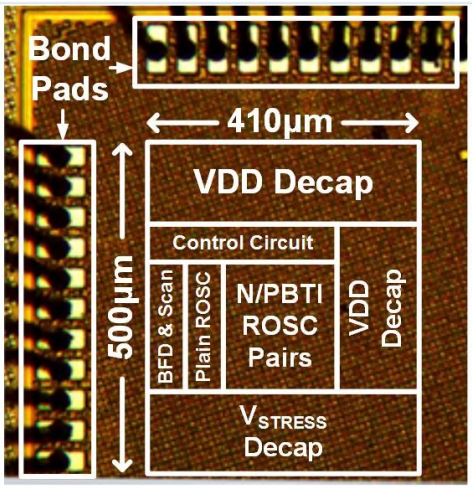


Figure 4.5: Phase alignment technique to reduce measurement time. The intentional phase lag of the reference signal makes the first count from the beat frequency detection block valid.

4.3 PBTI and NBTI Measurement Results



Process	HKMG CMOS
Core / IO Supplies	0.9V / 1.8V
Stress Voltage	0.9V, 1.4V, 1.8V, 2.2V
Measuremt Voltage	0.9V
Area	500x410µm^2
Δf Resolution	>0.01%
Measuremt Time	>400ns
DUT Type	Core Devices

Figure 4.6: Testchip die photo and chip feature summary. The chip was implemented in a high-k metal-gate process.

4.3.1 NBTI and PBTI under DC stress

A test chip was implemented in a high-k metal-gate process as the die microscopic photo and the specifications shown in Fig. 4.6. We firstly conducted the DC stress upon PBTI and NBTI with different stress voltages. Both PBTI and NBTI induced frequency shifts under DC stress closely follow a power law dependency as shown in Fig. 4.7. We observed that the magnitude of PBTI is 5X to 10X larger than that of NBTI at any given stress time for this process. Our measurement set up allows up to perform short time stress with the interval as low as 20µs, by which short time PBTI was measured at different voltage levels as shown in Fig. 4.8. Note that even down to tens of micro second

range, the PBTI still keeps its linearity in log-log scale. However, the time exponents are different from the long-term case shown in Fig. 4.7.

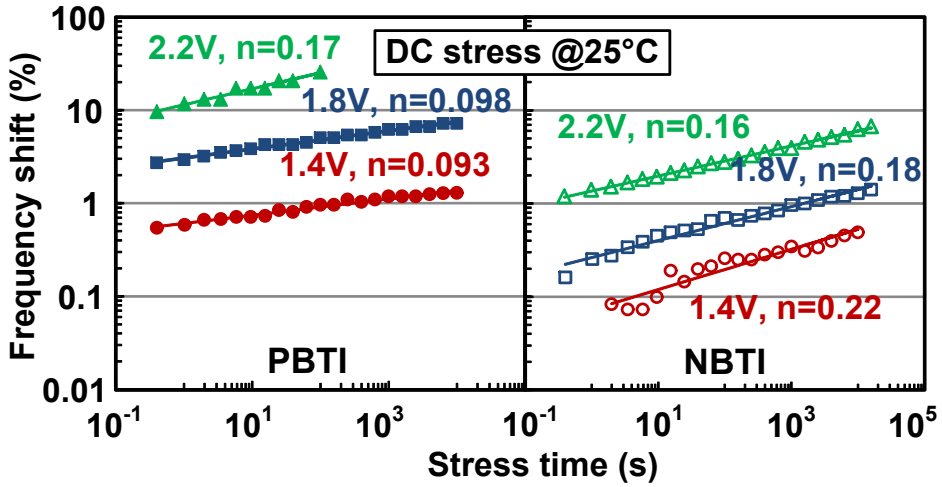


Figure 4.7: Measured PBTI and NBTI induced frequency shift under different DC stress voltages.

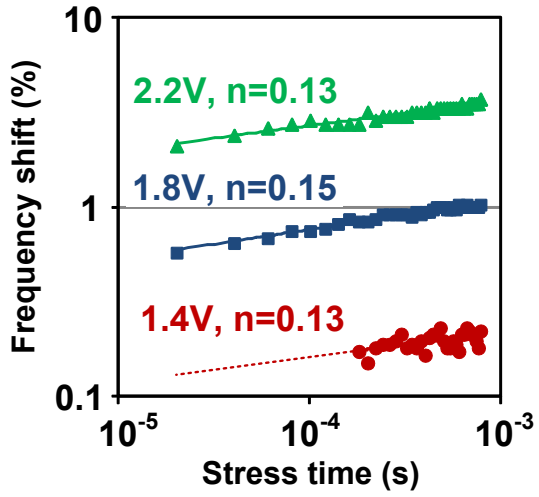


Figure 4.8: Short time stress measurement results of DC PBTI induced frequency degradation

The long term time exponent value extracted from the data in Fig. 4.7 for PBTI ranges from 0.09 to 0.12, which is slightly lower than the values reported in [15, 43, 58]. This can be attributed to the short measurement time of the proposed monitor circuit which eliminates the unwanted recovery in the early stress sample points resulting in a lower overall n value [35, 59].

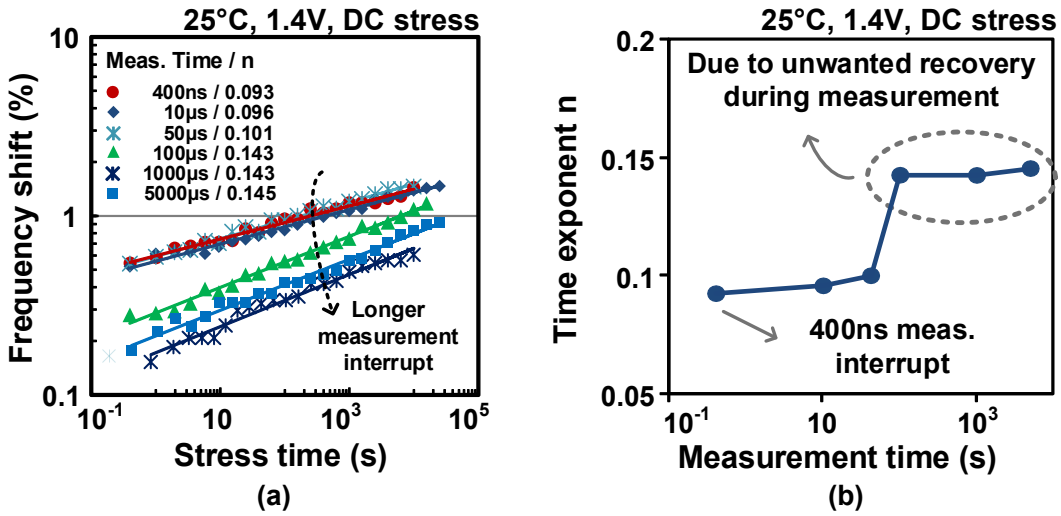


Figure 4.9: (a) PBTI induced frequency shift vs. stress time measured using different measurement times. (b) Time exponent n vs. measurement time.

In order to characterize the impact of the measurement interrupt on the overall frequency degradation to further verify the aforementioned point, we conducted the same stress test with controllable measurement interrupt interval as shown in Fig. 4.9. With longer measurement time, the magnitude of the frequency degradation is damped due to the recovery effect during the measurement, especially for the measurement time larger

than 50μs. The time exponent n increases with increasing measurement time with the 400ns measurement time returning the lower n value. A dramatic change in n value was observed between 50μs and 100μs, while the shift in other time region is mild, which might indicate the time constant for the dominant recovery mechanism is around that range. It is observed that the time slope for PBTI increases with increasing stress voltage, while the opposite trend was seen for the NBTI case.

4.3.2 AC stress with realistic recovery bias and temperature dependence

Repetitive stress-recovery sequences under two different recovery modes are compared in Fig. 4.10. Here, the realistic recovery is when a transistor is in an off state as given in Fig. 4.2, while the power down mode is when the supply is completely shut down. The realistic recovery bias gives a significantly stronger recovery rate compared to the power down case (86% vs. 51% after the third cycle), which can be attributed to the accelerated recovery induced by the larger V_{ds} [14, 57, 58]. With the negative bias applied at the drain side of the gate for NMOS (positive for PMOS), the potential barrier from the defect band to the doped silicon conduction band is lower, which result in an accelerated de-trapping process compared to zero bias case, as illustrated in the band diagram in Fig. 4.11. The stronger recovery under a realistic AC stress can also be seen in the DC to realistic AC stress frequency shift ratio in Fig. 4.12. The measured ratio for PBTI and NBTI were 3.2X and 18X, respectively. PBTI shows a smaller DC to AC ratio as compared to NBTI which is consistent with previous reports [15, 59]. Frequency shift measurements under a 200 MHz AC stress for different stress voltages are shown in Fig.

4.13, indicating that the PBTI time slope has a stronger voltage dependence compared to the DC stress case.

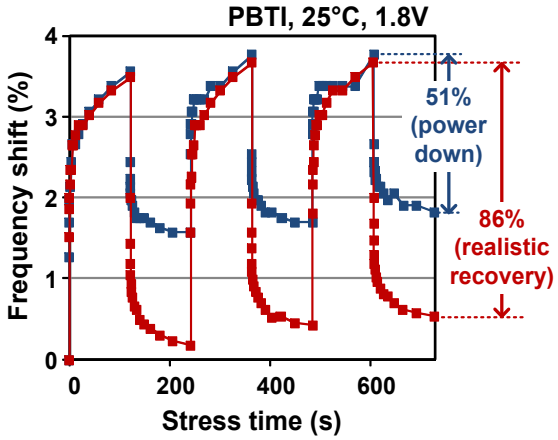


Figure 4.10: Periodic stress/recovery measurement results for power down and realistic recovery.

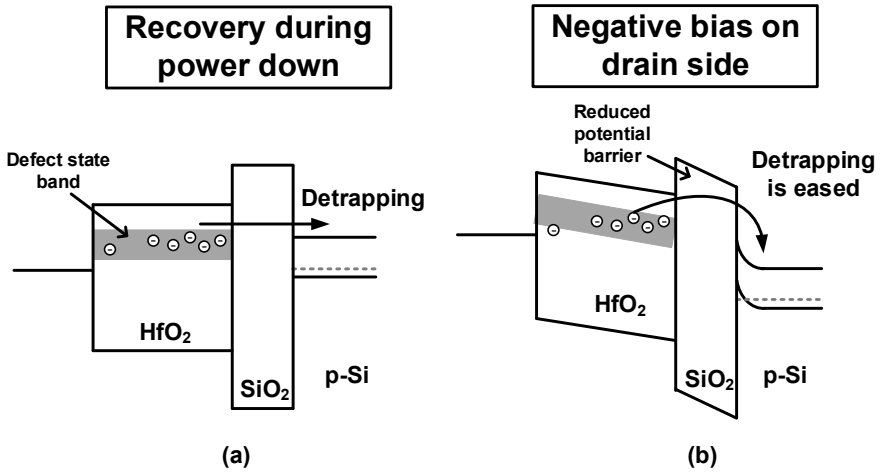


Figure 4.11: Band diagram for n-channel HKMG MOS at the drain side edge for (a) power down mode recovery with zero bias; (b) realistic recovery bias during OFF-

state, where drain induced band bending accelerates the detrapping effect through interface layer.

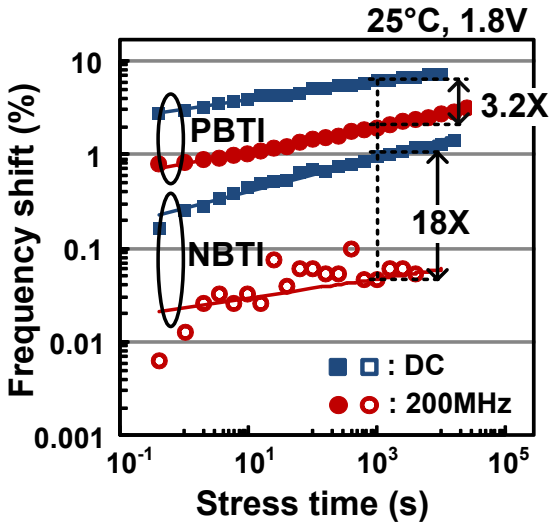


Figure 4.12: Comparison between DC and AC stress data.

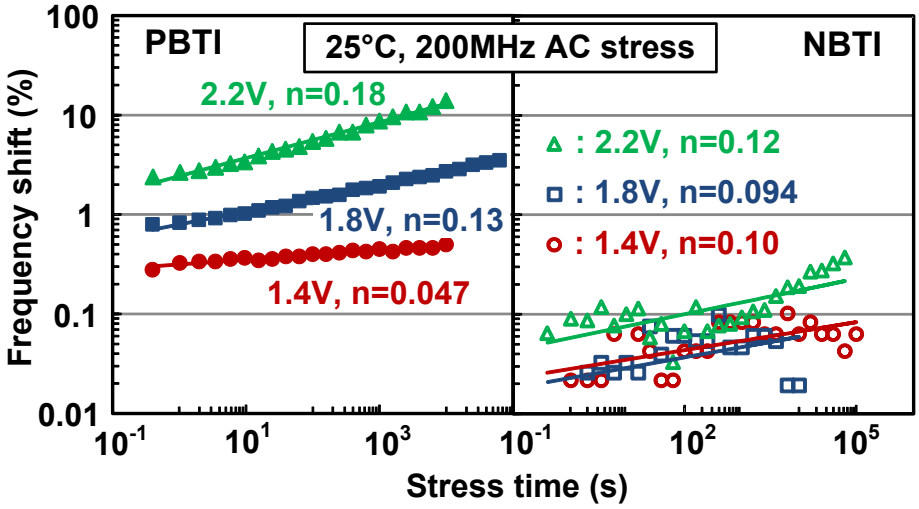


Figure 4.13: Measured PBTI and NBTI induced frequency shift under a 200MHz AC stress at different voltages.

To compare the impact of temperature on NBTI and PBTI, we measured frequency shifts at 25°C and 110°C, as plotted in Fig. 4.14, from which we can see that NBTI is more sensitive to temperature compared to PBTI. At 110°C, which is the typical processor operating temperature for logic circuits, the temperature acceleration effect for AC BTI is more severe compared to the DC case. In particular, the power law exponent increases with temperature in AC stress case for both PBTI and NBTI, suggesting the temperature dependency of time slope is more significant for short term stress and recovery mechanisms.

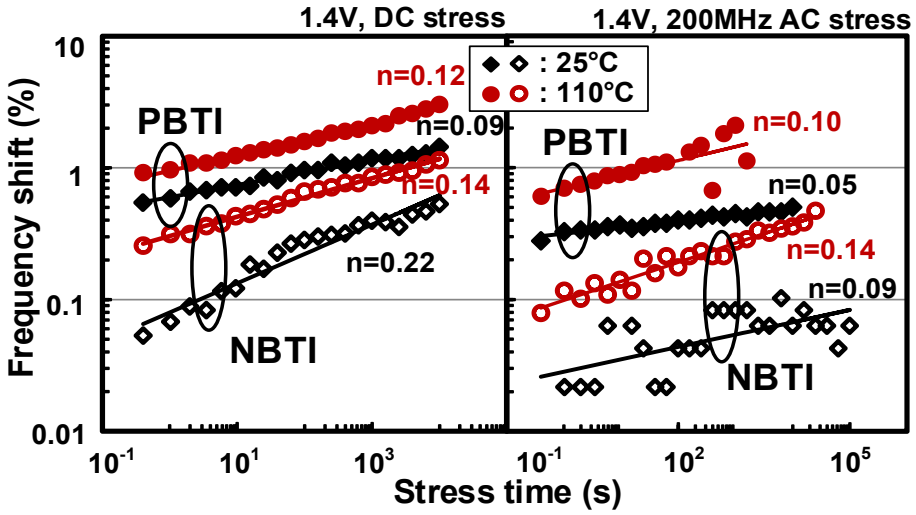


Figure 4.14: DC and AC stress data at 25°C and 110°C. The circuit was stressed at 1.4V and measured at 0.9V.

4.3.3 Long term recovery measurement

The frequency shift relaxation during long term recovery in power off mode roughly follows a -log(t) curve for both PBTI and NBTI for the given recovery conditions in Fig.

4.15 and Fig. 4.16. Both the magnitude and time slope for recovery increase with stress voltage, which is consistent with data presented in [14, 58]. Fig. 4.16 shows that recovery has a weak dependence on temperature which is in line with previous observations [58, 60].

The comparison between power off mode and the realistic bias mode was conducted for long-term PBTI recovery as shown in Fig. 4.17. Both recovery magnitude and the time slope are significantly accelerated in the realistic recovery situation, especially for the higher voltage. The ratio between power off mode and the realistic recovery mode is greater than 2 with recovery time larger than 1k seconds for 1.8V stress and recovery voltage.

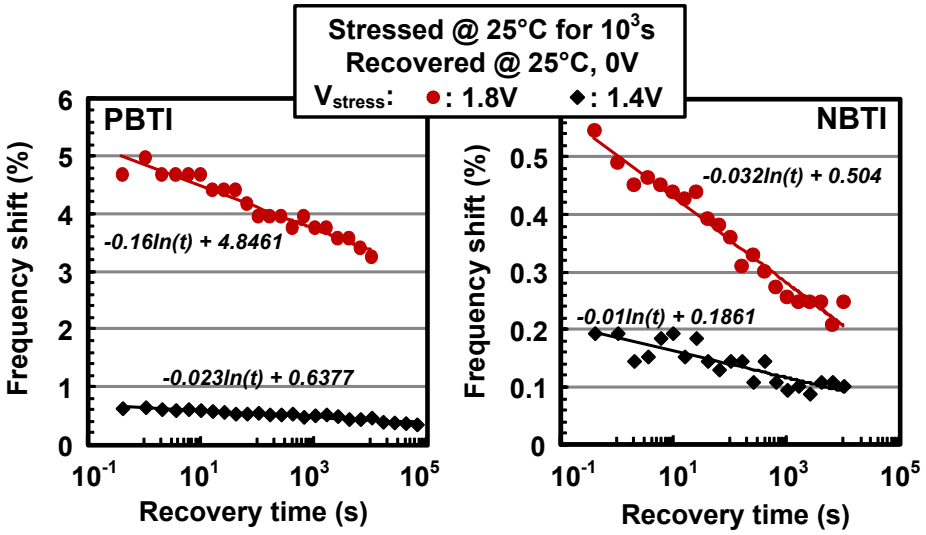


Figure 4.15: Frequency shift for PBTI and NBTI after the stress voltage is shut down. The recovery follows a log(t) dependence. The time slope depends on the stress voltage.

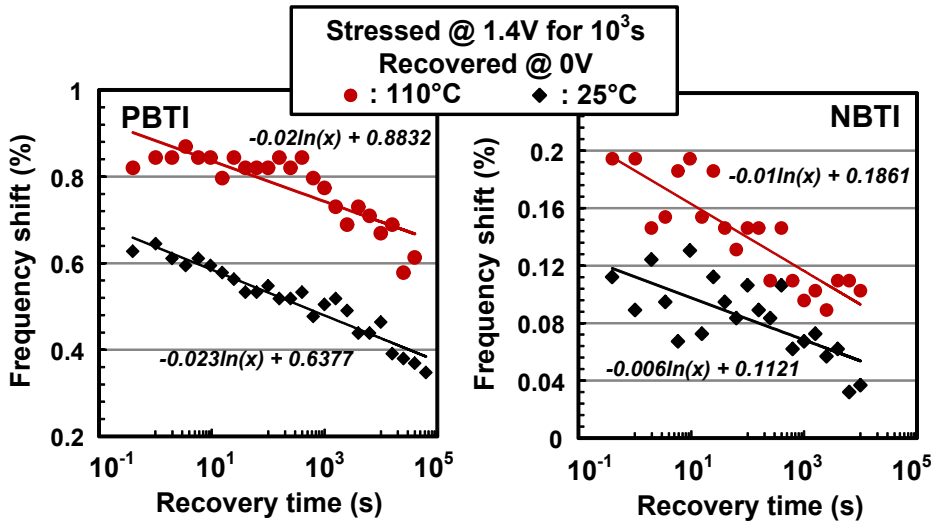


Figure 4.16: Frequency shift recovery data similar to Fig. 4.15 but for two different temperatures. Recovery time slope shows a weak dependence on temperature.

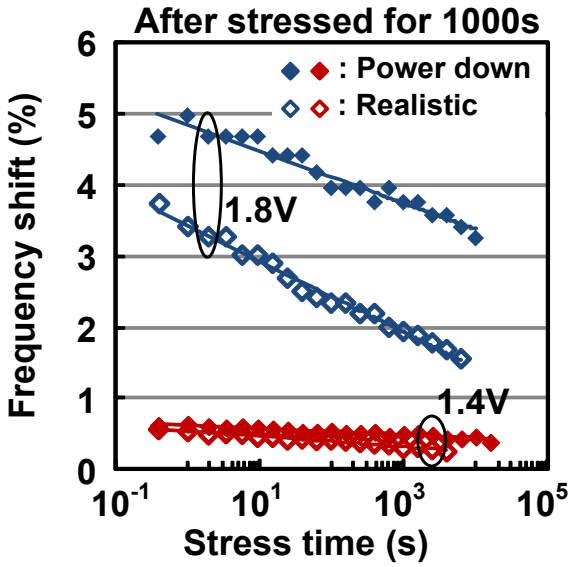


Figure 4.17: Long term recovery with power down mode and realistic recovery mode are compared for PBTI with two different stress voltages.

4.4 Calibration Method to Calculate Plain ROSC Degradation

In the proposed inverter design, additional switches and the pass gates are used to separate the impact of PBTI and NBTI. However, the extra R and C induced by these circuits make the measured frequency degradation result different with that from a Plain ROSC. It is necessary to map the data from the modified ROSC to a plain ROSC with a clean delay path so that the values more representing to standard logic circuit. Since the additional devices are un-stressed, their effect on ROSC delay is independent of aging, which makes it possible to use one point calibration method. This can be simply proved through the first order Elmore delay calculation.

The schematic of the modified inverter stage used in this design and the plain inverter stage is compared in Fig. 4.18. The propagation delay of a standard inverter stage can be expressed as:

$$t_d \propto (R_p + R_n)C_L \tag{4.1}$$

Where R_p and R_n are equivalent resistance of PMOS and NMOS during transition, C_L is the load capacitance at the output node.

The NBTI and PBTI induced shift can be expressed by the increment of equivalent transistor resistance ΔR_p and ΔR_n. The drivability of NMOS and PMOS should be equalized by proper sizing, i.e. $R_p=R_n=R_{eq}$. The resulting percentage delay shift of the plain inverter is:

$$\frac{\Delta t_{inv}}{t_{inv}} = \frac{\Delta R_p + \Delta R_n}{2R_{eq}} \tag{4.2}$$

The propagation delay for the proposed design can be expressed as:

$$t_d = R_{xg}C_3 + (R_p + R_n + R_h + R_f)(C_2 + C_3) + \left(R_h + R_f\right)(C_1 + C_2 + C_3) \tag{4.3}$$

Here, R_h and R_f are the on-resistance of the header PMOS and footer NMOS, C_1, C_2 and C_3 are the capacitance located at the output nodes pointed out in Fig. 4.18, and R_{xg} is the on-resistance of the transmission gate at the inverter output.

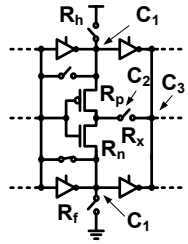
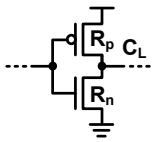
Schematic		
Measurable	$(\frac{\Delta f}{f})_{PBTI}$, $(\frac{\Delta f}{f})_{NBTI}$	$\frac{\Delta f}{f}$
Relationship	$\frac{\Delta f}{f} = \beta[(\frac{\Delta f}{f})_{PBTI} + (\frac{\Delta f}{f})_{NBTI}]$	
β	$\frac{1}{(R_h+R_f)(C_2+C_3)}[(R_h+R_f)(C_1+C_2+C_3) + (R_h+R_p+R_f+R_n)(C_2+C_3)+R_gC_3]$	

Figure 4.18: Overall degradation of a plain ROSC can be calculated using the separate PBTI and NBTI data by calibrating out the delay coefficient factor β introduced by the additional pass gates and switches.

After the selective stress, R_p or R_n is shifted by ΔR_p or ΔR_n and all the other devices are completely turned off thus free of any BTI stress during the stress mode. Assuming

the equalized pull-up and pull-down delay, i.e. R_p=R_n=R_{eq}, R_f=R_h=R_{sw}, the percentage delay shift caused exclusively by PBTI can be written as:

$$\frac{\Delta t_{PBTI}}{t} = \frac{\Delta R_p(C_2+C_3)}{R_{xg}C_3+(R_{sw}+R_{eq})(C_2+C_3)+R_{eq}(C_1+C_2+C_3)} \tag{4.4}$$

And its NBTI counterpart is:

$$\frac{\Delta t_{NBTI}}{t} = \frac{\Delta R_n(C_2+C_3)}{R_{xg}C_3+(R_{sw}+R_{eq})(C_2+C_3)+R_{eq}(C_1+C_2+C_3)} \tag{4.5}$$

Compare the equations above with plain ROSC results in Equ. 4.2, the percentage delay of plain ROSC can be expressed by the sum of PBTI and NBTI shift from the proposed circuit as follows:

$$\frac{\Delta t_{inv}}{t_{inv}} = \beta\left(\frac{\Delta t_{PBTI}}{t} + \frac{\Delta t_{NBTI}}{t}\right), \tag{4.6}$$

where, the parameter β which is the ratio between the PBTI + NBTI delay shift and the plain ROSC delay shift, representing the additional delay introduced by the pass gates can be expressed as:

$$\beta = \frac{R_{xg}C_3+(R_{sw}+R_{eq})(C_2+C_3)+R_{eq}(C_1+C_2+C_3)}{\alpha R_{eq}(C_2+C_3)} \tag{4.7}$$

Here, α is the parameter to differentiate the DC and AC stress case. For realistic DC stress, every other PMOS or NMOS is alternately stressed along a plain ROSC delay path, versus each PMOS or NMOS is stressed in the proposed ROSC design, which gives the factor of α=2 in DC case. But this difference does not appear in AC case as every devices is stressed, i.e., α=1 in AC case.

β is independent of aging as suggested by the above Elmore delay calculation as well as simulation results. HSPICE simulation results in Fig. 4.19 shows that the β value is almost a constant with the V_t shift greater than 0.5%. The jumping at very small V_t shift might due to the impact of the parasitics.

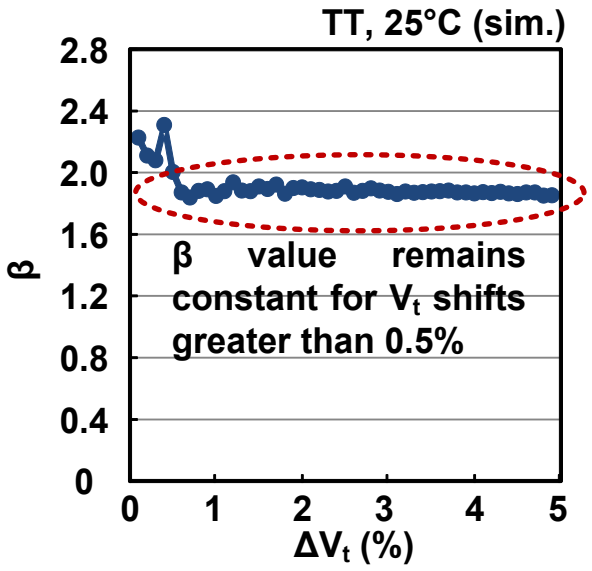


Figure 4.19: HSPICE simulation results show that β is independent of aging.

A plain ROSC was implemented in the same test chip in order to experimentally calibrate out the β value. The frequency shift results for the plain ROSC, and the separated PBTI and NBTI from the proposed circuit are measured from the same die under 1.8V, 200MHz AC stress condition at 25°C, as plotted in Fig. 4.20. We found the β value ranges from 1.93 to 2.09 for the entire stress period which further verifies its independency to device aging. The measured β value average can be readily used to

translate the other measured PBTI and NBTI results to the plain ROSC frequency shift with the same measurement condition.

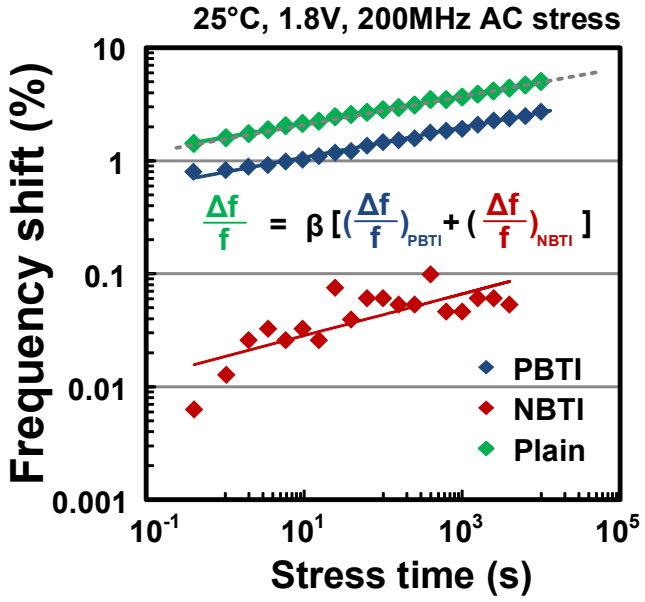


Figure 4.20: The β value can be calculated using the PBTI, NBTI, and combined aging data.

4.5 Conclusions

PBTI and NBTI effects are different in term of physical origin, intensity, and voltage and temperature dependence, thus their impacts on circuit performance differ. It is important to separately study these two effects. However, in a operating circuit, these two aging effects occur at the same time, so the overall degradation is their combination. A ROSC based on-chip monitor is proposed to separately characterize impacts from PBTI

and NBTI on logic circuit frequency. The realistic recovery effect is taken into account for the first time by using a modified ROSC structure. The proposed monitor has other major advantages such as hundreds of nanosecond measurement interrupt and picoseconds time resolution, compared to the previous designs. From the test results from test chips implemented under a high-k metal gate technology, we discovered that both PBTI and NBTI induced frequency shift follows a power law dependence (t^n) with stress time, with a smaller time exponent than reported results due to the short measurement achieved by this design. PBTI in this particular technology is 5X to 10X stronger than NBTI with a smaller recovery rate. A -log(t) dependence was observed for the recovery behavior with different voltage and temperature. Compared with the zero bias condition, a significantly accelerated recovery effect was observed for the realistic bias condition.

Chapter 5.

On-Chip Monitor Design for Characterizing Circuit-Level Electromigration

5.1 Introduction

As the leading back-end-of-line reliability failure mechanism of interconnect lines in VLSI microchips, electromigration (EM) is the process of metal-ion transport due to high current density stress in metals as shown in Fig. 5.1 [23-28]. As the metal-ions migrate due to the mechanical interaction by electron flux, metal depletion or extrusions occurs. The extrusions can results in shorts between interconnects, while the voids at the metal depletion can cause increases in interconnect resistance or catastrophic disconnections. Note that the direction of the electron flux is from the wide and short M2 to narrow and long M1 (down-stream), which is easier to failure compared to the case that the electrons move from wide and short M1 to narrow and long M2 (up-stream). The reason for this is because the void is more likely to form underneath the via for the down-stream case which causes a larger resistance shift [28]. The shift of resistance value typically jump at certain stress time and then increases progressively. The time when the resistance is over

the assigned criterion value (normally 5%) is defined as the time to failure (MTTF). MTTF is an important parameter to evaluate EM reliability, which can be fitted by lognormal distribution for both Al and Cu dual-damascene metallization [30, 31].

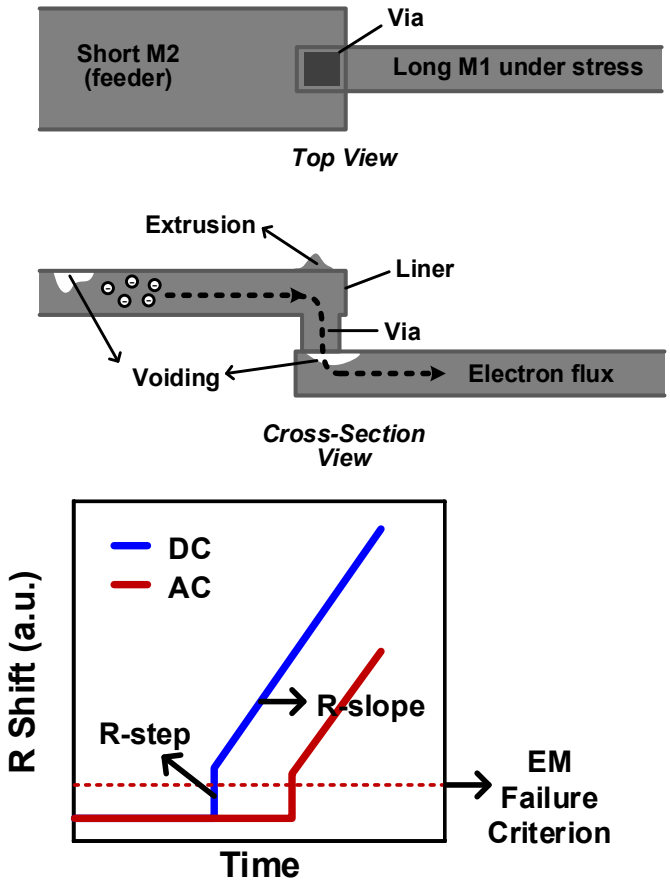


Figure 5.1: (upper) EM mechanism illustrated in an multilayer metallization for down-stream case. (lower) R vs. Time plot of DC EM and AC EM.

MTTF is related to the average current density and the absolute temperature of the interconnect by Black's equation [25]. Symmetric AC current is observed to cause EM failure even though the average current is zero as shown in Fig. 5.1, indicating a partial

instead of a complete recovery process occurs when the current switch to the opposite direction.

With the scaling of the technologies, metal interconnect widths have been reduced to deep submicron range, while the current has not been downscaled in a proportional fashion. This has resulted in an increase of current densities carried by interconnects as predicted by International Technology Roadmap for Semiconductors (ITRS) in Fig. 5.2 [61]. Therefore, the metal leads become more and more susceptible to EM failures.

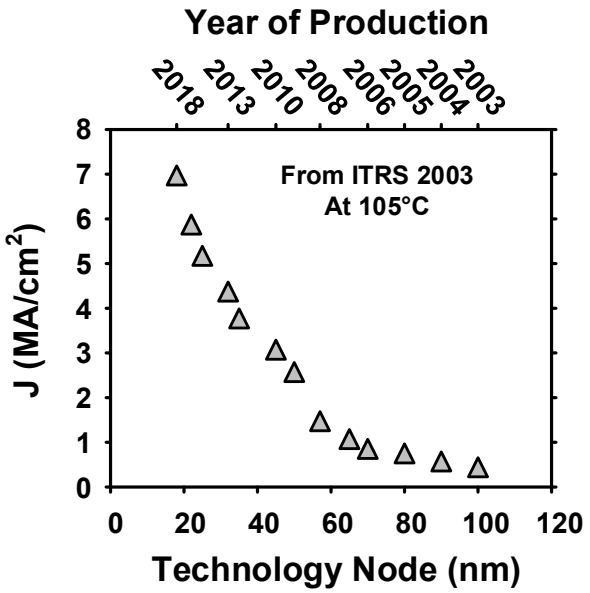
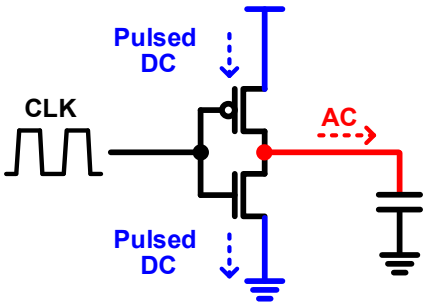


Figure 5.2: Current density delivered versus the technology nodes based on ITRS.

A common approach to model and qualify EM failure is based on the direct current (DC) stress tests. However, most interconnects in an actual circuit are operated and stressed under more complicated current stress patterns such as unipolar pulse (pulsed

DC) or bipolar pulse (AC) [62, 63]. As shown in Fig. 5.3, for a logic gate driven by a clock signal, the power grids encounter pulsed DC current stress, while signal lines suffer from AC current stress. Even though some work conducted by process engineers have found correlations between DC EM and pulsed current EM [23, 62, 64-66] based on the device probing method, none of these work shows EM behavior under the normal operation frequency (GHz). Other drawbacks of the probing method include unrealistic current waveform and the cumbersome test setup.



Type of Net	Current Type	Reliability Concerns
Power grid	Pulsed DC	EM
Signal net	AC	AC EM and Joule heating

Figure 5.3: Two current stress types in normal digital operations.

In this work, we proposed an on-chip ring oscillator (ROSC) based EM monitoring circuit for characterizing circuit level EM and its impact on circuit frequency. With the VCO and current drivers implemented on-chip, realistic current stress conditions can be readily provided with up to 10GHz stress frequency that would be impossible using the

previous probing method. The EM monitoring vehicle supports a bandwidth of stressing and testing 60 samples simultaneously, and provides four stress conditions during stress mode, and can operation frequency shift and resistance shift caused by EM.

5.2 Proposed On-chip EM Monitor Design

5.2.1 EM Test Array design

As shown in Fig. 5.4, the top level schematic diagram of the EM monitor is composed of the following parts: ten stressed ROSCs and one reference ROSC, the beat frequency detection (BFD) system, power switches, VCO and the control circuits. The stressed ROSC and the reference ROSC have identical layout structure to eliminate common mode noise during measurements. BFD system is adopted to detect the ROSC frequency degradation induced by EM at the interconnect lines between the inverter stages [35]. By comparing the frequency of the stressed ROSC and the fresh ROSC (reference), the beat frequency can be detected which largely improves the accuracy and reduces the measurement time. The schematic of the BFD circuit is shown in Fig. 5.4 and the detailed operation demonstration can be found in *Chapter 1.4*.

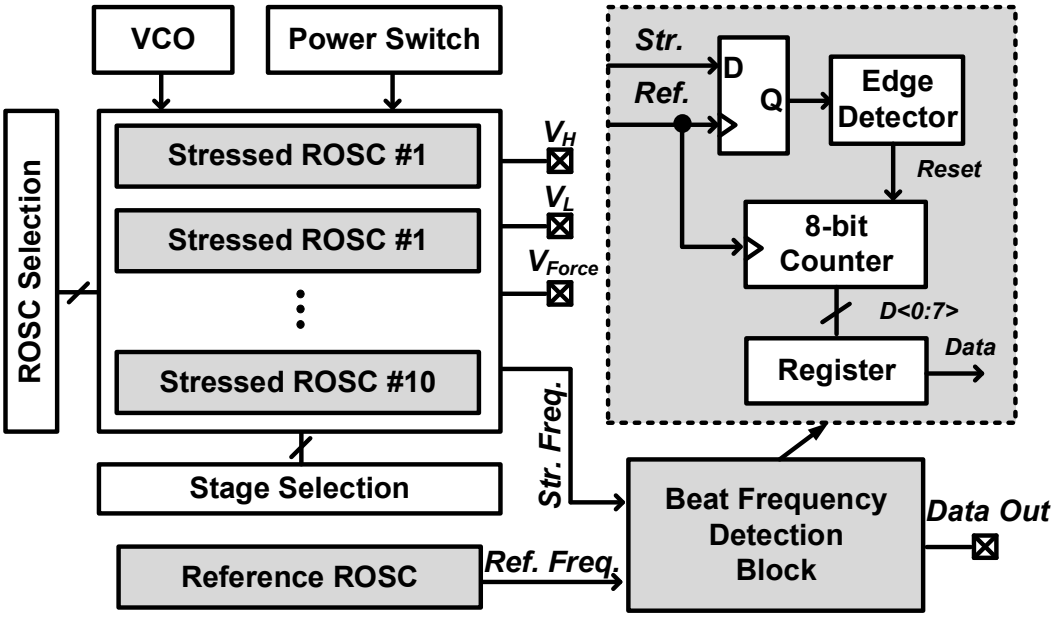


Figure 5.4: Top level schematic of the proposed EM monitor circuit. EM lines in 10 ROSCs can be stressed simultaneously, and beat frequency detection block is adopted to achieve accurate frequency measurement.

Fig. 5.5 illustrates the circuit diagram of a ROSC structure consisting of six inverter stage with long interconnects (EM lines) and a tri-state inverter to enable the oscillation. Note that the delay of tri-state inverter is less than 5% of that of a interconnect driver stage, so the impact of the last stage on the oscillation period is very small. For each interconnect driver, there are two additional drivers on each side of the EM line to providing different EM stress conditions. During the stress mode, the feedback loop is opened by turning off the last tri-state inverter, and the EM lines are stressed by drivers toggled by on-chip VCO which is capable to provide GHz stress signal. In the frequency measurement mode, the additional interconnect drivers are gated off, and the feedback

loop is closed to initiate the free run of the ROSC. For each stage, the resistance values of the interconnect can be measured by four-terminal Kevin method.

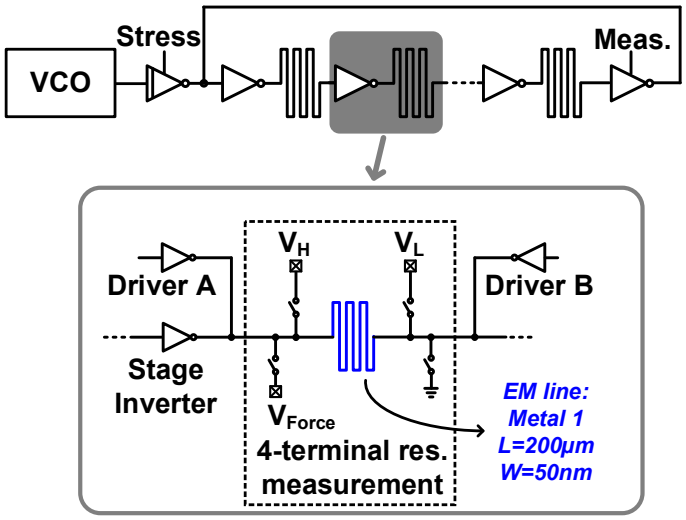


Figure 5.5: Schematic of the ROSC circuit design. Each inverter stage consists additional drivers to provide different stress condition and passgates connected to four probing terminals for resistance measurement.

The interconnect is dual-damascene copper/low-k in the 32nm process technology. Each EM line is a single straight M1 with a length of 200μm. EM line is sized to minimum width to maximized the current density, while the current feeder are made by wider M2 plates to form the downstream EM stress condition. Single via is used for the connection to EM line to create the worst case condition, while multiple vias are used to connect the M2 feeder plate to the driver and receiver, as shown in Fig. 5.6.

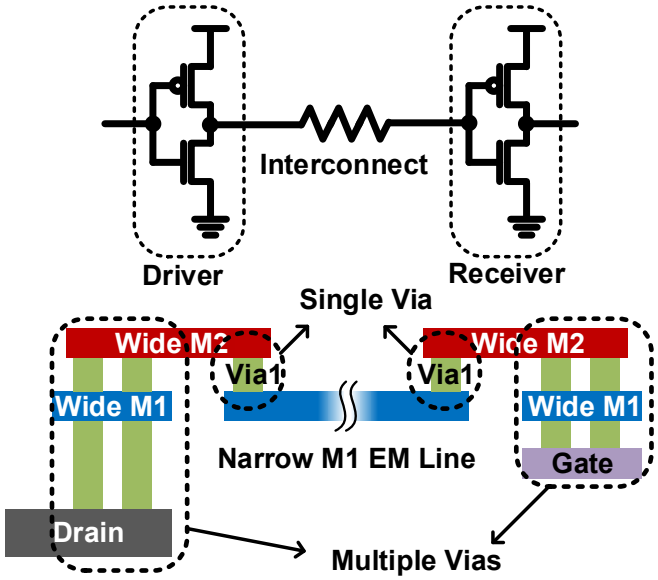


Figure 5.6: Schematic and cross-section view of the interconnect stage under test.

5.2.2 Supported stress and measurement modes

As listed in Fig. 5.7, the proposed design supports four different stress conditions: realistic AC, square AC, DC and square DC. The realistic AC current is the current scheme when a driver is used to charge or discharge the load during transition in digital operations. The square shape AC current waveform is not common in actual circuits, but it gives a larger current density to create an EM failure. The square DC case can be used to study the frequency dependence. And all the test results will be compared to DC stress result which will be set as the base line.

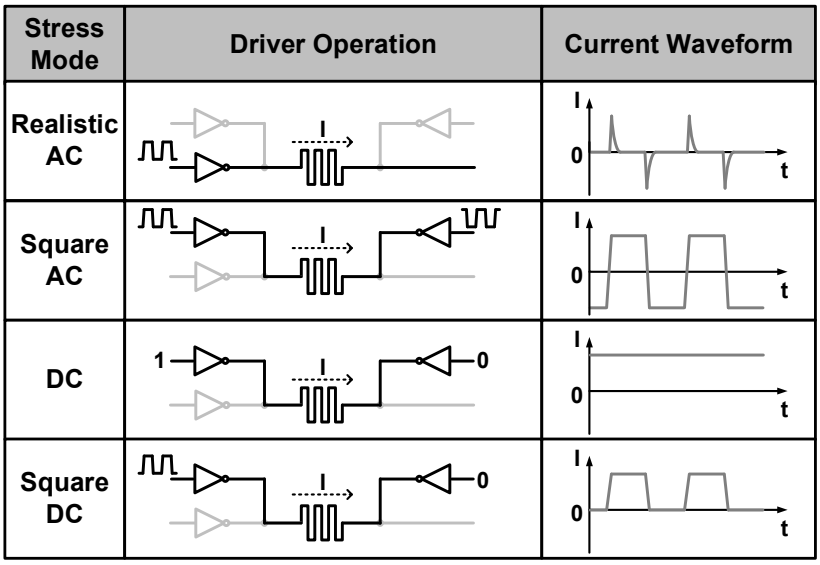


Figure 5.7: The circuit operation diagram and the current waveform for four different stress conditions.

The proposed EM monitor provides two types of measurement mode: the frequency measurement and the resistance measurement, as illustrated in Fig. 5.8. The frequency measurement captures the impact of EM induced interconnect resistance increase on the propagation delay. Another measurement mode is to directly measure the resistance of the interconnect for each stage. During the resistance measurement mode, only one of the six interconnects in a ROSC is selected by enabling the path gates at the four probing pins for Kelvin resistance measurement. V_{Force} and GND probe are used to force a current (I_{IN}) through the interconnect, and V_H and V_L pins are used to measure the voltage across the interconnect. The resistance can be calculated by the measured values as $(V_H-V_L)/I_{IN}$.

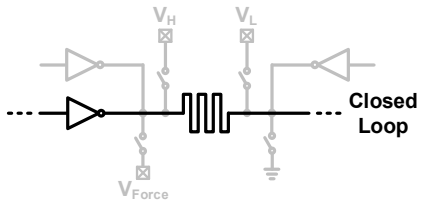
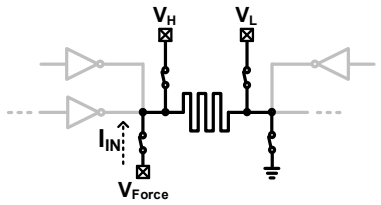
Meaurement Mode	Driver Operation	Methodology Description
Frequency Measurement		Beat Frequency Detection System for Monitoring the EM and BTI Induced Frequency Shift.
Resistant Measurement		Four Terminal Kevin Resistance Measurement: $R_{EM} = (V_H - V_L)/I_{IN}$

Figure 5.8: Two measurement modes capable of measuring frequency or resistance.

5.2.3 On-chip heater design

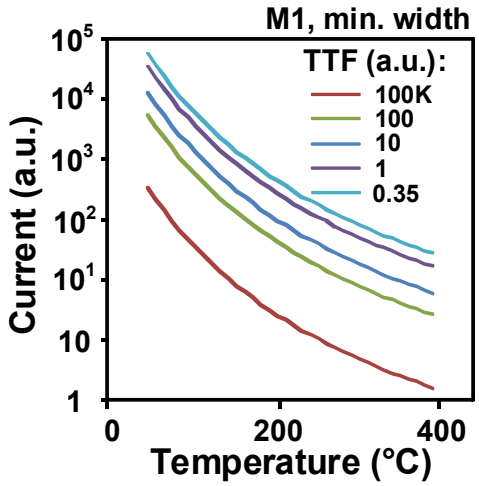


Figure 5.9: The EM failure conditions for different TTF are predicted by EM models provided by the foundry

The maximum current through the EM leads is limited to mA range by the VDD level and the interconnect resistance, elevated stress temperature is required to detect an EM failure in a reasonable stress time. The current and temperature stress condition can be calculated by using the EM models provided by the design manual, as plotted in Fig. 5.9. Based on the extrapolation, the stress temperature for our design needs to be higher than 300°C to keep TTF in the hour-range. In order to heat up the EM lead to the targeting temperature while keeping the transistors in the working temperature, on-chip heaters are normally used to provide a localized high temperature. The top-level layout of the circuit is shown in Fig. 5.10. The location of the interconnect lines are in the middle area which is covered by heater grid, while the drivers and other control circuits are placed on the sides, where the temperature is low.

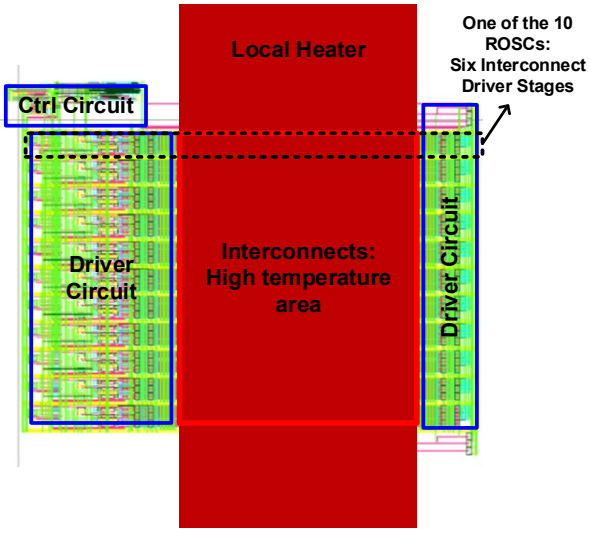


Figure 5.10: Top level layout of the EM monitor circuit.

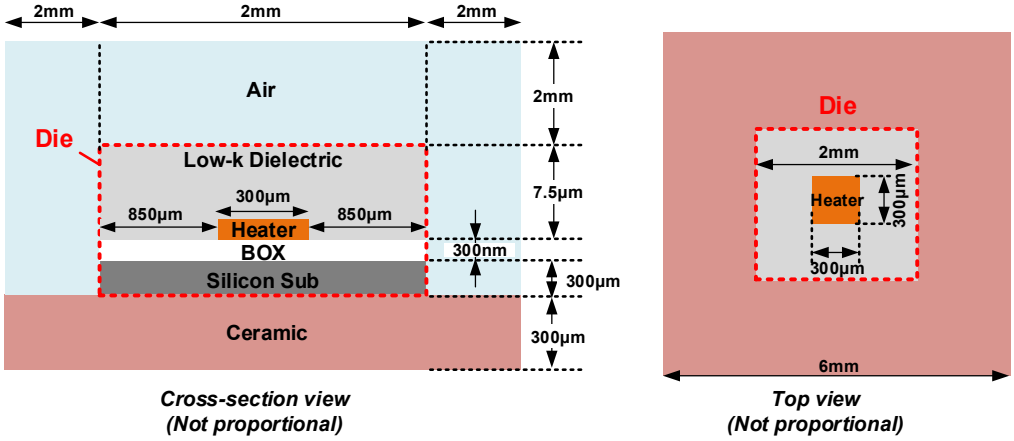


Figure 5.11: Package structure for ANASYS thermal simulation.

Layer	Material	Thermal Conductivity ($W*m^{-1}*K^{-1}$)	Thickness (µm)
Heater	Poly-silicon	400	0.04
Dielectric	Low-k dielectric	0.2	7.5
Substract	Silicon	15	300
BOX	Silicon dioxide	1.4	0.3
Package	Ceramic	200	300
Air	Air	0.0242	2000

Figure 5.12: Detailed layer information used for the thermal simulation.

To design a proper heater structure, ANSYS™ is used to conduct the thermal simulation with the finite element method. Poly silicon resistor is used to build the heater as it is less vulnerable to EM stress compared to metal. The cross-section and the top views of the ANASYS model are shown in Fig. 5.11. As can be seen, the surrounding air and ceramic floor are also included in the simulation as they will have thermal

interchange with the packaged chip. The die size is 2mm by 2mm and the package size is 6mm by 6mm. The detailed information on the thickness and thermal parameter for each layer is not provided by the foundry, so the numbers are based on our best guess listed in Fig. 5.12. To make the temperature of the targeting area consistent, parallel strip-shaped heater are used and voltages for resistor strips can be separately controlled. The layout, voltage application and the corresponding temperature distribution based on the ANSYS simulation are shown in Fig. 5.13.

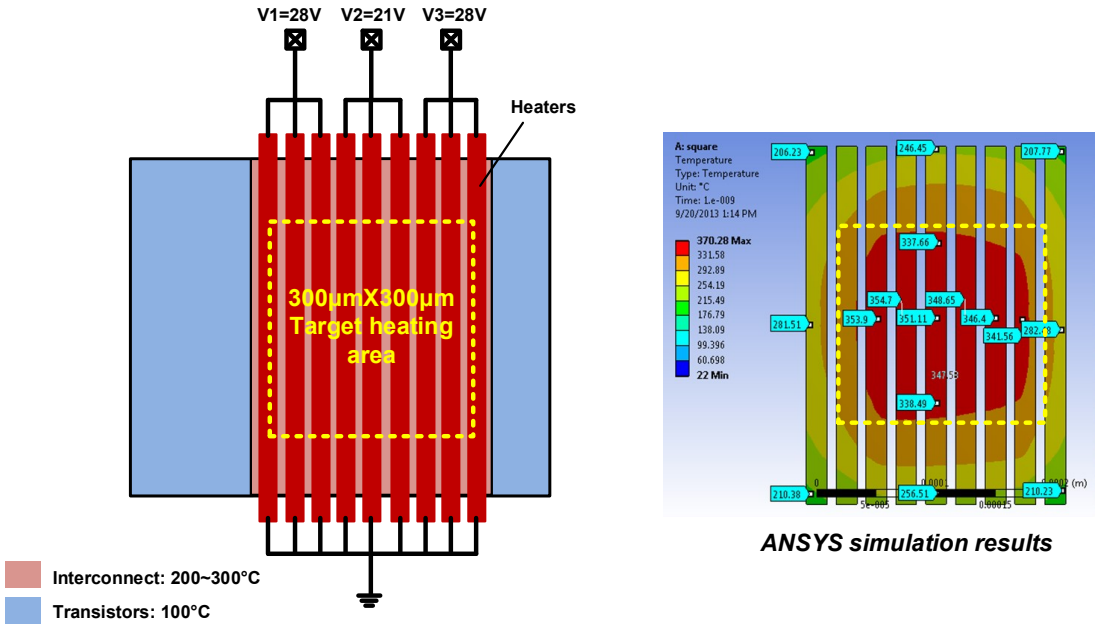
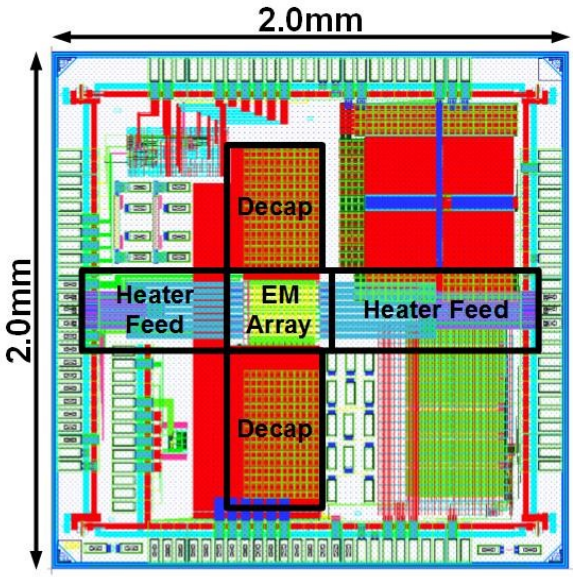


Figure 5.13: Layout of the poly heater and the ANASYS thermal simulation results.

5.2.4 Measurement plan

A testchip will be implemented with a 32nm SOI technology with the top level layout and the circuit specifications shown in Fig. 5.14. The yet to perform measurements will be conducted in the following steps: Firstly, the effectiveness of the on-chip heater will be checked. We will measure the metal resistance to calculate the temperature by using temperature coefficient of resistance (TCR). Then we will measure the TTF with different stress conditions and characterize the resistance shift cause by the EM effect. Based on the data, we will study the correlation between the realistic current stress versus the conventional DC based EM model, as well as the frequency dependency of the AC and pulsed DC EM lifetime.



Process	32nm SOI CMOS
VDD / IO Supplies	0.9V / 1.8V
# of ROSC	10
# of EM Stage	60
Current Direction	M2→M1 (downstream)
EM Line	M1, W=50nm, L=200μm
Area	Heater: 1500x256μm^2 Circuit: 520x1500μm^2
Target Temperature	310°C
Meas. Δf Resolution	0.01% ~0.1%

Figure 5.14: EM monitor testchip layout and design specifications.

5.3 Conclusion

In this chapter, an on-chip EM monitor for characterizing circuit level EM effect is proposed. It is the first known EM monitor to provide GHz frequency stress with realistic current waveform. Four different stress conditions can be provided by the proposed circuit. And both of the resistance and the frequency of the ROSC are supported. The purpose of the testchip is to study the circuit level EM with the high frequency stress. The experimental measurements and the further discussion will be conducted after the completion of this thesis.

Chapter 6.

Conclusions

The parametric shifts and circuit failures caused by device reliability issues such as Bias Temperature Instability (BTI), Hot Carrier Injection (HCI), and Electromigration (EM) have become more prevalent as electrical fields and current density continue to increase in scaled devices. Research, process and design groups have been devoting a significant amount of resources and efforts to better understand the aging mechanisms, and to explore strategies to conduct precise aging prediction. Our study in this field has been focused on developing accurate and efficient on-chip monitoring circuits to investigate different aging mechanisms and study their impacts on circuit parameters. In this thesis, a number of unique on-chip sensing systems have been proposed which provide us with significant advantages, such as pico-second timing resolution for usage condition stress, micro-second measurement interrupt to prevent unwanted recovery, and excellent immunity to voltage and temperature drifts. The odometer designs proposed in this thesis utilize standard logic gates and a simple scan based interface, making them suitable for integrating into an actual processor system.

In this thesis, four dedicated on-chip circuit designs that we have implemented over two generations of process technology to perform reliability tests on different types of circuits will be presented.

In Chapter 2, for the first time we explored the detailed aging behavior of interconnect paths. The degradation in the performance caused by BTI and HCI in the driver circuit exhibits a strong dependency on the interconnect RC parameters. To thoroughly understand this effect, firstly we demonstrated a 65nm interconnect odometer test vehicle that is able to accurately measure the frequency degradation of signal paths with different interconnect lengths. Our experimental results show that the frequency degradation caused by BTI decreases with increasing interconnect length; on the other hand, the HCI induced degradation peaks at around 500μm. The distinctly different interconnect dependences were explained by our circuit simulations that take in to account the effective stress time and stress voltage during signal transitions. We next proposed a closed loop modeling methodology which precisely captures the interplay between interconnect RC parameters and the changes in stress conditions. Finally, a case study of the closed loop model was discussed in which the delay shift for a practical interconnect path was minimized using optimal number of buffers.

Chapter 3 was focus on the effects of asymmetric BTI aging on the performance of circuits, particularly for various low-power logic circuits and memories with clock gating techniques. A simple and practical method was proposed to accurately measure the duty-cycle shift of a delay path based on the ring oscillator based beat-frequency detection

framework. In the second half of this chapter, a SRAM reliability detection vehicle was presented to detect the asymmetric aging impact on the read performance in an actual SRAM. Based on the statistical test results from SRAM reliability macro, we observed that the variation of read frequency increases with the stress intensity. Both of the average and the standard deviation of read frequency shift follow a power law dependence with the stress time. The asymmetric aging increases the wordline duty-cycle, which in turn reduces the sensing error caused by the short wordline duty-cycle.

A ROSC-based on-chip monitor is proposed in Chapter 4 to separately characterize impacts from PBTI and NBTI on logic circuit frequency. The realistic recovery effect is taken into account for the first time by using a modified ROSC structure. The proposed monitor also has many other advantages such as hundreds-of-nanosecond measurement interrupt and picosecond time resolution, compared to previous designs. From the test results from test chips implemented under a high-k metal gate technology, we discovered that both PBTI and NBTI induced frequency shift follows a power law dependence (t^n) on stress time, with a time exponent smaller than previous reported results, which is due to the short measurement achieved by this design. PBTI in this particular technology is 5X to 10X stronger than NBTI with a smaller recovery rate. A -log(t) dependence was observed for the recovery behavior with different voltages and temperatures. There is a significantly accelerated recovery effect with the realistic bias condition, compared with the zero bias condition.

In Chapter 5 we presented an on-chip EM monitor for characterizing the circuit level EM effect. This is the first known ROSC-based EM monitor to provide GHz frequency stress with realistic current waveforms. Four different stress conditions can be provided by the proposed circuit. The measurements of both the interconnect resistance and the ROSC frequency are supported. The purpose of the test chip is to study the circuit-level EM effect with high frequency stress and its impact on the frequency degradation of interconnect paths.

Bibliography

[1] R. Degraeve, M. Aoulaiche, B. Kaczer, P. Roussel, T. Kauerauf, S. A. Sahhaf*, et al.*, "Review of Reliability Issues in High-k/Metal Gate Stacks," in *IEEE Int. Symp. on Physical and Failure Analysis of Integrated Circuits*, 2008, pp. 1-6.

[2] Y. Lee, W. McMahon, N. Mielke, Y. Lu, and S. Walstra, "Managing Bias-Temperature Instability for Product Reliability," in *IEEE Int. Symp. on VLSI Technology, Systems and Applications*, 2007, pp. 1-2.

[3] Y.-H. Lee, N. Mielke, B. Sabi, S. Stadler, R. Nachman, and S. Hu, "Effect of pMOST Bias-Temperature Instability on Circuit Reliability Performance," in *IEEE Int. Electron Devices Meeting*, 2003, pp. 14.6.1-14.6.4.

[4] V. Huard and M. Denais, "Hole Trapping Effect on Methodology for DC and AC Negative Bias Temperature Instability Measurements in PMOS Transistors," in *IEEE Int. Reliability Physics Symp.*, April 2004, pp. 40-45.

[5] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, D. Roy, and A. Bravaix, "New Perspectives on NBTI in Advanced Technologies: Modeling & Characterization," in *IEEE Euro. Solid-State Device Research Conf.*, 2005, pp. 399-402.

[6] R. Vattikonda, W. Wang, and Y. Cao, "Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design," in *IEEE Design Automation Conf.*, 2006, pp. 1047-1052.

[7] M. Ershov, R. Lindley, S. Saxena, A. Shibkov, S. Minehane, J. Babcock*, et al.*, "Transient Effects and Characterization Methodology of Negative Bias Temperature Instability in pMOS Transistors," in *IEEE Int. Reliability Physics Symp.*, 2003, pp. 606-607.

[8] D. Ielmini, M. Manigrasso, F. Gattel, and M. G. Valentini, "A New NBTI Model Based on Hole Trapping and Structural Relaxation in MOS Dielectrics," *IEEE Trans on Electron Devices,* vol. 56, pp. 1943-1952, 2009.

[9] T. Grasser and B. Kaczer, "Evidence That Two Tightly Coupled Mechanisms Are Responsible for Negative Bias Temperature Instability in Oxynitride MOSFETs," *IEEE Tran. Electron Devices,* vol. 56, pp. 1056-1062, 2009.

[10] J. Keane, T. H. Kim, and C. H. Kim, "An On-Chip NBTI Sensor for Measuring PMOS Threshold Voltage Degradation," presented at the IEEE Int. Symp. on Low-Power Electronics, 2007.

[11] S. Aota, S. Fujii, Z. W. Jin, and e. al., "A New Method for Precise Evaluation of Dynamic Recovery of Negative Bias Temperature Instability," in *IEEE Int. Conf. on Microelectronic Test Structures*, April 2005, pp. 197-199.

[12] C. Schlunder, W. Heinrigs, W. Gustin, and H. Raisinger, "On the Impact of the NBTI Recovery Phenomenon on Lifetime Prediction of Modern p-MOSFETs," in *IEEE Int. Integrated Reliability Workshop*, October 2006, pp. 1-4.

[13] S. Rangan, N. Mielke, and E. Yeh, "Universal Recovery Behavior of Negative Bias Temperature Instability," *IEEE Int. Electron Devices Meeting,* pp. 14.3.1-14.3.4, 2003.

[14] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger*, et al.*, "Recent Advances in Understanding The Bias Temperature Instability," in *IEEE Int. Electron Devices Meeting*, 2010, pp. 4.4.1-4.4.4.

[15] S. Pae, M. Agostinelli, M. Brazier, R. Chau, G. Dewey, T. Ghani*, et al.*, "BTI Reliability of 45nm High-K + Metal-Gate Process Technology," in *IEEE Int. Reliability Physics Symp.*, 2008, pp. 352-357.

[16] C. Hu, S. C. Tam, F.-C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement," *IEEE Journal of Solid-State Circuits,* vol. 20, pp. 295-305, 1985.

[17] Z. Chen, K. Hess, J. Lee, E. Rosenbaum, I. Kizilyalli, S. Chetlur*, et al.*, "On the Mechanism for Interface Trap Generation in MOS Transistors Due to Channel Hot Carrier Stressing," *IEEE Electron Device Letters,* vol. 21, pp. 24-26, 2000.

[18] K. Hess, J. Lee, Z. Chen, J. W. Lyding, Y. Kim, B. Kim*, et al.*, "An Alternative Interpretation of Hot Electron Interface Degradation in NMOSFETs: Isotope Results Irreconcilable with Major Defect Generation by Holes?," *IEEE Trans. Electron Devices,* vol. 46, pp. 1914-1916, 1999.

[19] K. N. Quader, E. R. Minami, W.-J. Ko, P. K. Ko, and C. Hu, "Hot-Carrier-Reliability Design Guidelines for CMOS Logic Circuits," *IEEE Journal of Solid-State Circuits,* vol. 29, pp. 253-262, 1994.

[20] A. Acovic, G. La Rosa, and Y.-C. Sun, "A Review of Hot-Carrier Degradation Mechanisms in MOSFETs," *Microelectronics Reliability,* vol. 36, pp. 845-869, 1996.

[21] C. Schlunder, S. Aresu, G. Georgakos, W. Kanert, H. Reisinger, K. Hofmann*, et al.*, "HCI vs. BTI? - Neither One's Out," in *IEEE Int. Reliability Physics Symp.*, 2012, pp. 2F.4.1-2F.4.6.

[22] X. Wang, P. Jain, D. Jiao, and C. H. Kim, "Impact of Interconnect Length on BTI and HCI Induced Frequency Degradation," in *IEEE Int. Reliability Physics Symp.*, 2012, pp. 2F.5.1-2F.5.6.

[23] L. M. Ting, J. S. May, W. R. Hunter, and J. W. McPherson, "AC Electromigration Characterization and Modeling of Multilayered Interconnects," in *IEEE Int. Reliability Physics Symp.*, 1993, pp. 311-316.

[24] J. Tao, N. W. Cheung, and C. Hu, "Characterization and Modeling of Electromigration Failures in Multilayered Interconnects and Barrier Layer Materials," *IEEE Tran. Electron Devices,* vol. 43, pp. 1819-1825, 1996.

[25] J. R. Black, "Electromigration Failure Modes in Aluminum Metallization for Semiconductor Devices," *IEEE Proc.,* vol. 57, pp. 1587-1594, 1969.

[26] B.-K. Liew, N. W. Cheung, and C. Hu, "Projecting Interconnect Electromigration Lifetime for Arbitrary Current Waveforms," *IEEE Tran. Electron Devices,* vol. 37, pp. 1343-1351, 1990.

[27] J. Clement, "Vacancy Supersaturation Model for Electromigration Failure under DC and Pulsed DC stress," *Journal of Applied Physics,* vol. 71, pp. 4264-4268, 1992.

[28] A. S. Oates and M. H. Lin, "Void Nucleation and Growth Contributions to The Critical Current Density for Failure of Cu Vias," in *IEEE Int. Reliability Physics Symp.*, 2009, pp. 452-456.

[29] W. Yang, Z. Clik-Butler, H. H. Hoang, and W. R. Hunter, "Detection of Electromigration in VLSI Metalizations Layers by Low-Frequency Noise Measurements," in *IEEE Int. Electron Devices Meeting*, 1989, pp. 681-684.

[30] R. G. Filippi, G. A. Biery, and R. A. Wachnik, "Paradoxical Predictions and a Minimum Failure time in Electromigration," *Applied Physics Letters,* vol. 66, pp. 1897-1899, 1995.

[31] M. Gall, C. Capasso, D. Jawarani, R. Hernandez, H. Kawasaki, and P. S. Ho, "Statistical Analysis of Early Failures in Electromigration," *Journal of Applied Physics,* vol. 90, pp. 732-740, 2001.

[32] J. B. Velamala, V. Reddy, R. Zheng, S. Krishnan, and C. Yu, "On The Bias Dependence of Time Exponent in NBTI and CHC Effects," in *IEEE Int. Relibility Physics Symp.*, 2010, pp. 650-654.

[33] W. Weber, M. Brox, T. Künemund, M. Mühlhoff, and D. Schmitt-Landsiedel, "Dynamic Degradation in MOSFET's. II. Application in The Circuit Environment," *IEEE Tran. Electron Devices,* vol. 38, pp. 1859-1867, 1991.

[34] W. Jiang, H. Le, J. Chung, T. Kopley, P. Marcoux, and C. Dai, "Assessing Circuit-Level Hot-Carrier Reliability," in *IEEE Int. Reliability Physics Symp.*, 1998, pp. 173-179.

[35] J. Keane, X. Wang, D. Persaud, and C. H. Kim, "An All-In-One Silicon Odometer for Separately Monitoring HCI, BTI, and TDDB," *IEEE Journal of Solid-State Circuits,* vol. 45, pp. 817-829, 2010.

[36] T. Kim, R. Persaud, and C. H. Kim, "Silicon Odometer: An On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circuits," *IEEE Journal of Solid-State Circuits,* vol. 43, pp. 874-880, 2008.

[37] U. R. Karpuzcu, B. Greskamp, and J. Torrellas, "The BubbleWrap Many-Core: Popping Cores for Sequential Acceleration," in *IEEE Int. Symp. Microarchitecture*, 2009, pp. 447-458.

[38] E. Karl, P. Singh, D. Blaauw, and D. Sylvester, "Compact In-Situ Sensors for Monitoring Negative-Bias-Temperature-Instability Effect and Oxide Degradation," in *IEEE Int. Solid-State Circuits Conf.*, 2008, pp. 410-623.

[39] P. Singh, F. ZhiYoong, M. Wieckowski, S. Hanson, M. Fojtik, D. Blaauw*, et al.*, "Early Detection of Oxide Breakdownt through In-Situ Degradation Sensing," in *IEEE Int. Solid-State Circuits Conf.*, 2010, pp. 190-191.

[40] K. Hofmann, H. Reisinger, K. Ermisch, C. Schlunder, W. Gustin, T. Pompl*, et al.*, "Highly Accurate Product-Level Aging Monitoring in 40nm CMOS," in *IEEE Symp. VLSI Technology*, 2010, pp. 27-28.

[41] E. Saneyoshi, K. Nose, and M. Mizuno, "A Precise-Tracking NBTI-Degradation Monitor Independent of NBTI Recovery Effect," in *IEEE Int. Solid-State Circuits Conf.*, 2010, pp. 192-193.

[42] M. Chen, V. Reddy, J. Carulli, S. Krishnan, V. Rentala, V. Srinivasan*, et al.*, "A TDC-Based Test Platform for Dynamic Circuit Aging Characterization," in *IEEE Int. Reliability Physics Symp.*, 2011, pp. 2B.2.1-2B.2.5.

[43] J.-J. Kim, R. M. Rao, J. Schaub, A. Ghosh, A. Bansal, K. Zhao*, et al.*, "PBTI/NBTI Monitoring Ring Oscillator Circuits with On-Chip Vt Characterization and High Frequency AC Stress Capability," in *IEEE Symp. VLSI Circuits*, 2011, pp. 224-225.

[44] P.-F. Lu and K. A. Jenkins, "A Built-In BTI Monitor for Long-Term Data Collection in IBM Microprocessors," in *IEEE Int. Reliability Physics Symp.*, 2013, pp. 4A.1.1-4A.1.6.

[45] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy*, et al.*, "Review on High-k Dielectrics Reliability Issues," *IEEE Tran. Device Mater. Rel.,* vol. 5, pp. 5-19, 2005.

[46] J. Keane, D. Persaud, and C. H. Kim, "An All-In-One Silicon Odometer for Separately Monitoring HCI, BTI, and TDDB," in *IEEE Symp. VLSI Circuits*, 2009, pp. 108-109.

[47] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on The Temporal Performance Degradation of Digital Circuits," *IEEE Tran. Electron Devices,* vol. 26, pp. 560-562, 2005.

[48] R. Fernández, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodríguez, M. Nafría*, et al.*, "AC NBTI Studied in The 1 Hz -- 2 GHz Range on Dedicated On-Chip CMOS Circuits," in *IEEE Int. Electron Devices Meetings*, 2006, pp. 1-4.

[49] E. Takeda and N. Suzuki, "An Empirical Model for Device Degradation Due to Hot-Carrier Injection," *IEEE Tran. Electron Devices,* vol. 4, pp. 111-113, 1983.

[50] Y. Wang, H. Ahn, U. Bhattacharya, T. Coan, F. Hamzaoglu, W. Hafez*, et al.*, "A 1.1GHz 12uA/Mb-Leakage SRAM Design in 65nm Ultra-Low-Power CMOS with Integrated Leakage Reduction for Mobile Applications," in *IEEE Int. Solid-State Circuits Conf.*, 2007, pp. 324-606.

[51] M. B. Ketchen, M. Bhushan, and R. Bolam, "Ring Oscillator Based Test Structure for NBTI Analysis," in *IEEE Int. Conf. Microelectronic Test Structures*, 2007, pp. 42-47.

[52] K. Stawiasz, K. A. Jenkins, and P.-F. Lu, "On-Chip Circuit for Monitoring Frequency Degradation Due to NBTI," in *IEEE Int. Reliability Physics Symp.*, 2008, pp. 532-535.

[53] J. Keane, W. Zhang, and C. H. Kim, "An Array-Based Odometer System for Statistically Significant Circuit Aging Characterization," *IEEE Journal of Solid-State Circuits,* vol. 46, pp. 2374-2385, 2011.

[54] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "NBTI-Aware Synthesis of Digital Circuits," in *IEEE Design Automation Conf.*, 2007, pp. 370-375.

[55] M. Agostinelli, S. Pae, W. Yang, C. Prasad, D. Kencke, S. Ramey*, et al.*, "Random Charge Effects for PMOS NBTI in Ultra-Small Gate Area Devices," in *IEEE Int. Reliability Physics Symp.*, 2005, pp. 529-532.

[56] J.-J. Kim, R. Rao, S. Mukhopadhyay, and C.-T. Chuang, "Ring Oscillator Circuit Structures for Measurement of Isolated NBTI/PBTI Effects," in *IEEE Int. Conf. on Integrated Circuit Design and Technology and Tutorial*, 2008, pp. 163-166.

[57] S. Kupke, S. Knebel, G. Roll, S. Slesazeck, T. Mikolajick, G. Krause*, et al.*, "OFF-State Induced Threshold Voltage Relaxation after PBTI Stress," in *IEEE Int. Integrated Reliability Workshop*, 2012, pp. 95-98.

[58] A. Kerber, K. Maitra, A. Majumdar, M. Hargrove, R. J. Carter, and E. A. Cartier, "Characterization of Fast Relaxation During BTI Stress in Conventional and Advanced CMOS Devices With HfO2/TiN Gate Stacks," *IEEE Trans. Electron Devices,* vol. 55, pp. 3175-3183, 2008.

[59] K. T. Lee, W. Kang, E.-A. Chung, G. Kim, H. Shim, H. Lee*, et al.*, "Technology Scaling on High-K/Metal-Gate FinFET BTI Reliability," in *IEEE Int. Reliability Physics Symp.*, 2013, pp. 2D.1.1-2D.1.4.

[60] S. Deora, V. D. Maheta, A. E. Islam, M. A. Alam, and S. Mahapatra, "A Common Framework of NBTI Generation and Recovery in Plasma-Nitrided SiON p-MOSFETs," *IEEE Electron Device Letters,* vol. 30, pp. 978-980, 2009.

[61] Y. Park, *VMIC Seminar,* 2004.

[62] D. T. Blaauw, C. Oh, V. Zolotov, and A. Dasgupta, "Static Electromigration Analysis for On-Chip Signal Interconnects," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems,* vol. 22, pp. 39-48, 2003.

[63] O. Chanhee, D. Blaauw, M. Becer, V. Zolotov, R. Panda, and A. Dasgupta, "Static Electromigration Analysis for Signal Interconnects," in *IEEE Int. Symp. Quality Electronic Design*, 2003, pp. 377-382.

[64] D. G. Pierce, E. S. Snyder, S. E. Swanson, and L. W. Irwin, "Wafer-Level Pulsed-DC Electromigration Response at Very High Frequencies," in *IEEE Int. Reliability Physics Symp.*, 1994, pp. 198-206.

[65] J. Tao, N. W. Cheung, and C. Hu, "Metal Electromigration Damage Healing under Bidirectional Current Stress," *IEEE Electron Device Letters,* vol. 14, pp. 554-556, 1993.

[66] K. N. Tu, "Recent Advances on Electromigration in Very-Large-Scale-Integration of Interconnects," *Journal of Applied Physics,* vol. 94, pp. 5451-5473, 2003.