Array-Based On-Chip Hardware Monitors for Statistically Efficient Integrated Circuit Reliability Characterization

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

UNIVERSITY OF MINNESOTA

BY

NAKUL PANDE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Prof. Chris H. Kim, Advisor

September 2020

Acknowledgements

First of all, I would like to thank my advisor, Professor Chris H. Kim, for the consistent guidance and support I have received from him during the course of my doctoral studies. Few students are lucky to have a supervisor as Professor Kim and as such I have learnt a lot from him both personally and professionally.

Next, I would like to thank my PhD examination committee members: Professor Sachin Sapatnekar, Professor Steven Koester and Professor Antonia Zhai for their insightful comments, queries and suggestions, during my preliminary oral examination as well as during the course of my thesis review.

I would also like to thank all my colleagues at the VLSI Research Lab for their support during my graduate years. In particular, I would like to thank Chen Zhou for his help with the 16nm Electromigration test-chip tape-out and Luke Everson for his help with the neutron beam irradiation setup at the Los Alamos Neutron Science Center. I would also like to thank Ibrahim Ahmed for his good company, frequent motivation and support over the years.

I would especially like to thank Carlos Soria and Chimai Nguyen of the ECE IT team at the University Of Minnesota. Often times, I would run into tool issues at the very last minute of a tape-out and would reach out to them at odd times. I would receive help sooner than expected and needless to say, I would not have met any of the deadlines in time without their support.

I would like to thank my parents, Dr. S. Pande and Mr. Raj K. Pande for their unconditional love, constant prayers and incessant support which kept me sailing smoothly even in the toughest of times. Very few people are fortunate enough to have such parents.

Words cannot describe the gratitude I feel towards my spiritual gurus Neem Karoli Baba ji Maharaj and Mai ji who have bestowed upon me their grace since the day I was born. Their blessings have always made the impossible, possible in this lifetime.

I would like to thank the technique of insight meditation I have had the fortune to learn during the past year. Practicing this technique, I have come to appreciate the importance of developing a calm and equanimous mind, even in the midst of the storm, which has changed my outlook to research and life in general.

I am extremely grateful to my Professors at Indian Institute of Technology, Bombay, without who's guidance I would have never felt at ease with integrated circuits to begin with. I deeply thank Professor Dipankar Saha, Professor Swaroop Ganguly, Professor Dinesh Sharma, Professor Juzer Vasi and Professor Madhav P. Desai for introducing me to the world of microelectronics and teaching me the subtleties of semiconductor devices and integrated circuit design.

Finally, I would like to thank God, whom I know by different names, by who's will I have managed to survive thus far and who has bestowed upon me his grace my entire life.

Dedication

To my loving parents and to Kainchi Dham

Abstract

The miniaturization of feature sizes with every technology generation in accordance with Moore's Law coupled with the innovations in the transistor design have resulted in a significant performance improvement over the years. However, this has also introduced additional reliability concerns, such as increased on-chip current densities, self-heating effects, process variations, increased susceptibility to radiation-induced soft errors due to the increased device counts and the like. Accurate assessment of process reliability from multiple perspectives has therefore now become quintessential for every upcoming technology generation. To this end, the efforts in this thesis are focused towards circuit-based techniques aimed at making accurate reliability assessment on-chip feasible, and yet simple, by, 1) reducing the equipment complexity, 2) improving the die-area utilization (or the Devices-Under-Test (DUT) count), 3) improving test-efficiency by providing data from a statistically significant sample size in a limited amount of time, 4) ensuring representativeness to the real-case scenario and 5) making the mode selection and data collection easy using a simplistic digital on-chip control.

In this work we have proposed and implemented five such array-based characterization vehicles: The first one is a hardware monitor for studying the aging dynamics of diode-connected MOSFETs, the second is a test-vehicle for characterizing Electromigration (EM) induced resistance degradation on-chip, the third is a fully-digital EM characterization macro wherein monitoring the Bit-Error-Rate (BER) of the interconnect path is proposed as a new metric for capturing the impact of EM induced resistance shifts directly in terms of the interconnect path's signaling ability, the fourth one features a combination of capabilities borrowed from the previous two EM test-vehicles,

the motivation being the translation of a measured resistance change to an accompanying change in the measured BER of an interconnect path and finally, the fifth one is a highly scalable, standard combinational logic oriented Soft-Error-Rate (SER) and Single-Event-Transient (SET) pulsewidth characterization macro. Each of the proposed test-vehicles and measured data obtained thereof have demonstrated a methodology for easy and statistically efficient on-chip reliability characterization in comparison to the conventional industry-standard approaches, while also advancing the current state of the art.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The continuous downscaling of feature sizes has undoubtedly proven the key promised benefits of increased on-chip device densities, reduction in operating power, increased operating frequencies and the like. The change in transistor structure going from planar to FinFETs has further presented a vastly superior alternative from the perspective of power, performance, and area compared to the planar counterparts. However, the ultra-scaled feature sizes have also resulted in additional reliability issues such as increased on-chip current densities, in conjunction with localized thermal effects (self-heating) which are getting worse with every upcoming generation of FinFETs [1], [2] and have resulted in renewed Electromigration (EM) concerns for power and signaling interconnections alike. The per device Soft-Error-Rate (SER) has been reported [3], [4], [5] to decrease with the change in transistor structure from planar to FinFETs, however, on account of the increase in on-chip device densities as a consequence of miniaturization, radiation induced soft-errors continue to still remain a critical reliability concern. Furthermore, analog counterparts in a certain technology may have distinct biasing requirements to meet a set target criteria such as a certain gain, Signal-to-Noise Ratio (SNR) etc. and thus require a separate reliability evaluation. In addition, increased variations further complicate the overall reliability analysis, necessitating collection of failure statistics from a statistically significant sample size for the devices / circuits under test. Finally, for a confident reliability assessment, the scenario must also be representative of the realistic on-chip circuit counterpart, rather than the individual transistor or DUT level.

To this end, the efforts in this thesis focus on circuit-based test-vehicles which would make accurate reliability assessment on-chip feasible, and yet simple, by reducing the equipment complexity, improving the die-area utilization (or the DUT count), improving the test-efficiency in terms of the total test time and the number of DUTs tested, providing statistically significant data and ensuring representativeness to the real-case. A chapter-wise summary of the hardware monitors implemented as part of this work is presented below:

1.1 An Array-Based Characterization Vehicle for Investigating the Impact of Feedback on the Aging Dynamics of Diode-connected MOSFETs

This work presents measured data corresponding to a comprehensive reliability characterization of diode-connected MOS transistors. An array-based test-vehicle, with the specific aim of quantifying the impact of feedback on the aging dynamics for the circuit configuration of interest was proposed and implemented in a 65nm Low-Power (LP) process. Through detailed measurement data obtained using the test-vehicle we, (1) characterize the impact of feedback on the aging rate and compare it to the no-feedback case & (2) evaluate the efficacy of iterative simulations for lifetime projection in such scenarios in comparison to the method based on the universality of hot carrier degradation extended to the case featuring feedback [6].

1.2 Array-Based Hardware Monitors for Detection & Characterization of On-Chip Electromigration (EM) Effects

Electromigration (EM) is essentially the movement of the material constituting the interconnect, as a consequence of stress over a period of time. This chapter presents three test-vehicles: a resistance degradation sensing monitor, a bit-error-rate degradation monitor, and a third, with a combination of capabilities borrowed from the previous two, for capturing the impact of EM effects on the interconnect resistance and its signaling capability simultaneously over time under stress.

1.2.1 Characterizing Electromigration Effects in a 16nm FinFET Process Using a Circuit Based Test Vehicle

This work showcases measured data corresponding to direct-current (DC) stress induced Electromigration (EM) phenomenon, characterized using on-chip circuits for interconnect test structures fabricated in a 16nm FinFET process. An array-based test vehicle featuring parallel stress and 4-wire Kelvin sensing capabilities is presented, employing wires with distinct feature sizes and metal stacks as the Devices-Under-Test (DUTs). Accelerated stress testing is achieved using on-chip using metal heaters positioned directly above the DUTs, which provides precise local temperature regulation in conjunction with fast cycling between stress and measurement temperatures. Test chip data captures several EM effects ranging from abrupt and/or progressive failures depending on

DUT geometry, temporary healing effects, circuit-interconnect interplay, and process variation impacting EM lifetime [7].

1.2.2 Electromigration-Induced Bit-Error-Rate Degradation of Interconnect Signal Paths Characterized from a 16nm Test Chip

An array-based test-vehicle for tracking bit-error-rate (BER) degradation of signal interconnects subject to DC electromigration (EM) stress was implemented in a 16nm FinFET process. A unit interconnect path comprises five identical interconnect stages where each wire is driven by inverter based buffers. Accelerated EM stress testing is achieved entirely on-chip using metal heaters located directly above the Devices-Under-Test (DUTs) and separate stress circuits driving both ends of the wire. BER measurement results from four individual interconnect paths are presented and analyzed.

1.2.3 A 22nm Array-Based EM-Induced Bit-Error-Rate Degradation Monitor featuring DUT Resistance Sensing & Stress Capabilities with added flexibility in Mode-Control

This section builds upon the work presented in the former subsections. A new EM monitor designed a 22nm FDSOI process node is presented, featuring the combined capabilities of the test-vehicles described in the previous subsections, the primary motivation being the translation of a measured resistance shift to an associated degradation in the BER of an interconnect path. The vehicle features added flexibility in mode-control which allows for monitoring additional effects such as Joule heating at a target stress current or the local DUT temperature with the on-chip heaters activated. Furthermore, the five wires comprising a datapath can be selected independently for stress, allowing for a

further detailed investigation in the associated resistance and BER shift as a consequence of one or multiple wires degrading over time, for a given interconnect path.

1.3 A High-Density, Array-Based, Soft-Error-Rate & Single-Event-Transient Pulse-Width Characterization Macro meant for Standard Combinational Logic

This work presents neutron radiation induced Soft Error Rate (SER) statistics and the corresponding Single-Event-Transient (SET) pulse-width distributions, together with a detailed analysis thereof, revealing a multitude of circuit parameters impacting the soft-error susceptibility of standard combinational logic in advanced CMOS nodes. A high density, array-based, soft-error characterization vehicle is presented, featuring standard logic gate chains of varying lengths. Neutron irradiation data obtained from gate variants employing devices with distinct channel widths and threshold voltage flavors is analyzed at multiple supply voltages, ranging from nominal down to near-threshold. Supplemented with first-order simulations, measured results obtained from test structures implemented in a 65nm planar CMOS technology node reveal the complex interplay between factors such as supply voltage, node capacitance, restore current ($I_{RESTORE}$), gate topology and logic chain length responsible in contributing towards the collective soft error susceptibility of a standard gate type, which constitutes the main focus of this work. In addition, the easy process portability of the proposed macro is demonstrated through implementation in a 16nm FinFET process [8].

Chapter 2. An Array-Based Characterization Vehicle for Investigating the Impact of Feedback on the Aging Dynamics of Diode-connected MOSFETs

2.1 Introduction

Most Analog / Mixed-Signal systems typically leverage feedback of some sort to ensure output precision in the face of changing circuit or load conditions. Of late, there has been growing interest towards understanding the impact of feedback on the aging mechanics in these circuits [9]. A popular method for modeling aging induced parameter shifts thus far has been the constant power law model t^n, where t is the stress time and n a constant. However, recent studies have shown that aging models such as the constant power law model may need to be augmented with a time-varying stress condition for circuits employing feedback [9], [10], [11] [12]. One such example of a circuit exhibiting accelerated aging due to time-varying stress is the ubiquitous diode-connected MOSFET. Fig. 2.1 illustrates the aging of this circuit under the influence of constant current stress. The degradation, which is predominantly due to Hot Carrier Injection (HCI), causes the V_{TH} and hence, the drain current (I_D) of the device to degrade overtime. However, owing to the presence of feedback in the system, the diode voltage (V_{DIODE}) must increase correspondingly to maintain the same initial stress current across the device. This leads to a time varying stress condition across the transistor, aggravating the HCI aging and in turn causing the device parameters to degrade rapidly over time.

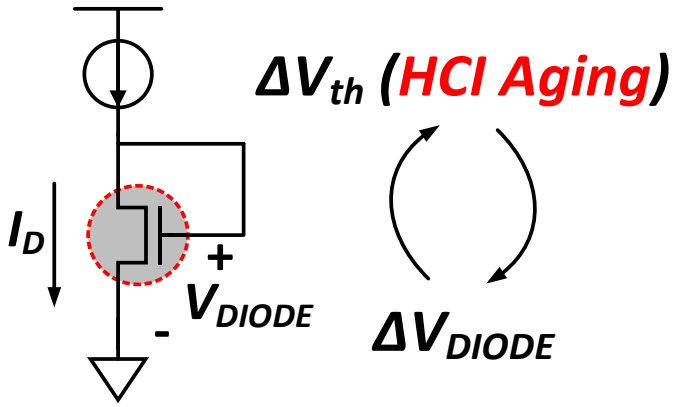
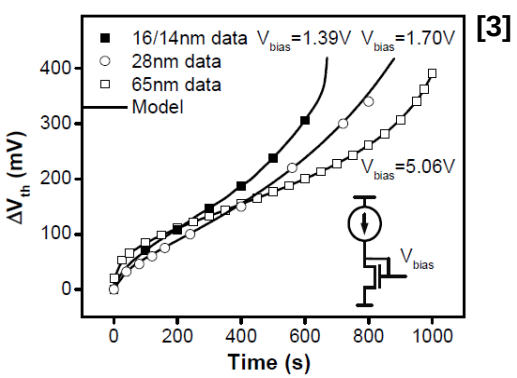


Figure 2.1: Diode-connected MOS under constant current stress resulting in accelerated V_{TH} shift due to increasing diode voltage with time

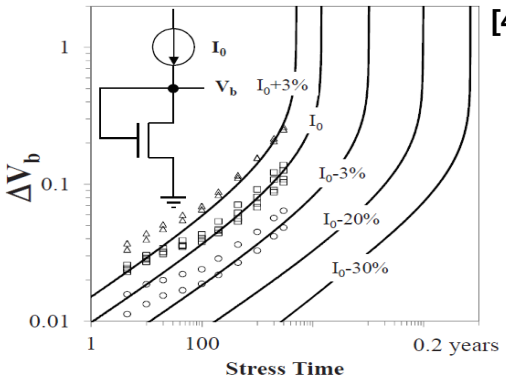
Prior research has proposed iterative simulation frameworks that account for the change in the stress condition over time [9], [10], [11]. However, previous efforts have several drawbacks such as: (i) lack of comprehensive experimental data from a real test-chip and (ii) prediction results suggesting runaway behavior which could not be experimentally verified due to HCI saturation (Fig. 2.2). These limitations can lead to gross overestimations of aging for high reliability applications and hence, improper design margins when extrapolating down to usage conditions.

To this end, we present a 1-dimensional array-based test-structure, implemented in a 1.2V 65nm Low Power (LP) process, featuring diode-connected MOS transistors as the Devices-Under-Test (DUTs) and offering the flexibility to investigate the aging behavior for a particular DUT type in either absence or presence of feedback. Our aim is twofold: (1) experimentally characterize the impact of feedback on the aging rate and compare it to

the no-feedback case & (2) evaluate the efficacy of iterative simulation frameworks for lifetime projection in such scenarios.



[3]



[4]

Figure 2.2: Prior Art [10], [11]: Lack of comprehensive test-data does not clearly specify the extent of model fit and owing to the temporal universality of HCI degradation [6] requires long-term stress results at lower bias voltages to actually verify if only gradual aging without the possibility of runaway [10] occurs at voltages lower than the critical voltage detailed in [10].

2.2 Test Structure & Measurement Methodology

2.2.1 Proposed Array-Based Test Vehicle

Fig. 2.3 illustrates the schematic of the unit stress cell with the DUT housed within. The array consists of a single row of such stress cells with dedicated scan bits per cell to either independently enable stress for one or multiple cells or enable sensing of the drain and/or the source voltages for a particular DUT of interest.

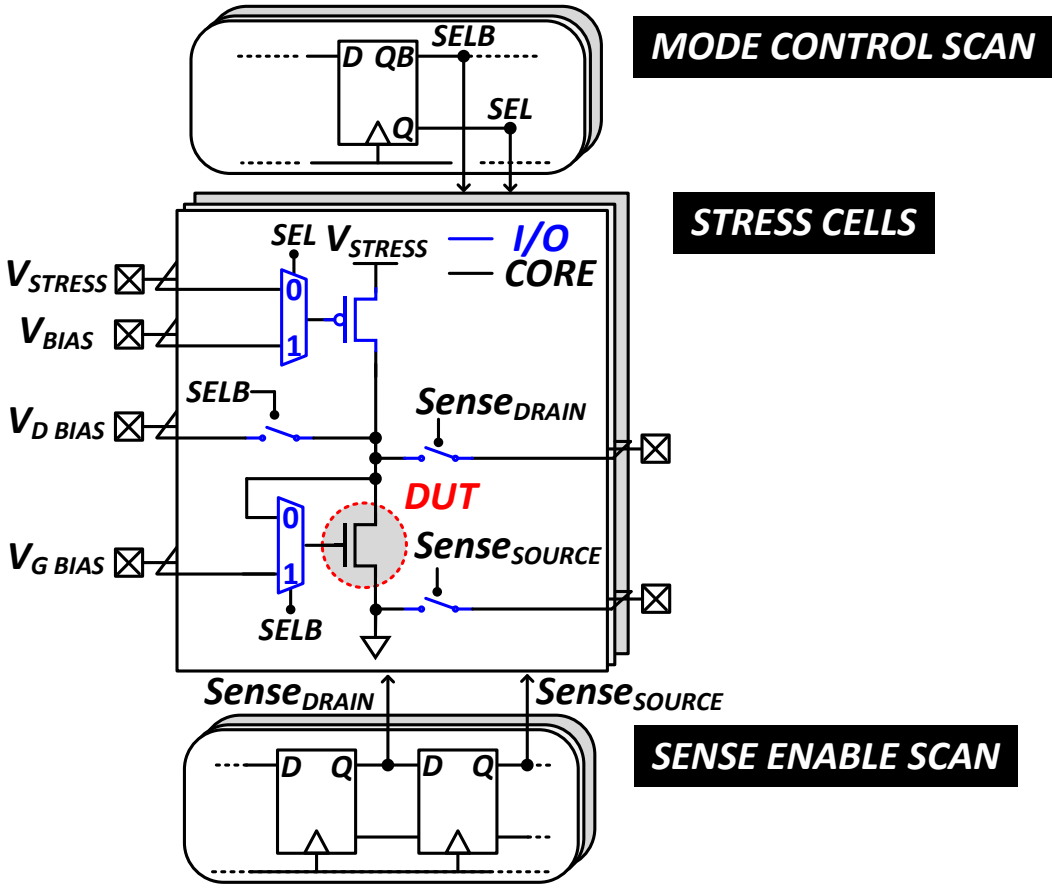


Figure 2.3: Overview of the proposed test-structure detailing the unit-cell.

Each stress cell contains a wide (W/L=30µm/1.2µm) thick oxide (3.3V) PMOS header which can be either (i) turned on fully to enable DUT selection when using the source

measure unit (SMU) for applying a given stress current or voltage, or (ii) used as a near ideal on-chip current source with an elevated headroom and fixed bias to mimic a real test case scenario. Control signals to the stress cells come from the scan chains followed by level shifters for up-converting to the stress voltage. Three shared analog inputs V_{BIAS}, $V_{D\ BIAS}$ and $V_{G\ BIAS}$ are used for biasing the stress cells whereas V_{STRESS} works as the power supply of the entire array. The stress cells can be programmed using the mode control scan chain to keep multiple devices under one of 'STRESS' (SEL = 1) or 'IDLE' (SEL = 0) modes as illustrated in Fig. 2.4.

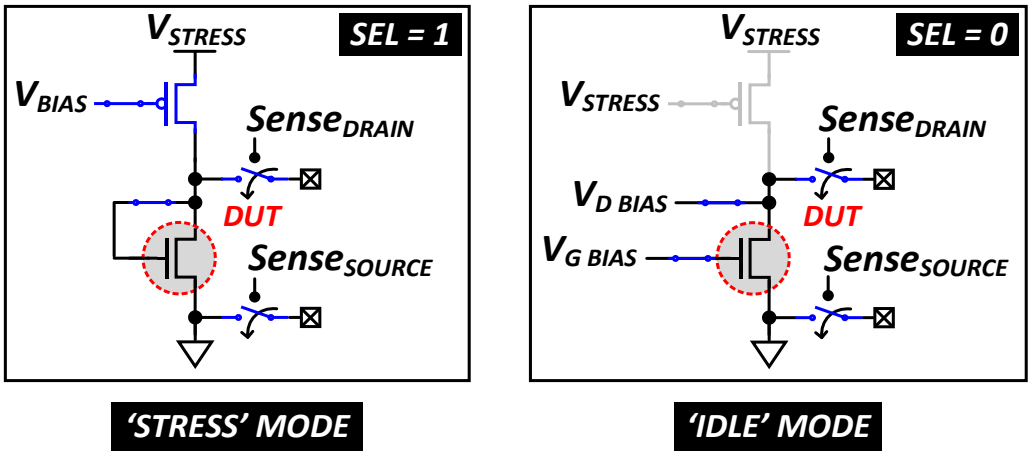
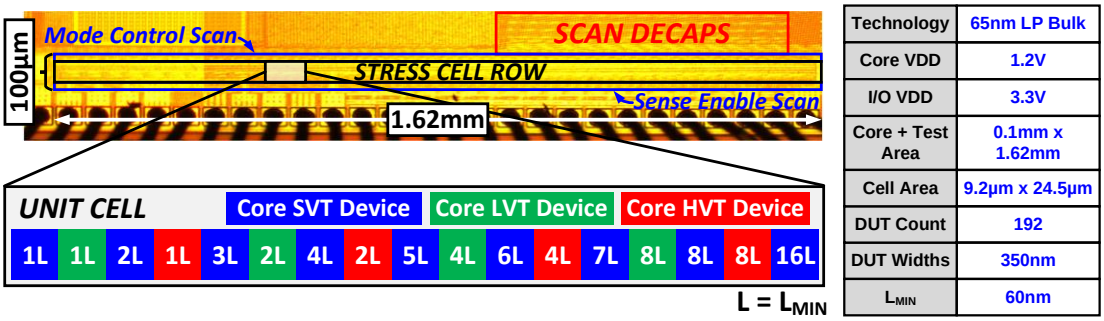


Figure 2.4: Illustration of the different operation modes for the proposed stress cell.

A total of 17 different DUT types were implemented in a single unit block which was then instantiated over to form the 1-dimensional array (Fig. 2.5). Fig. 2.5 also shows the die microphotograph of the 65nm test chip along with a description of the different DUT types implemented in this work.



Technology	65nm LP Bulk
Core VDD	1.2V
I/O VDD	3.3V
Core + Test Area	0.1mm x 1.62mm
Cell Area	9.2µm x 24.5µm
DUT Count	192
DUT Widths	350nm
L_{MIN}	60nm

Figure 2.5 Die micrograph & implementation summary along with the distinct DUT types housed within a tileable unit.

2.2.2 Measurement Methodology

Fig. 2.6 A illustrates the equivalent circuit configuration for our constant current stress experiments. The PMOS header is turned fully on, the target stress current is sourced using the SMU and the DUT is configured as a diode with the drain and source sensing enabled. Fig. 2.6 B shows the corresponding stress-measurement cycles. Following a time zero measurement, in a cycle of 60 sec, the first 58 sec are used for subjecting the DUT to

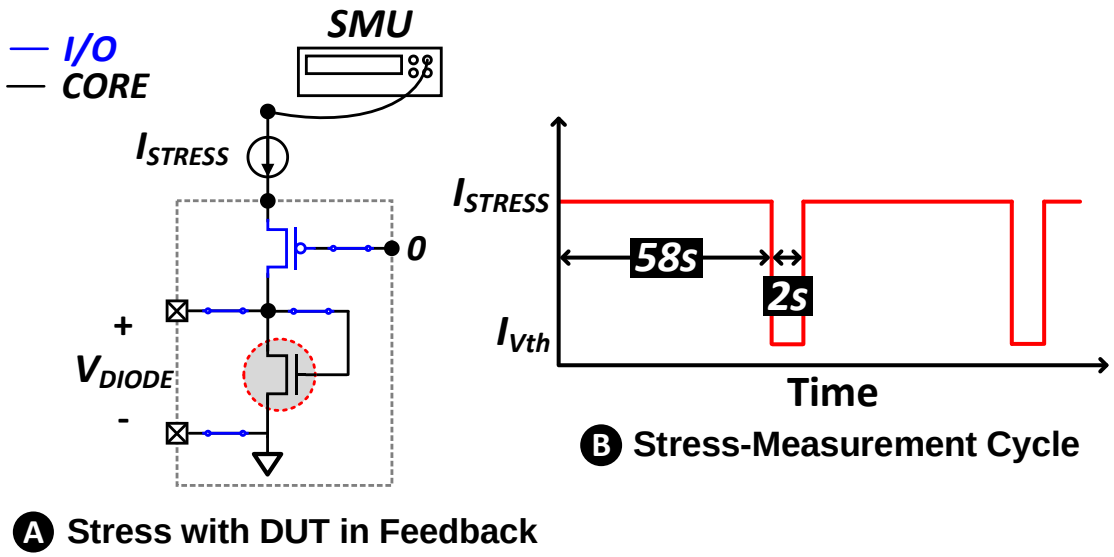


Figure 2.6: (a) Illustration of the DUT subjected to a constant-current / feedback stress & (b) the corresponding stress-measurement cycle.

a constant current stress, with the drain and source voltages recorded every 2 sec and the last 2 sec used to switch the SMU to a lower current for sampling the threshold voltage using the method of constant current.

For constant voltage stress experiments, a similar methodology can be adopted wherein the PMOS header is kept fully on and the desired stress voltage is sourced using the SMU. However, this leads to a complication: as the DUT ages overtime, the decrease in DUT current leads to a reduced drop across the PMOS device creating instead an accelerating stress condition like the constant current case. To ensure constant voltage at all times, a feedback loop is implemented in software (Fig. 2.7) which monitors the sensed V_{DIODE} every few seconds and compares its value with the target stress voltage to adjust the SMU output voltage accordingly. Fig. 2.8 shows the measured V_{DIODE} results with and without the correction mechanism implemented. Less than 0.5% (10mV) of variation in the sensed value of V_{DIODE} for 2.4V of stress was observed post-correction.

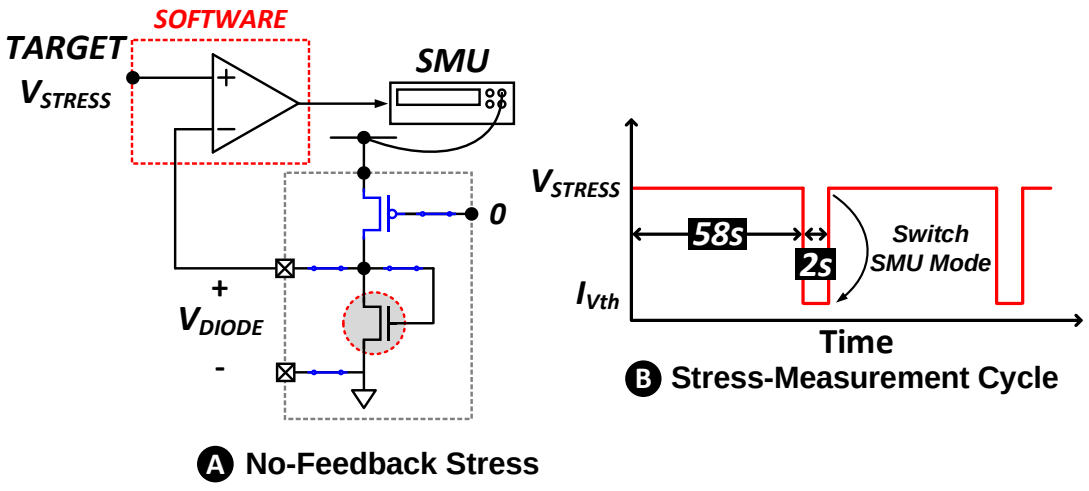


Figure 2.7 (a) Illustration of the DUT subjected to a constant-voltage / no-feedback stress & (b) the corresponding stress-measurement cycle.

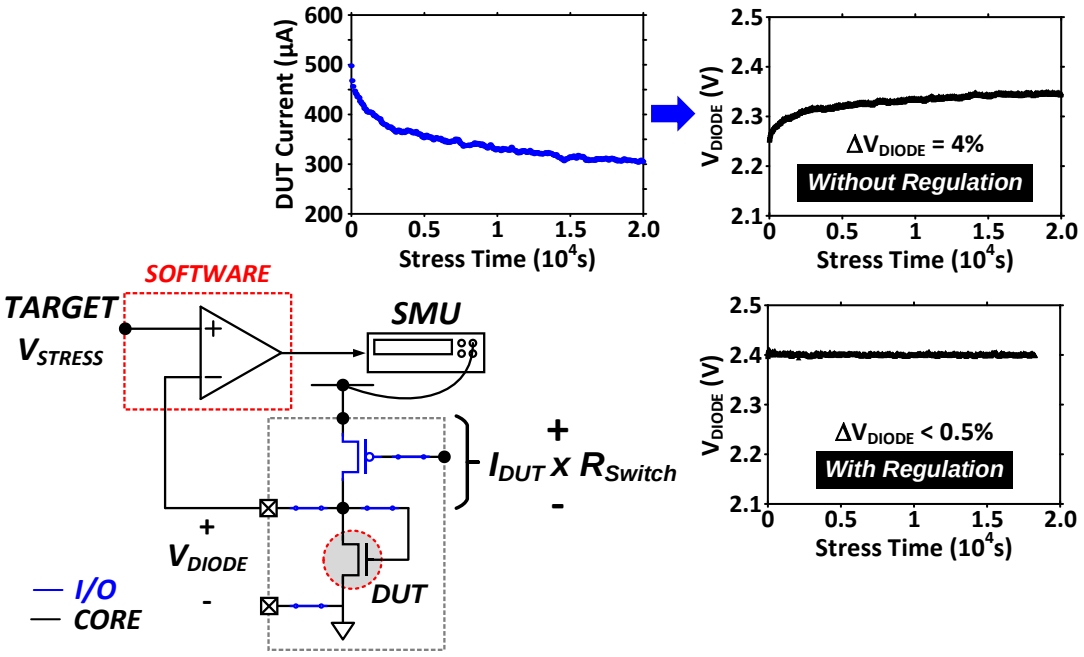


Figure 2.8: (a) Measured V_{DIODE} with and without the correction loop implemented in software for the constant-voltage stress case.

2.3 Measurement Results and Aging Prediction

2.3.1 Measured V_{DIODE} and V_{TH}

Fig. 2.9 compares the evolution of V_{DIODE} and V_{TH} for the constant current and constant voltage stress experiments. Although the results shown are for the case of minimum channel length High V_{TH} (HVT) DUTs, similar trends were observed for the Standard V_{TH} (SVT) and Low V_{TH} (LVT) DUTs as shown in Fig. 2.10. As can be seen from Fig. 2.9, the aging is more pronounced for the constant current case due to the increasing stress voltage overtime. Looking at the degradation traces for both the stress conditions, it can be concluded that: (1) the degradation, which is primarily due to HCI, saturates over time and, (2) the onset of saturation depends on the initial value of the stress bias. These observations are in alignment with the bond-dispersion model [13] and point to the universality of the degradation [14], presented comprehensively in [15]. Bottommost figures in either column in Fig. 2.9 show the measured fresh and post-stress diode I-V curves and are reflective of the high V_{TH} degradation in presence of feedback.

Once the aging rate saturates, the device (for constant current stress) continues to age at a slower rate until oxide breakdown after which the current drawn by the gate causes a drop across the transmission gate connected between the drain and gate. This is evinced by both the jump in V_{DIODE} (needed to force the same current on account of reduction in the gate bias due to IR drop in the transmission gate) and drop in the measured V_{TH} (owing to the parallel current path drawing a portion of the small current needed for measuring V_{TH}). For constant voltage case, the range of stress values applied in our experiments (<= 2 x Nominal VDD) were not sufficient to cause an oxide breakdown (nearly 3.3V, as can

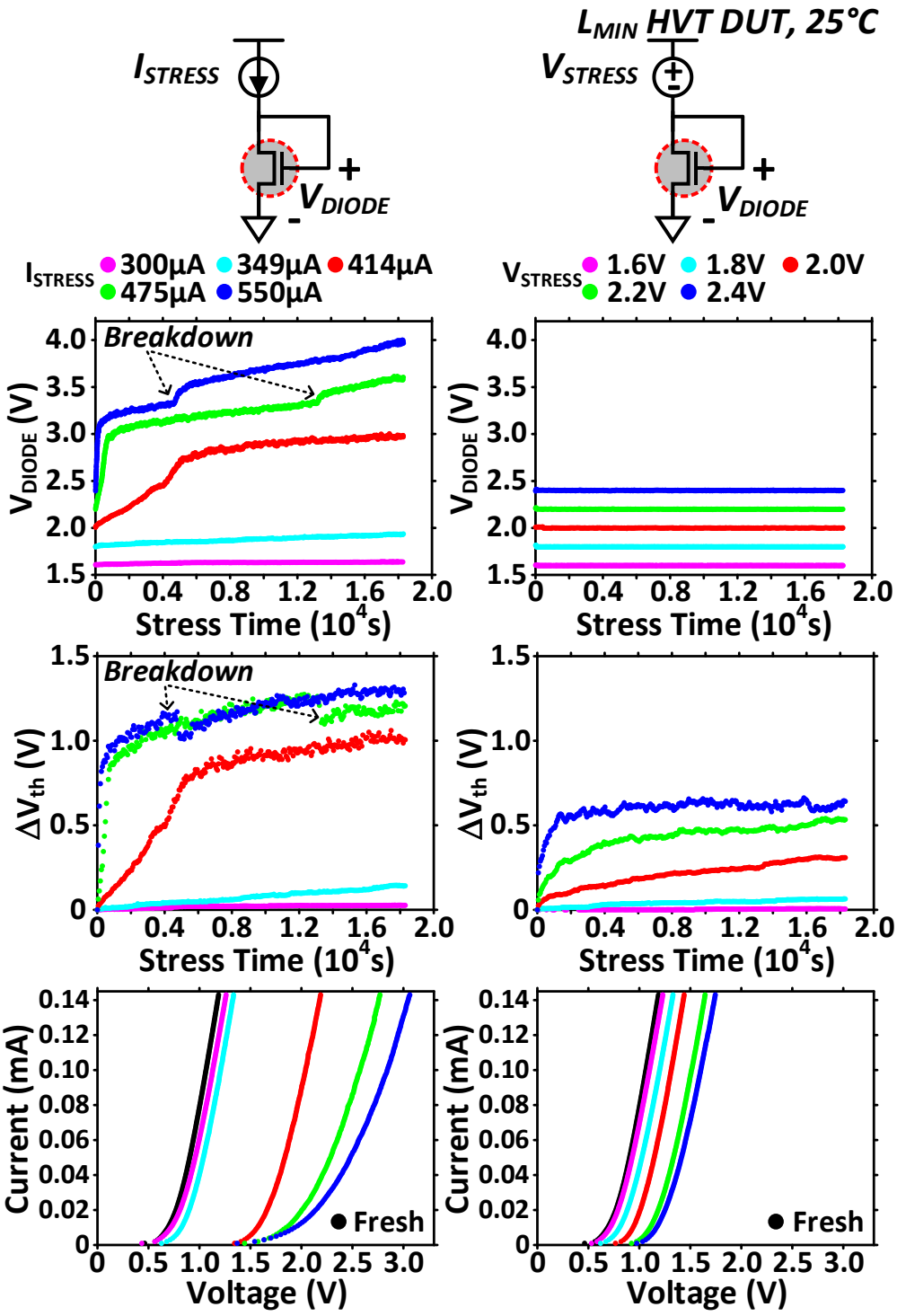


Figure 2.9: Shift in V_{DIODE} and V_{TH} for constant current stress (left column) and constant voltage stress (right column) for identical values of starting stress biases. V_{TH} was monitored by applying a small fixed bias current and measuring the diode voltage. Bottommost figure shows the fresh & post-stress diode I-V curves.

be seen from the constant current case). Longer channel length devices showed substantially higher resilience to HCI degradation even in presence of feedback (Fig. 2.10).

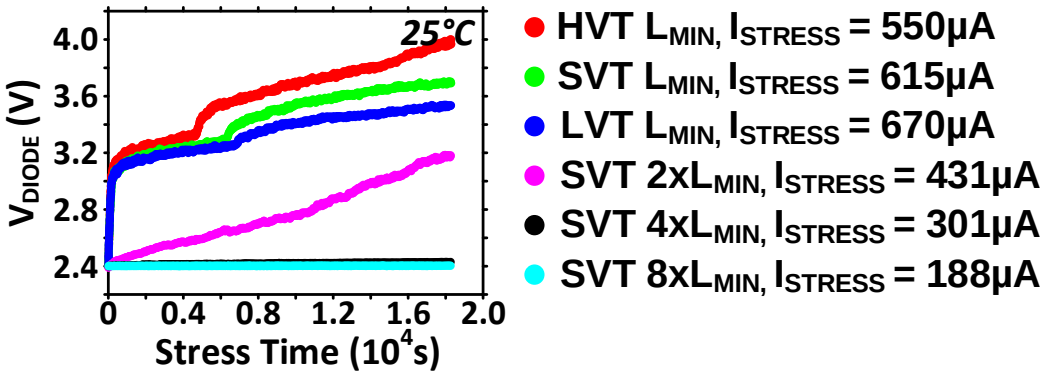


Figure 2.11: Measured V_{DIODE} for core DUTs with different V_{TH} (LVT, SVT, HVT) and channel lengths (L, 2L, 4L, 8L) under constant current stress for identical values of starting stress bias (= 2.4V).

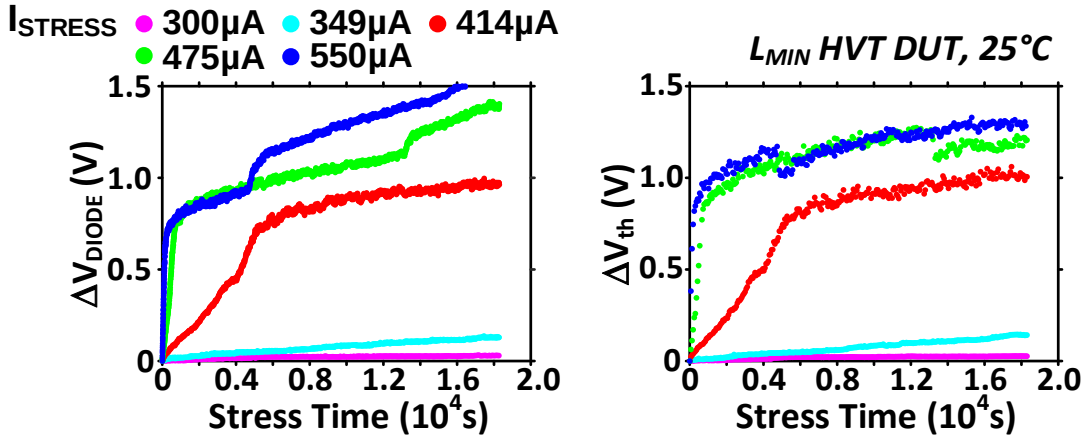


Figure 2.10: Measured V_{DIODE} for core DUTs with different V_{TH} (LVT, SVT, HVT) and channel lengths (L, 2L, 4L, 8L) under constant current stress for identical values of starting stress bias (= 2.4V).

Fig. 2.11 shows the measured shift in diode voltage for the constant current case plotted alongside the shift in V_{TH}. As can be observed from Fig. 2.11, the shift in V_{DIODE} closely follows the shift in V_{TH}. The insight from this result was used to update V_{DIODE} at

each simulation time step for the iterative version of the model used for the constant current stress (discussed in the next subsection).

2.3.2 Modeling the Constant Voltage & Constant Current induced degradation

Fig. 2.12 shows the results of using the simple constant power law model [16] for modeling the constant voltage induced V_{TH} degradation (left) at two different stress biases. The model provides a good fit to the experimental data at lower stress voltages and/or for shorter stress durations but does not predict the onset of saturation occurring for longer stress times or at higher stress biases [14]. When an iterative methodology (similar to [11]) is employed to update the stress condition for the constant current case, the model correctly predicts the initial super-exponential trend but fails to account for the onset of saturation, thus resulting in a limited fit depending upon the applied stress bias (more at lower voltages or shorter stress durations and less at higher stress biases or longer durations) This is particularly concerning for the constant current (feedback) case where starting from a higher stress bias can cause a high initial aging rate followed quickly by the onset of saturation after which the device continues to age at a much slower rate (Fig. 2.12, Right). Hence, an alternative method is needed for modeling the entire degradation characteristics including saturation. This would be especially useful for high reliability applications with longer operational lifetimes. Fig. 2.13 encapsulates the details of the iterative flow (implemented in MATLAB) used for the constant current stress case.

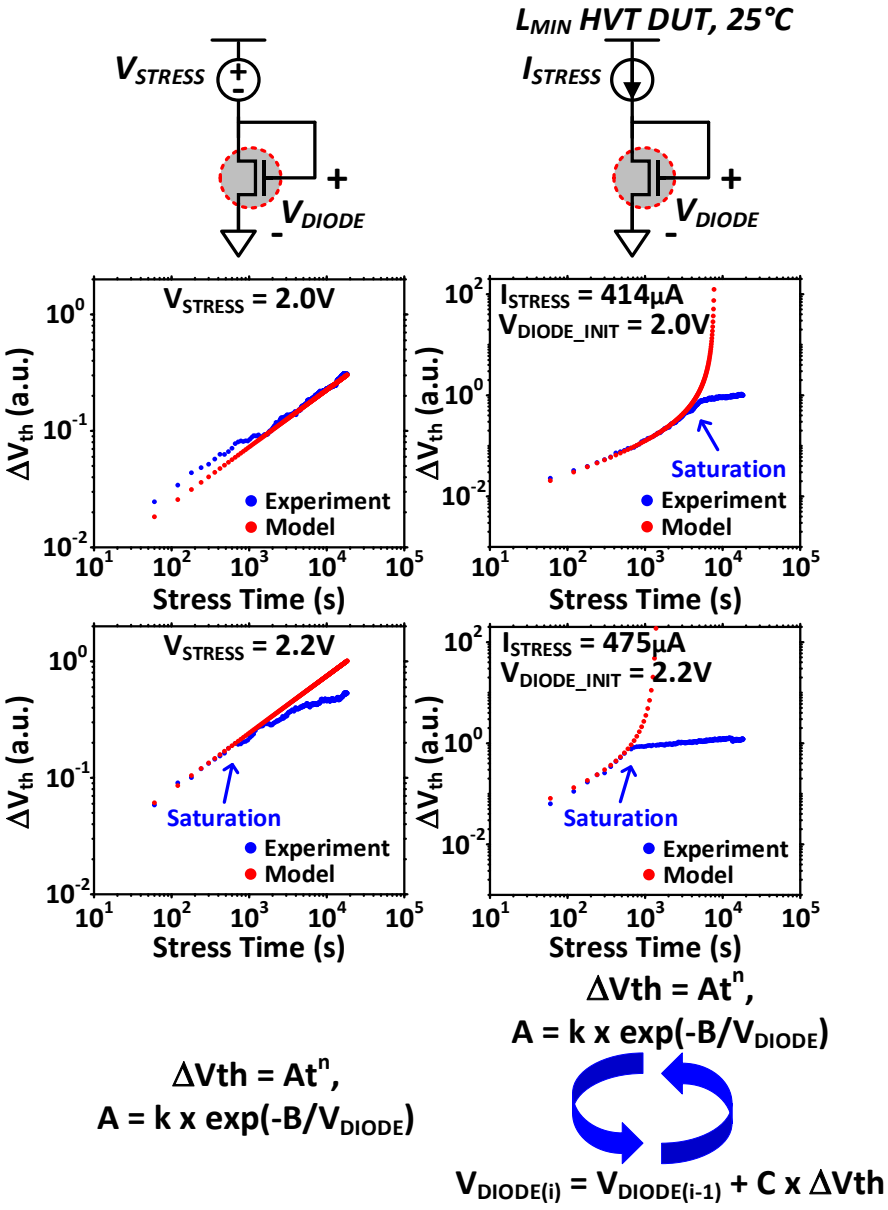


Figure 2.12: Using the constant power law model [8] in conjunction with an iterative approach leads to an accurate fit with the constant current stress data as shown in the figure. One limitation of this method is that it cannot incorporate the onset of saturation. This figure shows measurement data from Fig. 5 plotted on log scale along with the modeled data. Where n = 0.49, B = 26.5, k = 1400 & C = 0.93 (close to 1, as expected from Fig. 6). Constants derived from curve fitting the constant voltage stress results were used as it is for the constant current case with the stress condition updated iteratively (similar to [11]) at each simulation time-step (60 sec, similar to our measurement).

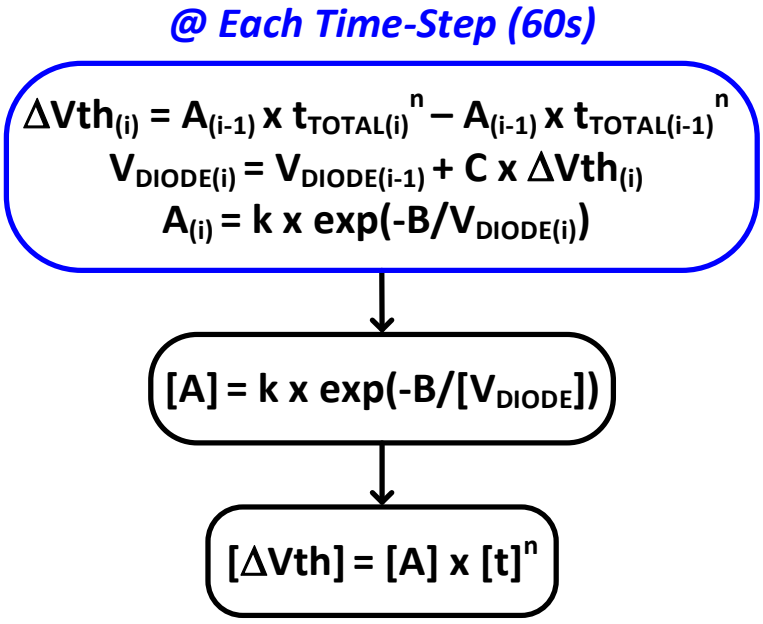


Figure 2.13: Iterative flow detail for the constant current model in Fig. 2.12 (Right) implemented in MATLAB, A_0 = A (Constant voltage case).

2.4 Application of the method based on the universality of hot carrier degradation to the diode-connected FET

In this section we explore the usefulness of the temporal universality of HCI degradation [14], also referred to as lateral scaling, in the context of circuits with varying stress conditions. Fig. 2.14 shows the original as well as the time-scaled versions of the V_{TH} degradation curves corresponding to different values of stress biases for the constant voltage and the constant current stress experiments. Upon time scaling the measured results on the log scale, we can clearly observe the universality of degradation (i.e. the time-scaled versions all lie on different parts of a common curve) holding for both the constant voltage (a known result, [14]) and the constant current (feedback) stress cases.

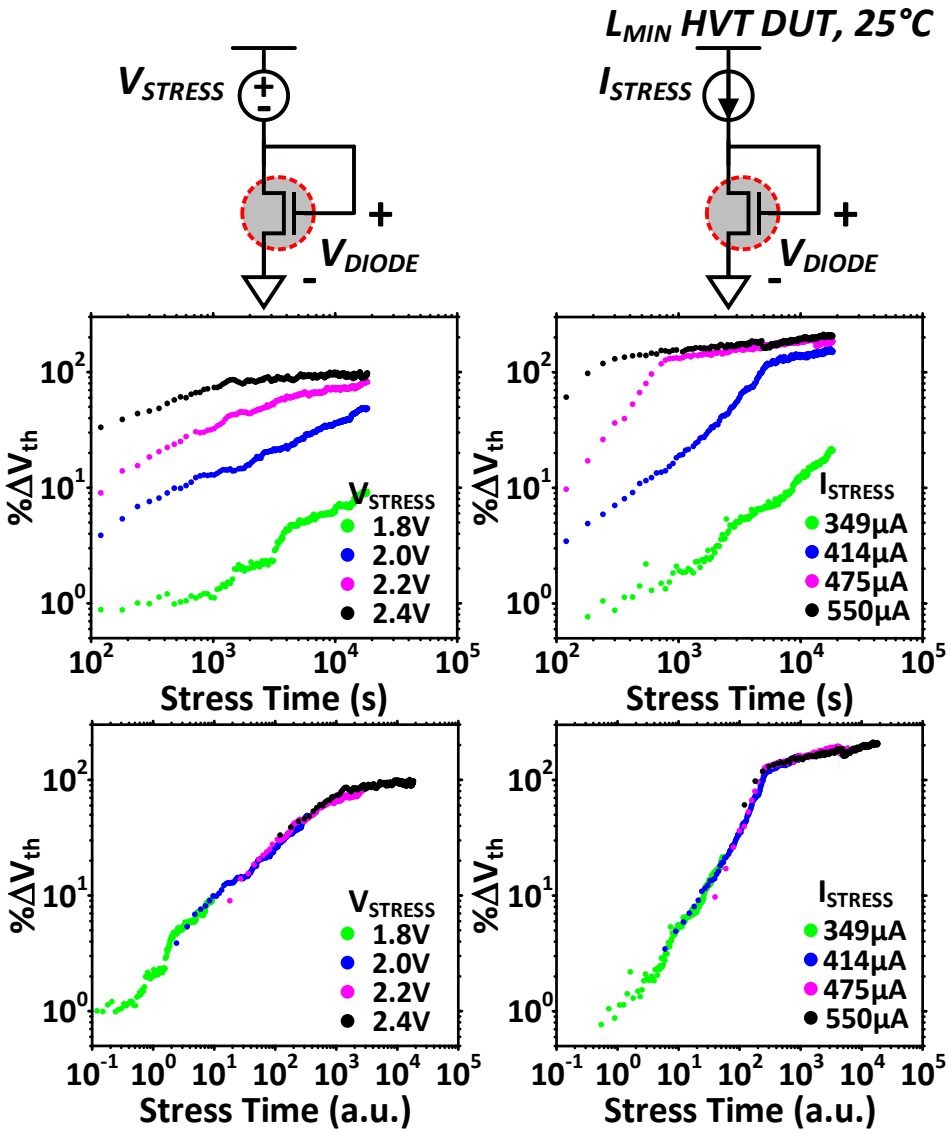


Figure 2.14: Time Scaling the test-data results in universal curves [14] for both constant voltage and constant current (feedback) stress cases. This figure shows the original (top) and time-scaled (bottom) versions of measurement results from Fig. 2.9 plotted on a log scale.

Fig. 2.15 then illustrates the application of this method to the feedback case. Following a few short-term measurements at different stress currents, the results are laterally scaled along the time axis. The projection for scaling factors [17] is then derived from the measured values of 'initial' stress biases (V_{DIODE_INIT}), corresponding to the

different values of stress currents and the scale factors obtained in the previous step. Using this projection, accurate lifetime prediction including the onset of saturation becomes possible for a diode usage voltage (for instance 1.0V, as illustrated in Fig. 2.15 (d)).

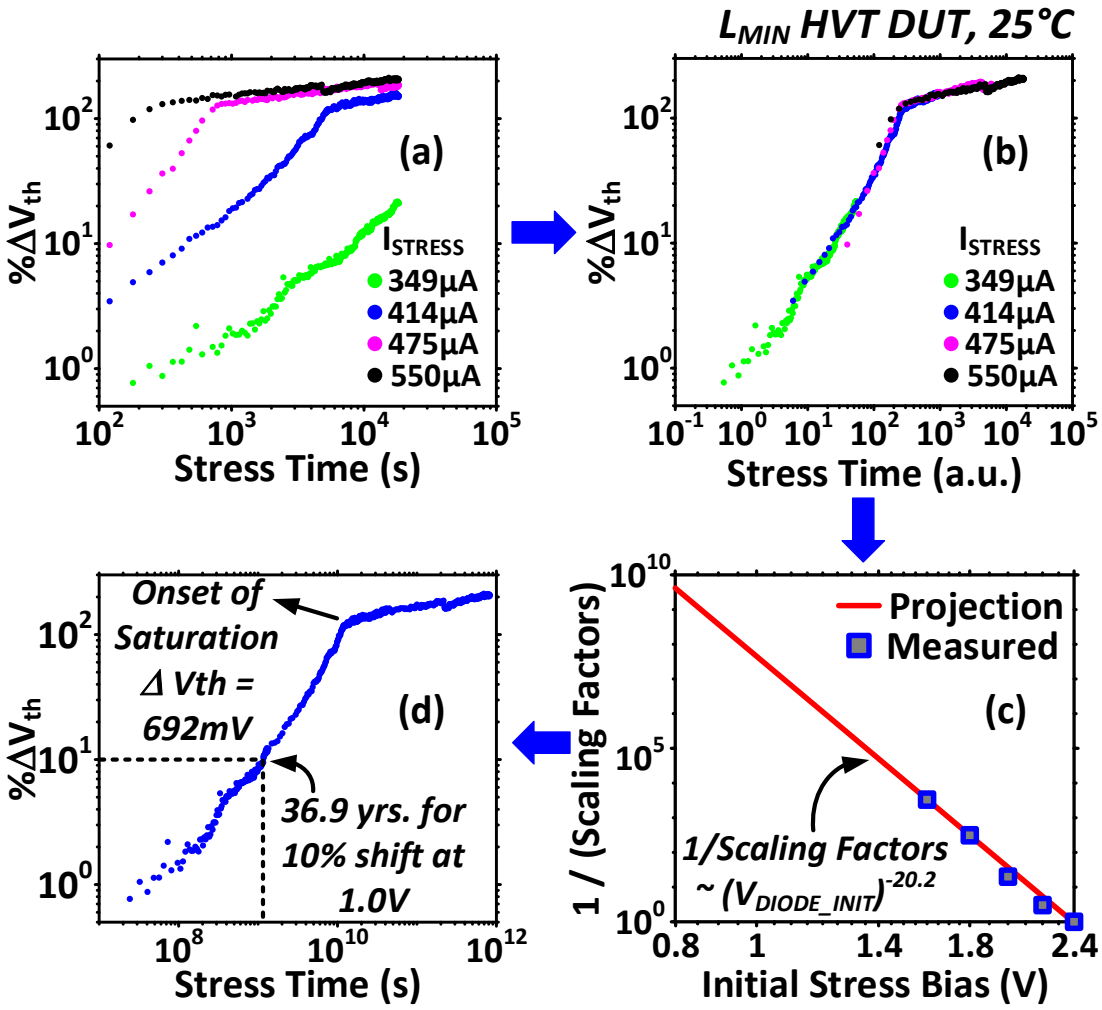


Figure 2.15: Application of the method based on universality of degradation [6, 9] to the diode-connected DUT: (a) Short-term stress measurements under constant current stress; (b) Universal curve obtained by time scaling the measurement results in (a) ; (c) Voltage acceleration for scaling factors [17] obtained using the 'initial' values of diode stress biases (V_{DIODE_INIT}) developed corresponding to stress currents in (b) / Fig. 2.9 (Left Column). (d) Aging prediction at a usage voltage (=1.0V) including the onset of saturation.

It is also possible to extract the scaling factors in terms of the stress current (I_{STRESS}). Fig. 2.16 graphically illustrates this.

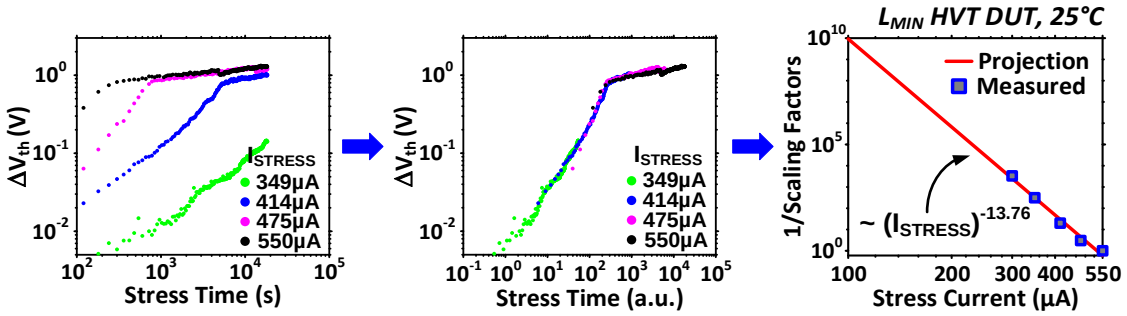


Figure 2.16: In certain applications, bias current extrapolation may be more relevant rather than the diode voltage. Universality can also be directly employed to stress current in such scenarios.

Finally, Fig. 2.17 & 2.18 present the real-time long-term stress data for lower initial stress biases such as 1.6V (300uA) and 1.8V (349uA) showing the onset of saturation for longer stress times and time scaled versions of the same, (using the scaling factors derived from the short-term measurement data / projection) fitting well to the universal curve. On account of the DUTs experiencing a common range of stress voltages (V_{DIODE}) for the constant current case, there is a greater overlap between the individual degradation traces in presence of feedback. Hence, even fewer short-term measurements (which are also quick, on account of accelerated aging) are now needed to construct the full universal curve. It should also be noted from these figures that the V_{TH} undergoes the same total shift overtime starting at a lower bias (owing to the universality of degradation) and thus the bias itself would undergo the same shift as for a higher initial value albeit in a longer time. Hence, the risk of a significant deviation overtime in the initially applied value of bias still remains even for lower initial biases.

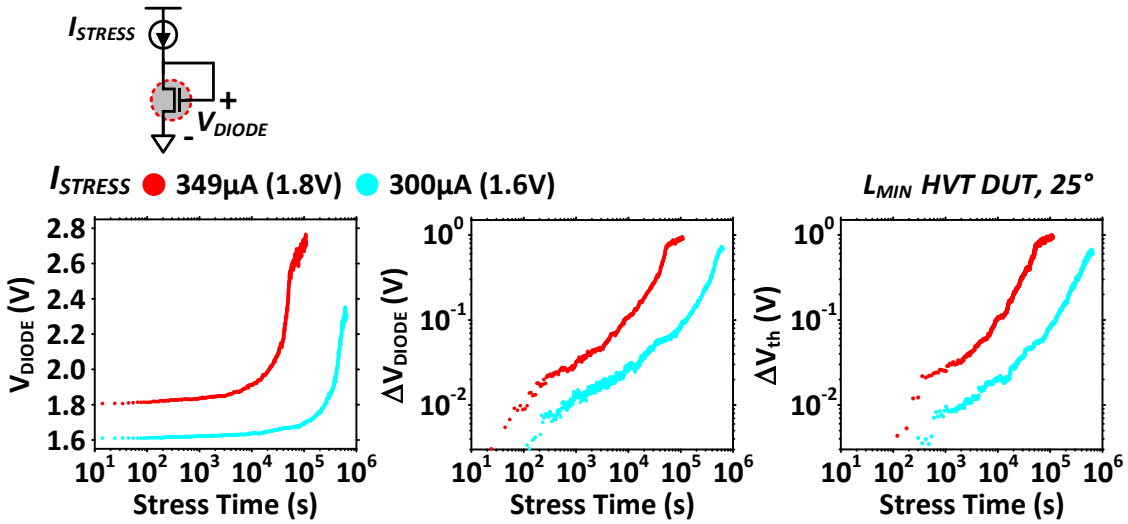


Figure 2.17: Measured V_{DIODE}, ΔV_{DIODE} & ΔV_{TH} at lower initial stress biases viz. 1.6V & 1.8V.

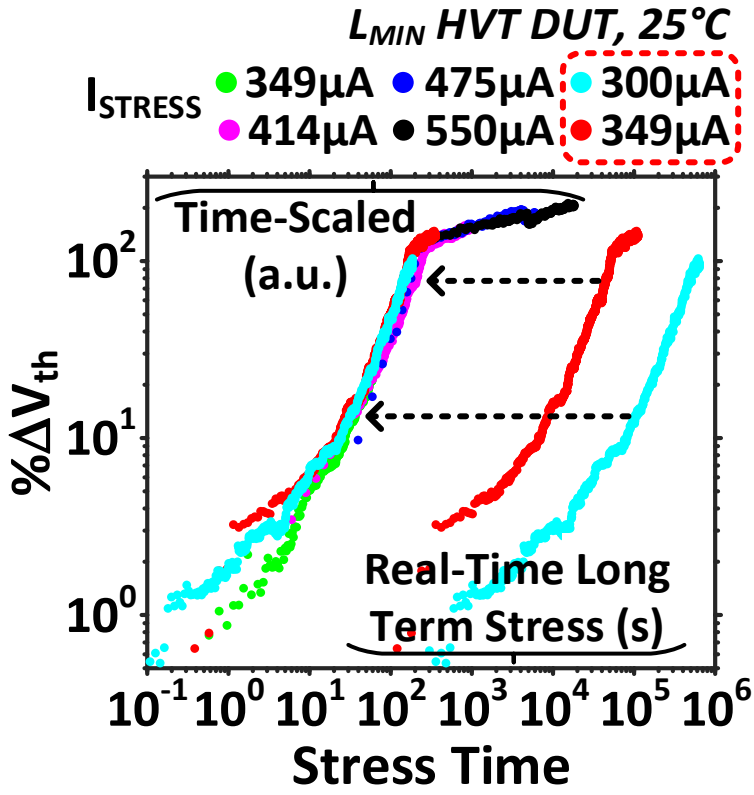


Figure 2.18: Real-time long-term stress data at diode biases of 1.6V (cyan) and 1.8V (red) fits well using the projected scaling factors derived in Fig. 2.15. Then, using the inverse of scale factors and unrolling the Universal Curve in time provides an accurate estimate of the aging dynamics and hence, shift in the bias for an initial value of stress bias applied to the diode.

2.5 Summary

This chapter presented comprehensive test-data showcasing the detailed impact of feedback on aging for a diode-connected MOSFET. We presented an array-based characterization vehicle intended specifically to quantify the impact of feedback on aging. The efficacy of iteratively updating the stress condition for lifetime prediction in such scenarios was also evaluated and it was observed that simple models that don't account for saturation provide a limited fit (depending on the stress level or the stress duration) and predict non-realistic super exponential rates thereafter. Finally, we discussed the application of the method based on the universality of hot carrier degradation to the feedback stress case. Owing to greater overlap between the individual degradation curves coupled with accelerated aging in presence of feedback, only a few short-term measurements at different stress currents are sufficient to construct a universal curve that can accurately predict the aging dynamics including saturation given the initial stress bias or stress current of the diode-connected device.

Chapter 3. Array-Based Hardware Monitors for Detection & Characterization of On-Chip Electromigration (EM) Effects

3.1 Characterizing Electromigration Effects in a 16nm FinFET Process Using a Circuit Based Test Vehicle

3.1.1 Introduction

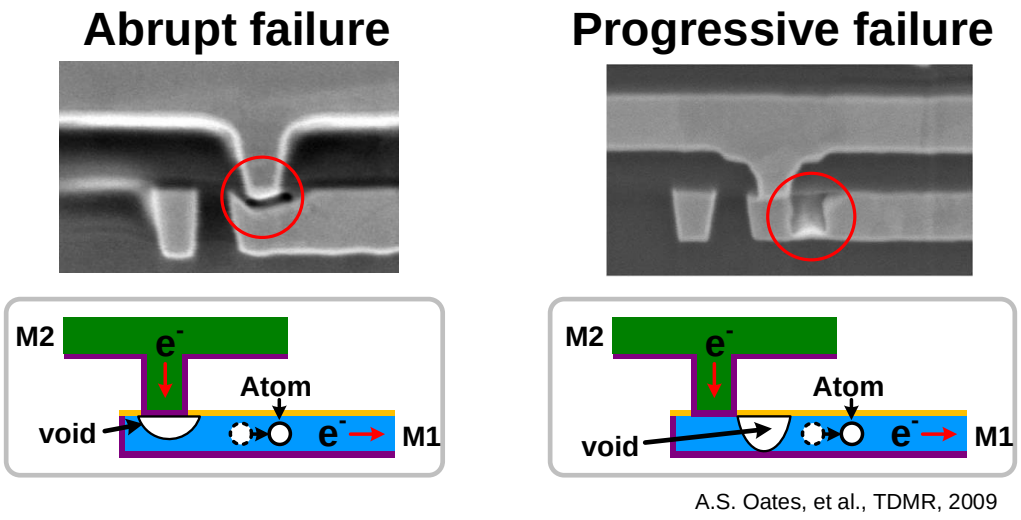


Figure 3.1: Electromigration in interconnects resulting in unique void morphologies associated with distinct resistance shift signatures [18], [19], [20].

Electromigration is essentially the displacement of the material constituting the interconnect, as a consequence of stress overtime. The associated material depletion results in distinct void morphologies which may either cause an abrupt or a gradual change in the resistance of the conductor overtime (Fig. 3.1). The shift in resistance, whether gradual or abrupt, is a strong function of temperature, the stress current density, the direction of

applied stress, as well as mechanical forces operating within the wire. For VLSI circuits, this may translate to a reduced supply voltage across circuits in a power-grid or an increased delay for signaling interconnections. Traditionally, EM effects were characterized by monitoring the resistance of individual wires under stress. Since EM lifetime is a stronger function of temperature than voltage or current, for accelerated testing, the stress temperature must be raised to above 300°C using an extensive oven-based test setup. Furthermore, to avoid statistical artifacts, EM data must be collected from a large sample population of DUTs, which is time consuming due to the limited number of wires that can be stressed together in existing setups. In addition to this, the available test area on the shuttle is largely limited by the area of pads, which imposes an upper limit on the number of test-structures feasible on-die.

To overcome the aforementioned testing challenges of EM, this work presents the first circuit based EM characterization test vehicle in a FinFET technology, using on-chip circuits to stress interconnects with different geometries. Our proposed approach utilizes an on-chip heater to establish and maintain a stress temperature locally, using the Temperature-Coefficient-Of-Resistance (TCR) of the metal. A fully automated test flow stresses multiple DUTs parallelly using an on-chip heater while also periodically keeping track of their resistances using only a few shared pads. The presented test vehicle offers a practical on-chip EM characterization approach, enabling efficient data collection from a large population of DUTs and providing deeper insights into underlying effects.

3.1.2 Proposed EM Characterization Macro

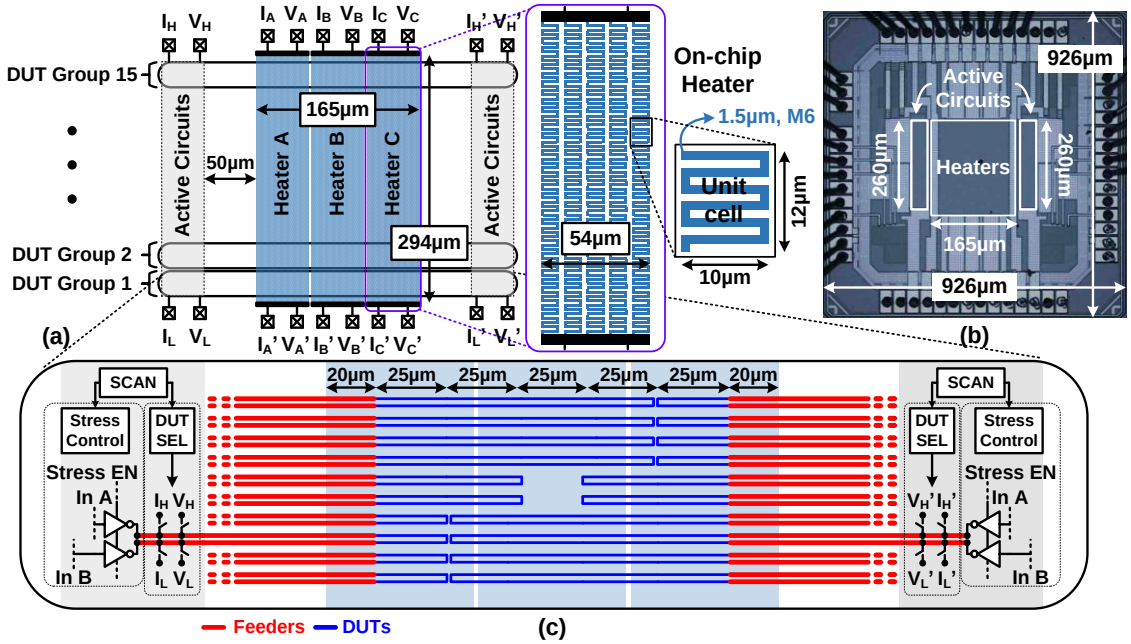


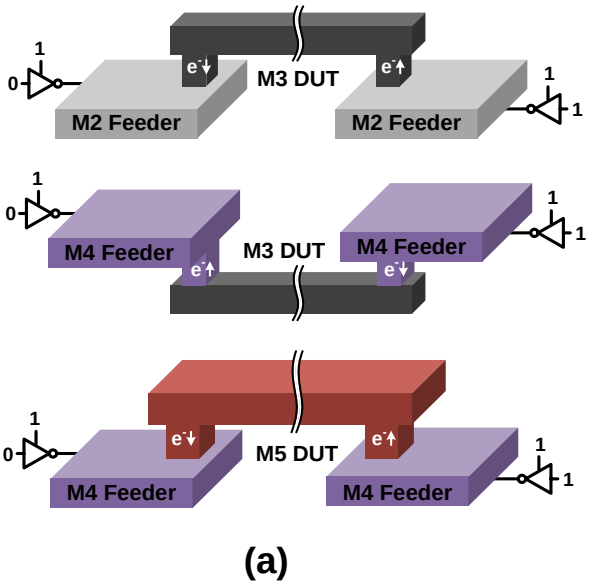
Figure 3.2: (a) Overview of the 16nm EM test-vehicle and on-chip heater specifics, (b) Die-micrograph and (c) Unit Tile-able DUT group details.

Fig. 3.2 presents an overview of the test-vehicle design and chip implementation specifics. The overall architecture follows an array-based scheme, with the unit tile-able block (Fig. 3.2(c)) housing the DUTs together with the active (stress, measurement and mode-control) circuits comprising the left and the right arrays. The heating area is centrally positioned, with the on-chip heaters employing wide Metal-6 (M6) tracks implemented in a snake-like fashion and uniformly covering the entire area underneath to ensure similar heat distribution throughout. The local die temperature was measured using the on-chip heater structures through 4-terminal Kelvin sensing. The DUTs are routed underneath the heating area using wide feeder tracks to minimize IR drop (Fig. 3.3) and are folded to maximize the area utilization. The active circuits on either side are placed far away from the central heating area to avoid leakage issues during measurement as well as to prevent

damage to the core circuits during stress. The DUT resistance measurement circuitry comprises 4-wire sensing switches per test structure, implemented using I/O based transmission gates as switches, (owing to leakage concerns) which share common output pads on either side: I-High (I_H), I-Low (I_L), V-High (V_H), V-Low (V_L) for the left array, and I'_H, I'_L,V'_H,V'_L for the right array. A one-hot encoded scan sequence is used for selecting the DUTs serially during the measurement window. The stress circuit comprises wide tri-state drivers, enabled using control signals from the scan-chain, with the inputs at either end driven to '1/0' and '0/1', respectively (Fig. 3.3) to result in a constant unidirectional current flow across the DUT during stress mode. A total of 15 DUT groups, each consisting of 10 wires on either side (total 300 DUTs) are tiled to cover the entire area underneath the heaters. Local stress generation and control provides the flexibility to stress one or several DUT groups simultaneously. The scan chain is implemented using D flip flops employing staggered two-phase clocking to prevent hold time issues during operation at elevated temperatures.

Fig. 3.3 illustrates the interconnect test structures with different geometries implemented in this work. Each of the test structures is composed of wide metal feeder lines connected using multiple vias which ensures redundancy at the driver end and a single via at the DUT end. This combination then implements an upstream / downstream for the case of M2 Feeder – M3 DUT (M2-M3) and M4 Feeder – M5 DUT (M4-M5), and a downstream / upstream for M4 Feeder – M3 DUT (M4-M3). The M3 as well as M5 DUTs are implemented using the minimum width allowed by the process. The via width for M3 lines is the same as the DUT width whereas for M5 lines the contact is fully enclosed. Wire

lengths of 50µm, 100µm and 200µm were implemented for each of the DUT types.



Process	16nm FinFET
Core, I/O VDD	0.8V, 1.8V
Chip Area	926µm x 926µm
Design Area	291µm x 388um
DUT Types	M2-M3, M4-M3, M4-M5
Wire Lengths	50, 100, 200µm
Wire Widths	Minimum
DUT Count / Chip	300

(b)

Figure 3.3: Interconnect structure, stress details and implementation summary.

3.1.3 Test Setup and Experiment Methodology

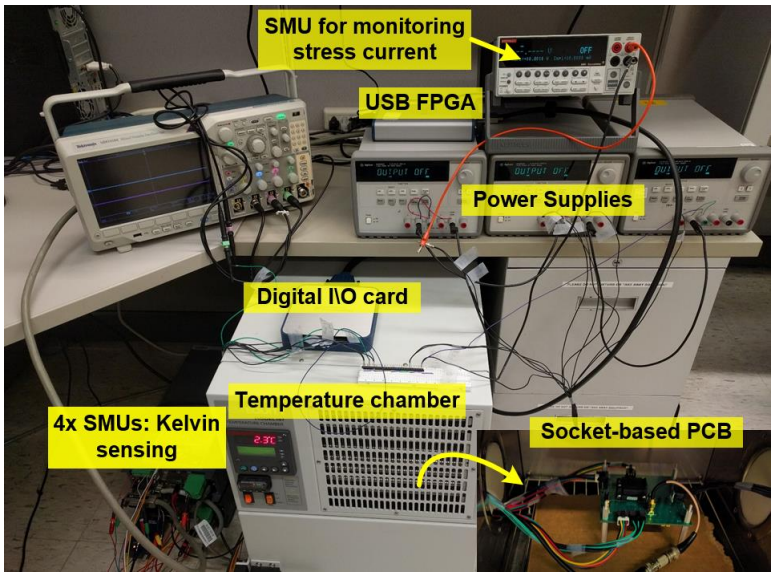


Figure 3.4: Measurement setup.

Fig. 3.4 shows the measurement setup used for the DC stress experiments. The test chip is interfaced using a socket-based PCB (inset), placed inside a temperature chamber cooled to 20°C (discussed shortly). An FPGA based control is used for sequencing the scan-chain. Kelvin sensing for the heaters and the DUT is carried out using Source-Measure Units (SMU).

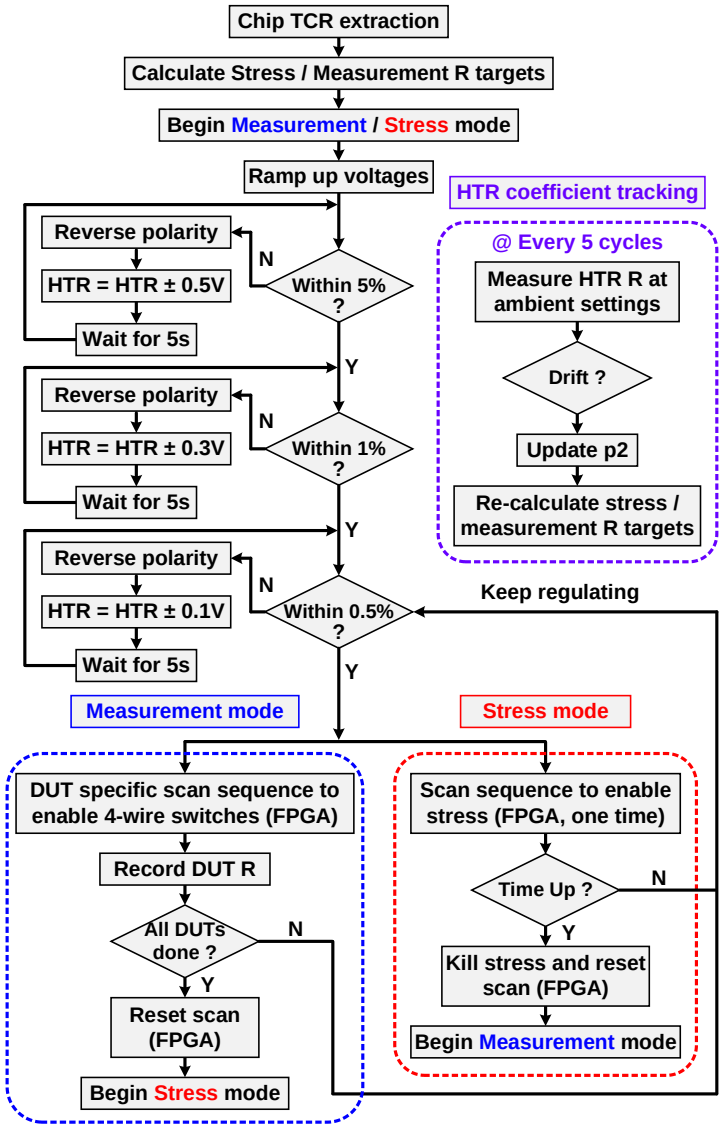


Figure 3.5: Automated test flow for TCR extraction, on-chip temperature regulation, stress control, and data collection.

Figure 3.5 illustrates the detailed test sequence. In this work, the on-chip metal heaters themselves are used to establish and monitor a desired stress temperature using the Temperature Coefficient of Resistance (TCR) method. Therefore, the first step in the experiment flow begins with the TCR extraction for the individual heaters. Fig. 3.6(a) presents the individual heater resistance distributions recorded from 11 test chips at 0°C showing a mean heater resistance of 44.71Ω and a standard deviation of 1.78Ω. Fig. 3.6(b) shows the extracted TCR measured by ramping up the chamber temperature from 0°C to 100°C in steps of 20°C as well as the extrapolation using the linear model to desired stress temperature of 325°C. The measurements demonstrate excellent linearity, however slight variations in the fresh heater resistance at 0°C (p2 parameter) as well as the TCR coefficient p1 result in dissimilar resistance values for the same target stress temperature, emphasizing the importance of precise per-heater TCR characterization. Furthermore, leakage in the DUT 4-wire switches at high stress temperatures corrupts the resistance measurement, therefore the temperature needs to be brought down during measurement. We chose a measurement temperature of 100°C as no leakage effects were detected in the sensing switches at this temperature.

With the stress and measurement targets extrapolated, the individual heater voltages are ramped up in varying steps to reach within 0.5% of the targeted resistance values. A proportional-integral-derivative (PID) loop implemented in software monitors and regulates each on-chip heater temperature separately. A tight temperature control is especially needed for recording DUT resistances to suppress measurement artifacts, and the loop ensures that all individual heaters stay close to their target temperature prior to

each sampling. The heater voltage polarity is also reversed every 5 seconds, to prevent EM induced heater failures [18]. Once the resistance values are measured, the heaters are ramped up to stress temperature (325°C) for 20 minutes with the cycle continuously looped. Fig. 3.7 presents the measured heater resistance values as well as the extrapolated stress temperatures during the repetitive measure-stress cycles (right y-axis).

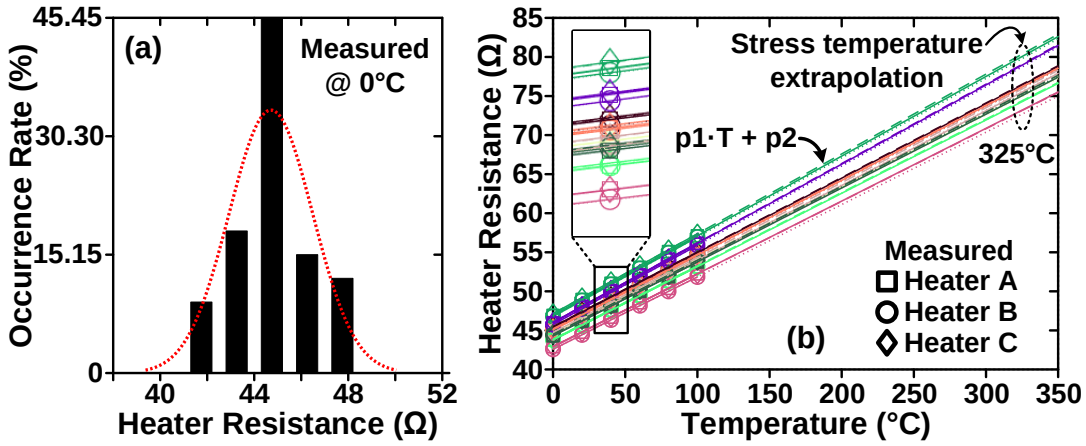


Figure 3.6: (a) Time 0 heater resistances (b) Measured TCR.

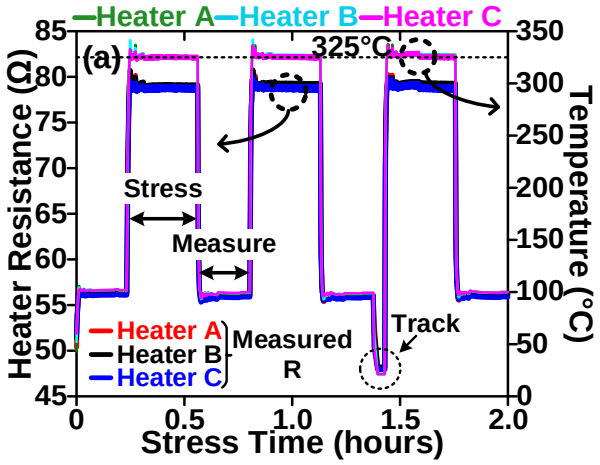


Figure 3.7: Temperature cycling during stress & measurement windows.

Fig. 3.8 shows the average heater voltages and currents over a stress cycle measured from multiple dies at an ambient temperatures of -20°C and +20°C. A lower ambient temperature, although useful from the perspective of cooling the active circuits, imposes higher power requirements on the heaters to generate the same stress temperature leading to shorter heater lifetimes. From Fig. 3.8, it can be noted that heater B sandwiched between heaters A and C, draws the lowest current of the three, attributed to the additional heat generated from both sides. Fig. 3.9 showcases the changes in heater resistance over time and the measured pre/post stress TCR. The heater resistance degrades over time which can be attributed to the sharp temperature cycling in this work, causing the heater to draw larger currents with longer stress time. The change in heater resistance is periodically tracked by a separate loop (Fig. 3.5) and the stress & measurement targets are updated accordingly.

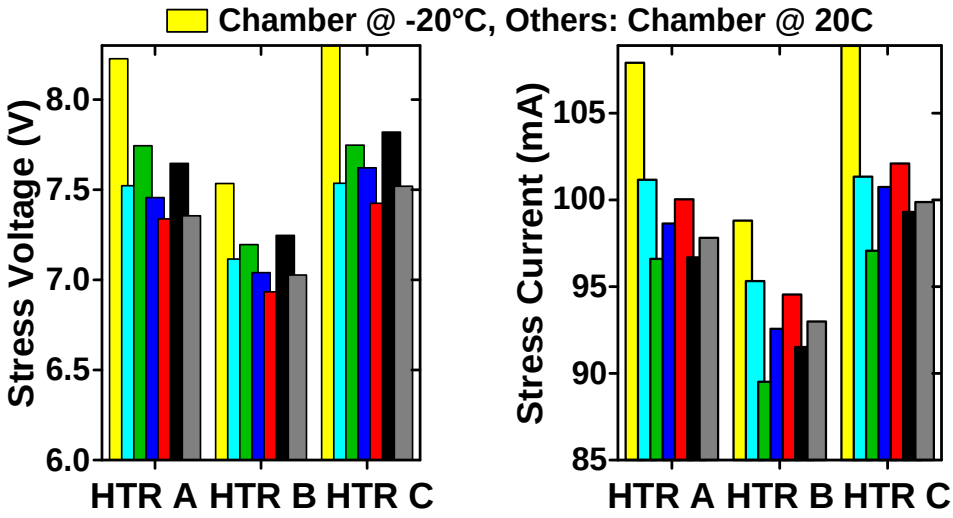


Figure 3.8: Measured heater voltages & currents at 325°C from multiple dies.

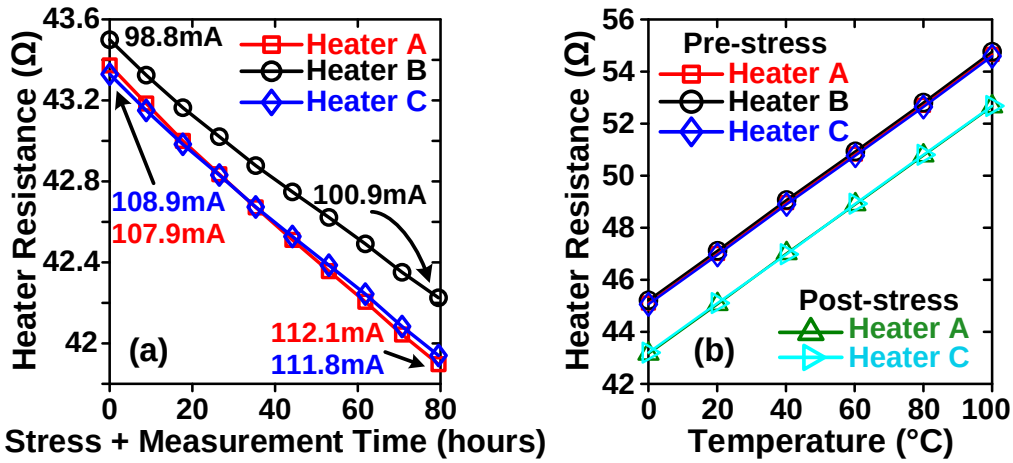


Figure 3.9: (a) Degradation in heater resistance over time and (b) Pre / post measurement TCR.

3.1.4 Discussion Of Measured Results

Fig. 3.10 - 3.12 present the sampled resistance traces, the corresponding lognormal failure distributions as well as the T-0 / fresh resistance distribution data collected from a total of 70 wires corresponding to shorter 50µm flavor for each of the distinct wire types. The stress voltage of the driver circuit was 1.5V and the stress temperature was 325°C. Abrupt (3-10X) jumps in resistance were observed in most cases for each of the narrow M2-M3, M4-M3 wires featuring via enclosure on only one side, with some cases exhibiting jumps as high as 16-25X, to rare GΩ order shifts. These failures can be attributed to 'slit void' formations directly underneath the via as shown in [19]. Compared to our previous 32nm results in [18] where resistance jumps were typically more than 100X, the resistance change was smaller in this process.

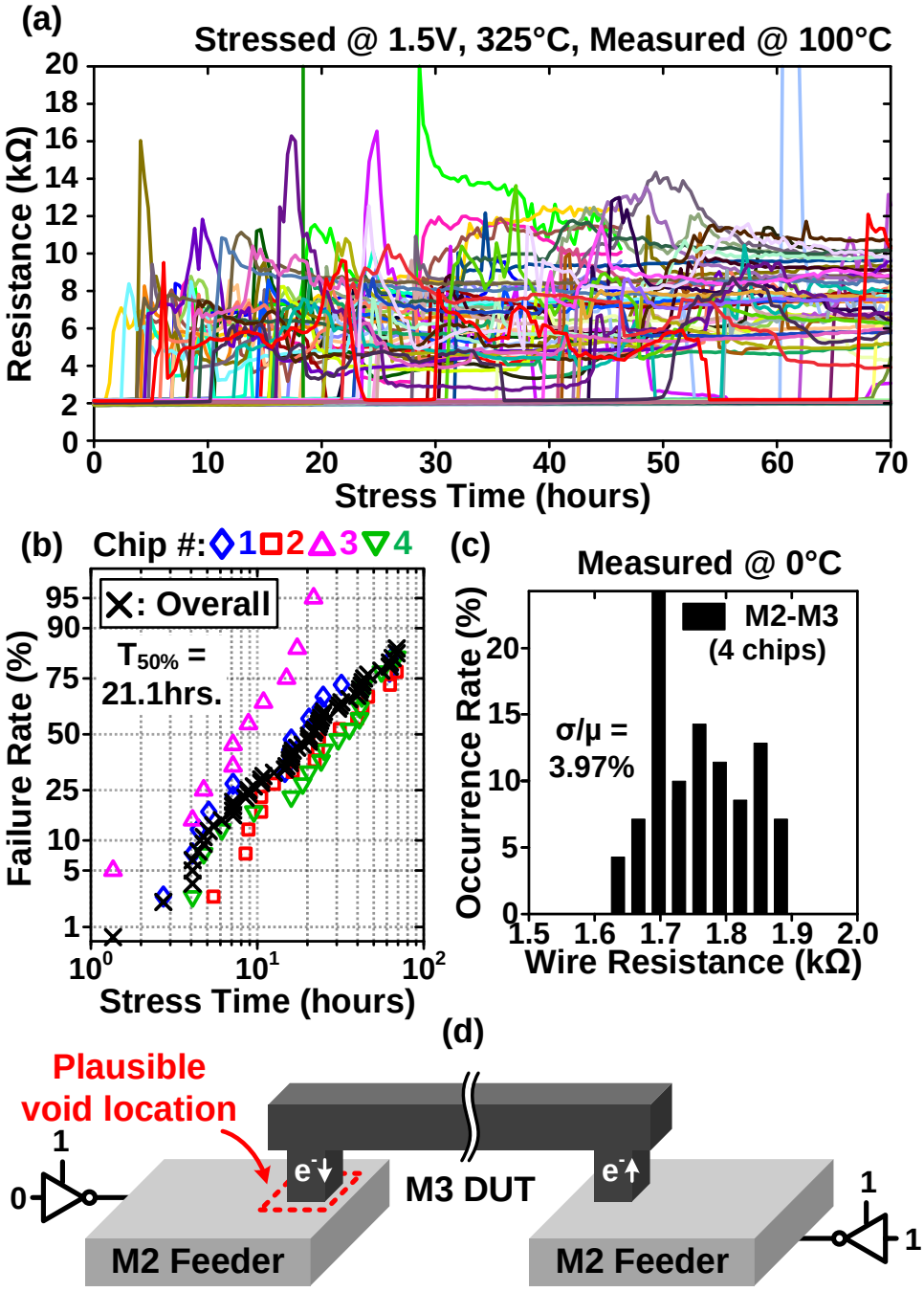


Figure 3.10: (a) Sampled M2 feeder – M3 DUT resistance traces from a total of 70 wires over the stress duration, (b) the corresponding lognormal failure distribution data mined from 4 chips, (c) Measured T-0 / fresh wire resistance distributions from the 4 test-chips & (d) Plausible void nucleation site.

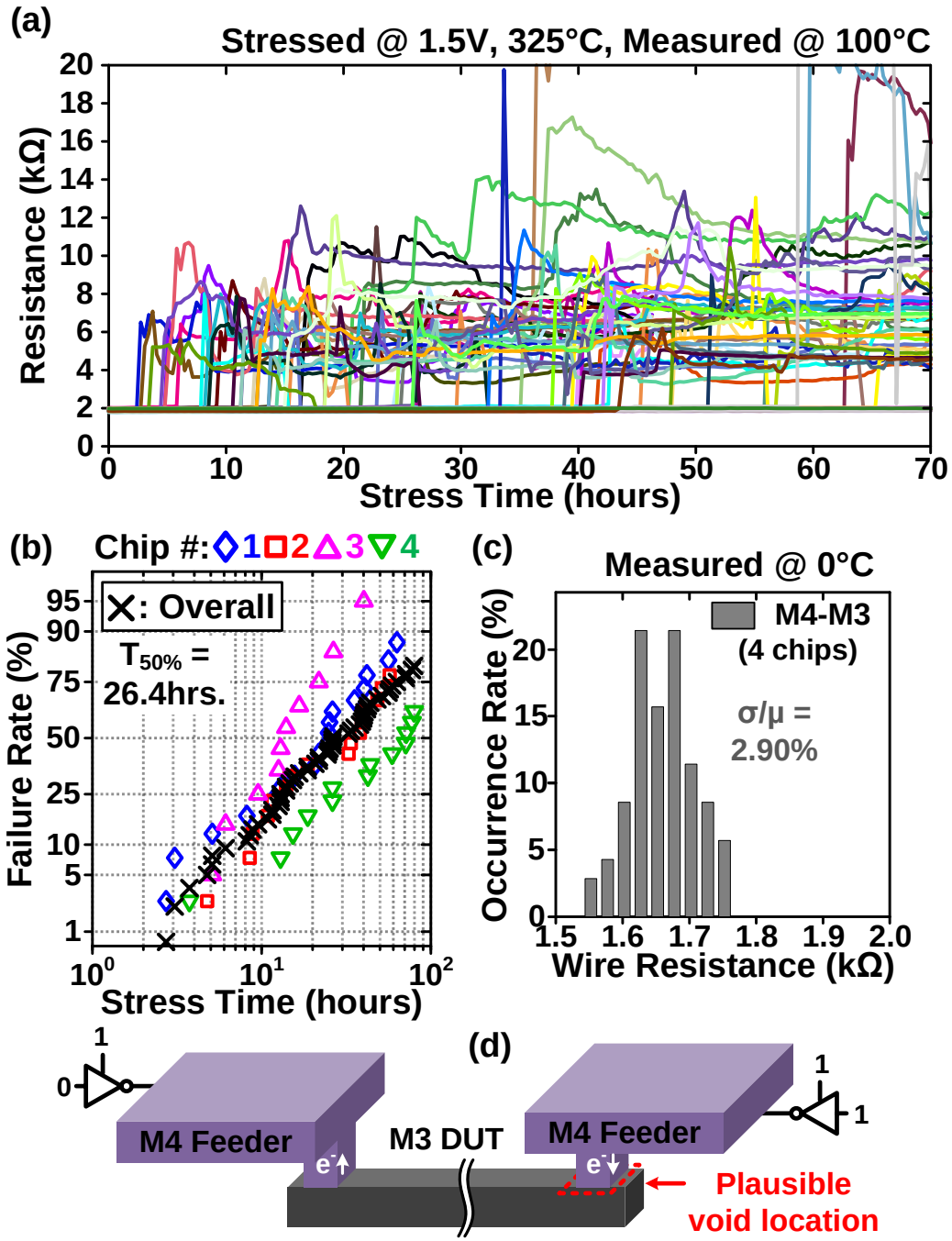


Figure 3.11: (a) Sampled M4 feeder – M3 DUT resistance traces from a total of 70 wires over the stress duration, (b) the corresponding lognormal failure distribution data mined from 4 chips, (c) Measured T-0 / fresh wire resistance distributions from the 4 test-chips & (d) Plausible void nucleation site.

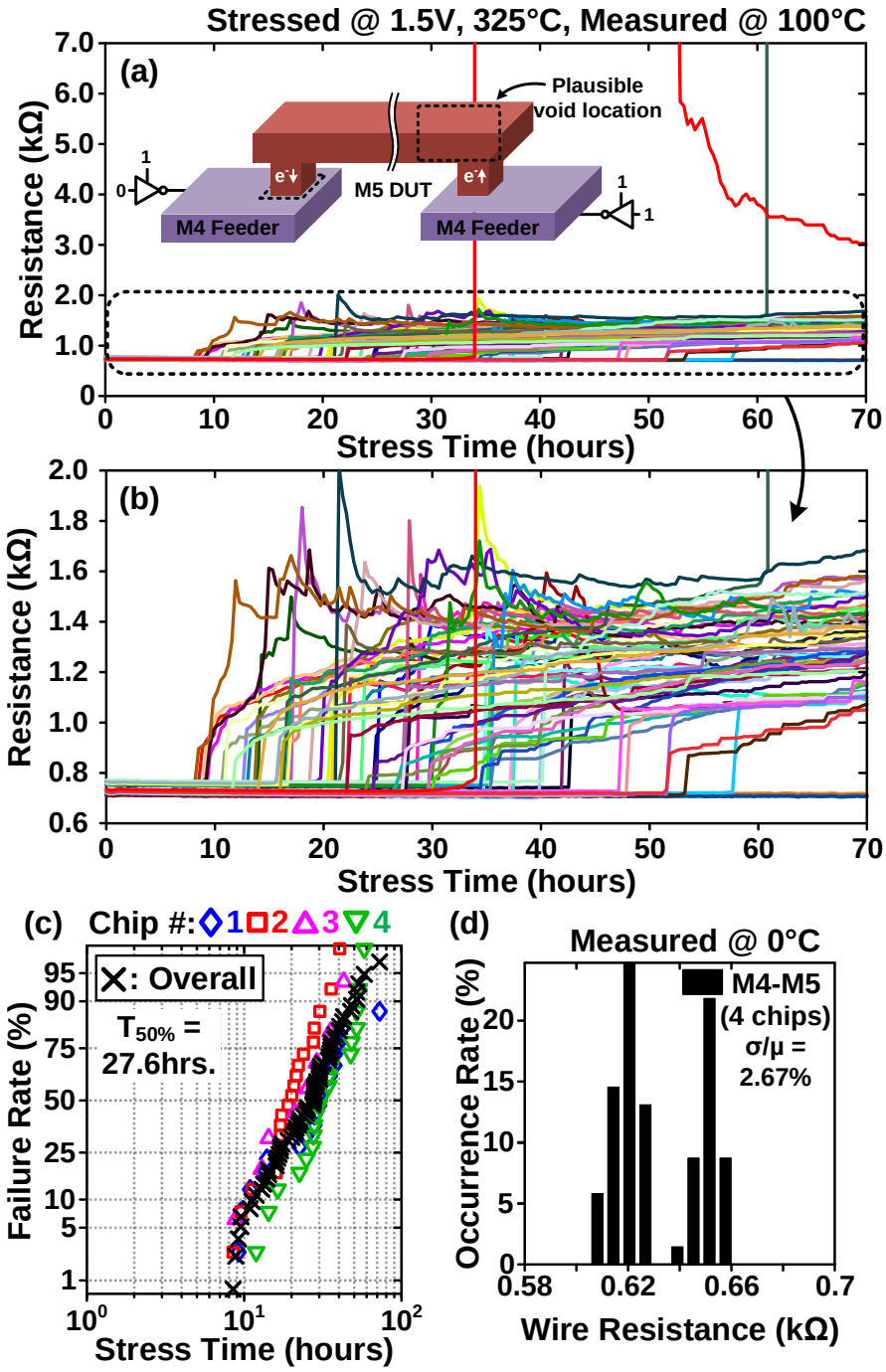


Figure 3.12: (a) Sampled M4 feeder – M5 DUT resistance traces from a total of 70 wires over the stress duration and plausible failure location (inset), (b) zoomed view of resistance traces, (c) the corresponding lognormal failure distributions & (d) Measured T-0 / fresh wire resistance distributions from the 4 test-chips.

The slightly higher Median-Time-To-Failure (MTTF) for the M4 Feeder – M3 DUT (26.4 hours) in comparison to the M2 Feeder – M3 DUT combination (21.1 hours) can be possibly attributed to the more regular grain structure in the narrower M3 DUT, making it more resilient to EM as compared to the wide M2 Feeder in case of the latter.

Taking a closer look at the failure mode for these two cases, it is observed that the resistance jumps up instantaneously, then continues to increase or decrease overtime. In a few select cases, it is also observed that the wire resistance recovers back to its original value for certain intervals of time before an EM failure happens again. This behavior was also observed in our previous 32nm HKMG test-chip and could be attributed as a possible consequence of the fast temperature cycling when switching between the stress and measurement modes (100°C – 325°– 100°C …) or due to mechanical back stress forcing the material to reconnect [18] (For a further detailed discussion, refer to [18]).

In contrast, the wider M4-M5 wires (Fig. 3.12) featuring a full contact enclosure exhibited relatively few abrupt failures (attributed to slit-void formations) and predominantly progressive monotonic shifts (void nucleating within the wire) in the DUT resistance overtime despite featuring a similar stress current density. Based on the observed failure trends, the plausible failure locations can be on the downstream side underneath the wide M4 feeder (abrupt) or within the M5 wire itself, predominantly on the upstream end (progressive). The relatively fewer abrupt failures in this case can be explained by the wider via size (1.25X) leading to lower stress in the M4 feeder underneath. Predominance of

progressive failures can be attributed to with the possible lack of the well-defined (more regular) grain structure in the wide M5 wire (2.5X wider compared to M3 DUTs) causing a higher susceptibility to EM effects as compared to the M3 DUTs.

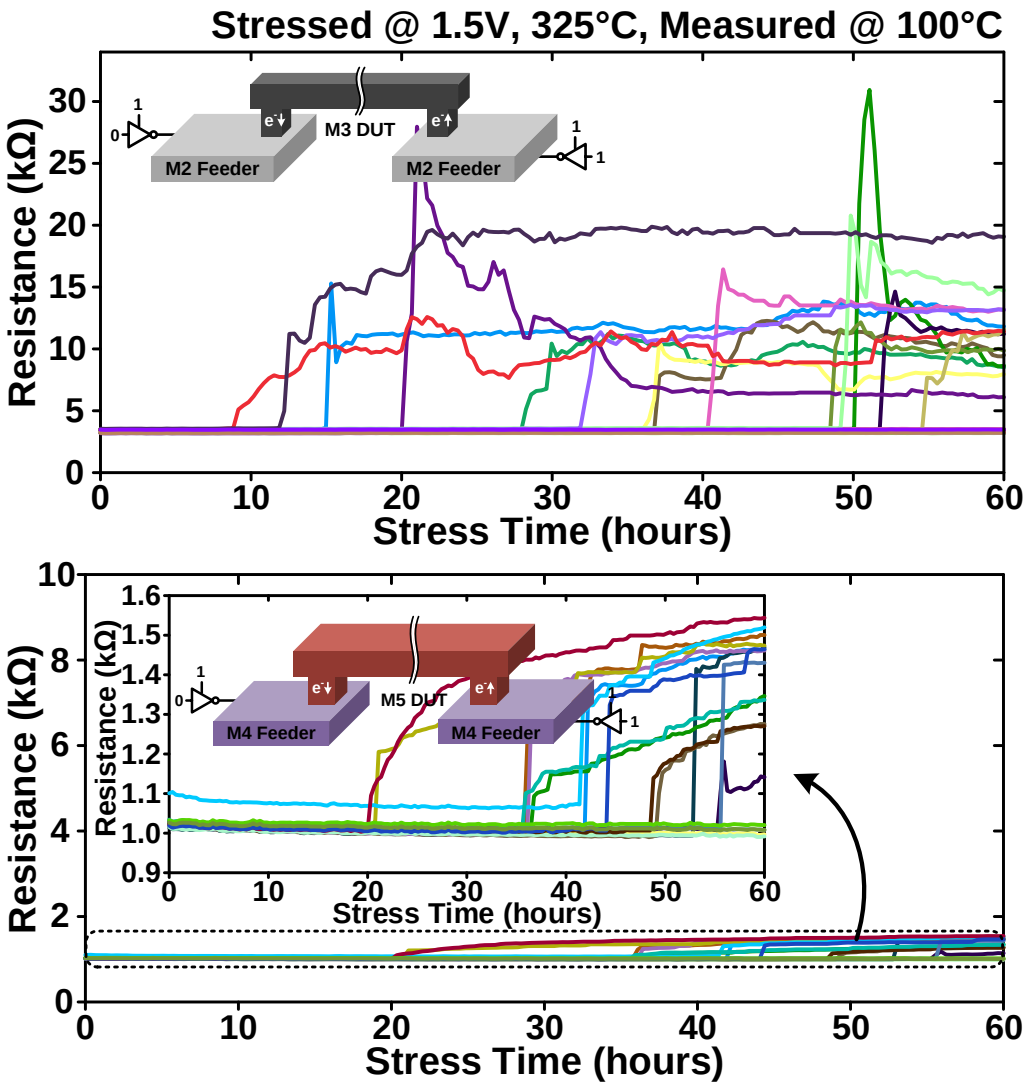


Figure 3.13: Sampled resistance traces for the same stress duration from 100µm wires.

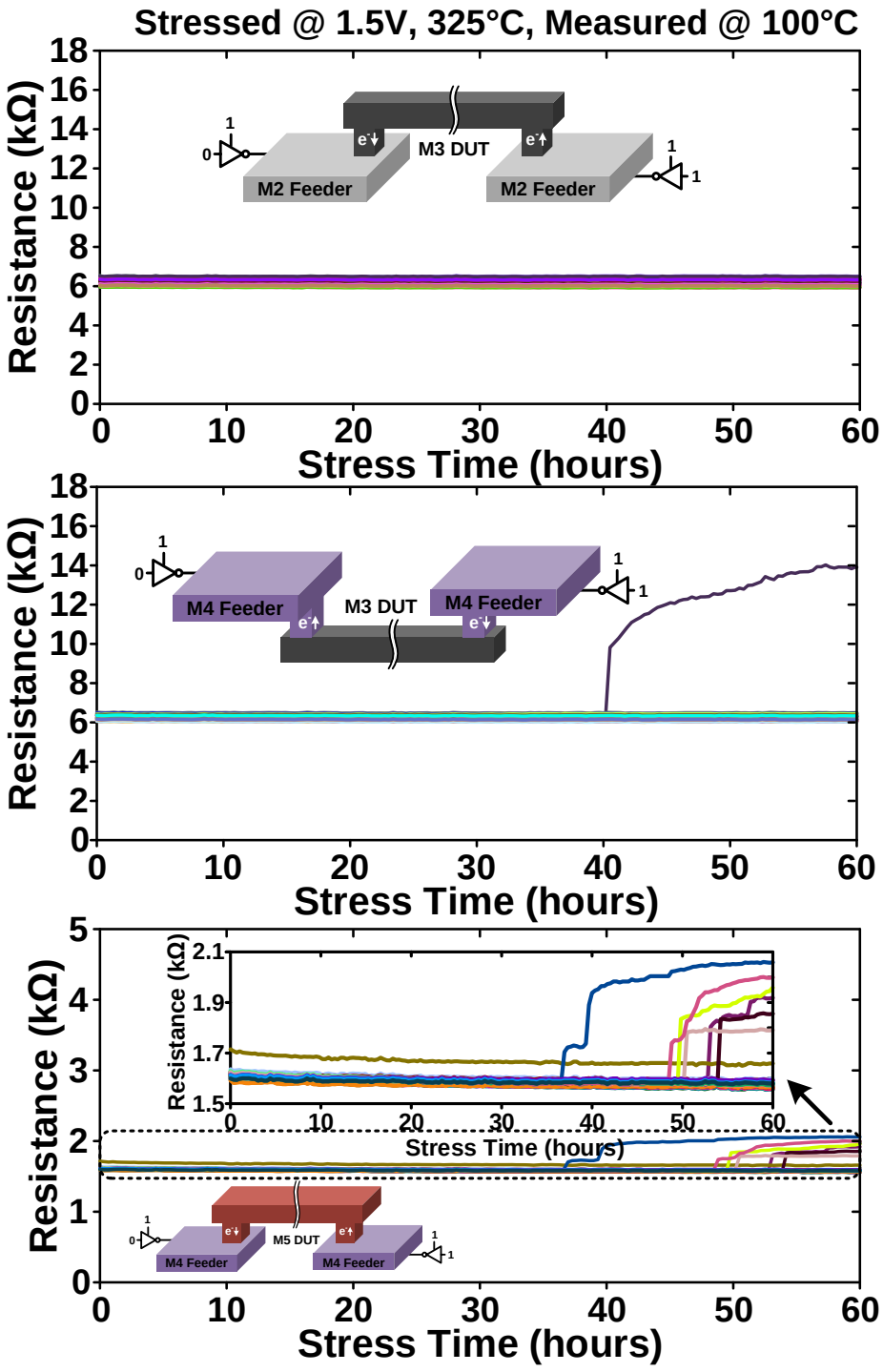


Figure 3.14: Sampled resistance traces for the same stress duration from 200µm wires.

Fig. 3.13 – 3.14 present the resistance traces sampled form the longer 100µm and 200µm wires. Very few (100µm) to negligible failures (200µm) are observed for the case of longer wires over the same stress duration as their 50µm counterparts, primarily due to lower stress current density, a consequence of the constant voltage stress condition used in this work.

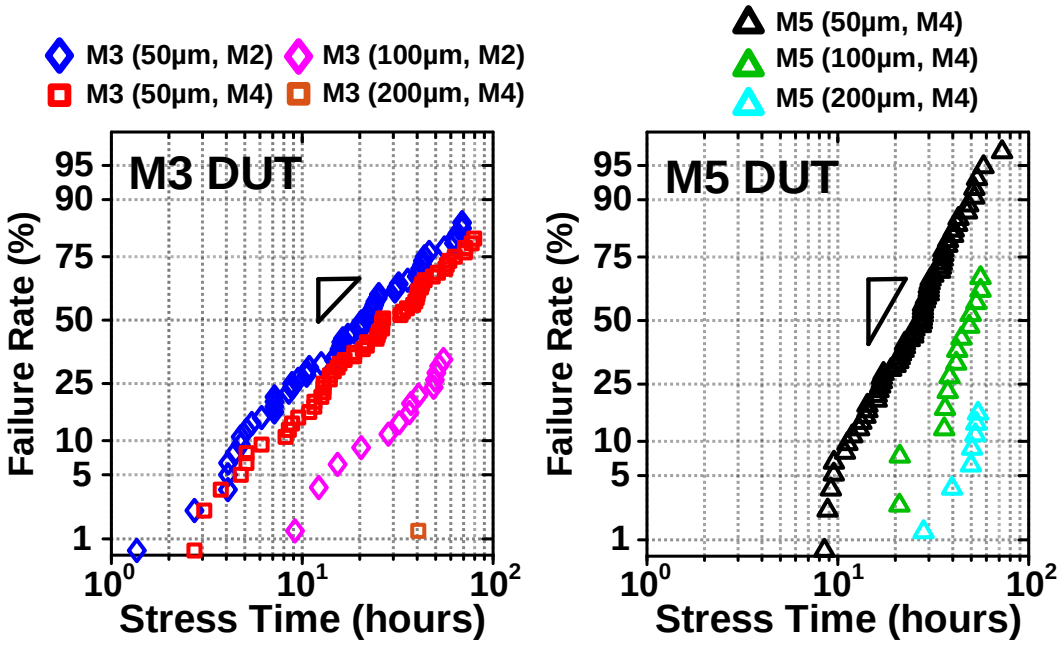


Figure 3.15: Abrupt vs. Progressive lognormal failure distributions for the two distinct DUT types viz. M3, M5 used in this work.

Finally, Fig. 3.15 presents a comparison between the abrupt (predominant failure mode for the narrower M3 DUT) and progressive (predominant failure mode for the M5 DUT) EM failure distributions. The abrupt failure distributions owing to 'slit-void' formations were observed to have a wider distribution in comparison to the progressive failures owing to 'trench voids' nucleating within the wire in agreement with previously published findings [19], [20].

3.1.5 Conclusion

This section showcased measured data corresponding to direct-current (DC) stress induced EM phenomenon, characterized using on-chip circuits for interconnect test structures fabricated in a 16nm FinFET process. An array-based test vehicle featuring parallel stress and 4-wire Kelvin sensing capabilities was presented, employing wires with distinct feature sizes and metal stacks as the Devices-Under-Test (DUTs). Accelerated stress testing was achieved entirely on-chip using metal heaters positioned directly above the DUTs, which provided precise local temperature regulation in conjunction with fast cycling between stress and measurement temperatures. In addition to the improved test-efficiency together with the reduction in die-area and equipment complexity as compared to the traditional industry-standard oven-based approach, known EM failure modes ranging from abrupt and/or progressive were captured from different DUT geometries and metal stacks, together with effects such as temporary healing as well as the circuit-interconnect interplay, making the approach suitable for accurate MTTF extraction in realistic scenarios on-chip, especially in context of future technology nodes.

3.2 A Digital-Intensive Hardware Monitor in 16nm for Characterizing Electromigration-Induced Bit-Error-Rate Degradation of Interconnect Signal Paths

3.2.1 Introduction

In the previous section, an array-based hardware monitor based on resistance sensing was presented for efficient characterization of EM effects on-chip. In this section,

we explore for the first time, the impact of EM on the BER of a signal interconnect path using a dedicated array-based test vehicle. Traditionally, single wire based DUTs have been used to characterize EM lifetime where the resistance shift is monitored under elevated temperatures, with either a constant stress current or a constant stress voltage. In contrast, the proposed test structure employs a number of novel on-chip monitoring circuit techniques including on-chip heaters, a bit pattern generator, and an on-chip BER monitor to simplify the overall test setup and allow high precision characterization of EM induced resistance shifts using only digital circuits. Measurement data collected from a 16nm test chip reveals unique insights into EM induced signal path degradation that was not available prior to this work. Our experimental studies suggest that monitoring the BER of an interconnect path could be used as a new metric for capturing EM induced resistance shifts in a real system.

3.2.2 BER-Based EM Degradation Monitor Design

Fig. 3.16 presents an overview of the proposed test-chip architecture detailing the on-chip metal heater and the interconnects comprising the datapath routed underneath it, together with the error sampling monitors local to a group and the measurement circuitry specifics. The test-chip is similar in organization to our previous EM characterization macro [7], with the heating area consisting of the three on-chip metal heaters centrally located, and the active circuits kept away from the heater edges, to avoid any possible damage to core circuits during stress. A unit tileable cell-based approach is followed throughout, both in designing the heaters as well as the left and right arrays housing the local datapath circuits and the error sampling monitors.

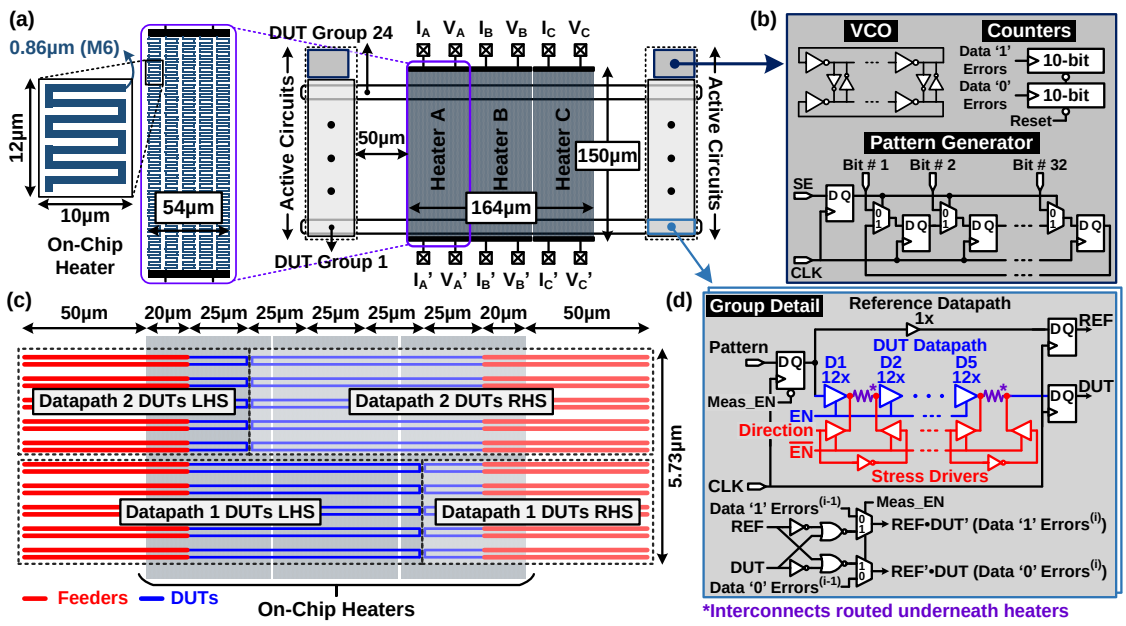
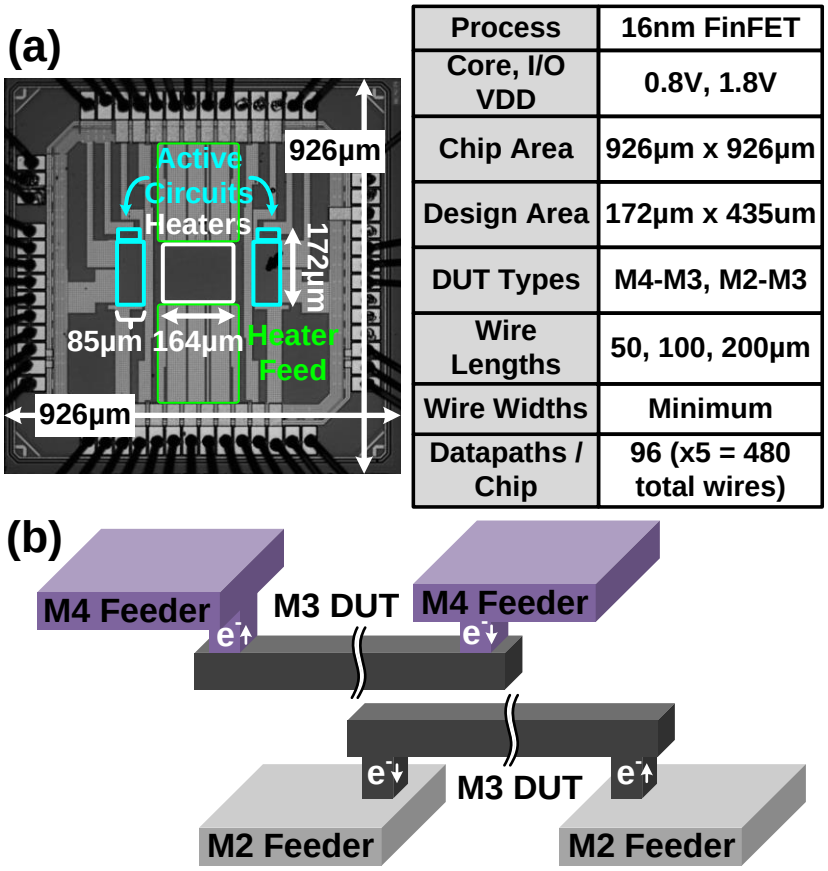


Figure 3.16: (a) Overview of the 16nm EM induced BER shift characterization monitor and on-chip heater specifics; (b) Measurement circuit consisting of VCO, error counters, and 32-bit data pattern generator; (c) Datapath routing details comprising one DUT group underneath the heating area; (d) Unit tileable DUT group details: Reference and DUT datapaths, with separate error counters for data '0' and data '1'.

Figure 3.16(a) further illustrates the details of the on-chip metal heater, constructed by cascading the unit-cell employing wide M6 tracks, coiled symmetrically in a serpentine fashion to ensure uniform heat distribution underneath. The local heater temperature is both sensed and regulated using the 4-wire Kelvin sensing arrangement embedded per heater. Fig. 3.16(b) presents the measurement circuitry, located towards the top corner of either array, housing the clock and pattern generator, together with the 10-bit counters for separately accumulating the incoming locally sampled errors, corresponding to data '0' and data '1', for a selected datapath in the array. The clock generator is a supply-tuned, cross-

coupled ring-based voltage-controlled-oscillator (VCO), whereas the pattern generator is implemented as a 32-bit circular shift-register, with a scan-based parallel load capability.

The DUT groups consist of two separate datapaths per array (Fig. 3.15 (c, d)), each further consisting of five identical stages comprising a tri-state buffer driving an identical flavor of the interconnect per stage. Fig. 3.16(c) illustrates the layout details for one such DUT group, with the interconnect test structures routed underneath the heaters using wide feeder tracks to minimize the IR drop and prevent EM problems outside of the DUT. Furthermore, the DUTs themselves are folded and mirrored so as to maximize the area utilization. The BER sampling circuity local to a datapath (Fig. 3.16(d)) consists of a reference datapath comprising a signal buffer, along with the actual datapath consisting of the 5 interconnect stages, both driven by the same incoming pattern sequence. Their outputs are then sampled on the next rising edge of the clock to generate the REF and DUT signals. To separately monitor data '1' and data '0' error rates, we perform REF•DUT' and REF'•DUT logic operations which increment the two 10-bit counters, respectively. The interconnects are DC stressed using separate tri-state stress drivers placed at both ends of the DUTs, with programmable control to allow for a unidirectional current flow in the desired polarity for stressing the DUTs. A scan-based control is used for selecting one or multiple datapaths under parallel stress, automatically disabling the corresponding datapath drivers. Additionally, a 1-bit control (MEAS_EN) allows for selection of a single datapath for the BER measurement pre- and post-stress.



Process	16nm FinFET
Core, I/O VDD	0.8V, 1.8V
Chip Area	926µm x 926µm
Design Area	172µm x 435um
DUT Types	M4-M3, M2-M3
Wire Lengths	50, 100, 200µm
Wire Widths	Minimum
Datapaths / Chip	96 (x5 = 480 total wires)

Figure 3.17: a) Die photo and chip feature summary; (b) metal test structure specifics

Fig. 3.17(a) presents the die microphotograph and chip feature summary. The design is implemented in a 16nm FinFET process with a total of 96 datapaths implemented per chip. The DUTs comprising the datapath are implemented as standard upstream-downstream and downstream-upstream interconnect test structures with a long M3 wire (Fig. 3.17(b)). Wire lengths of 50, 100, and 200µm were implemented while their widths are kept at minimum allowed by the process. The non-minimum sized feeders routing the DUTs are connected using multiple vias at the driver end (to ensure redundancy) and a single via at the receiver end.

3.2.3 Test Setup and Experiment Flow

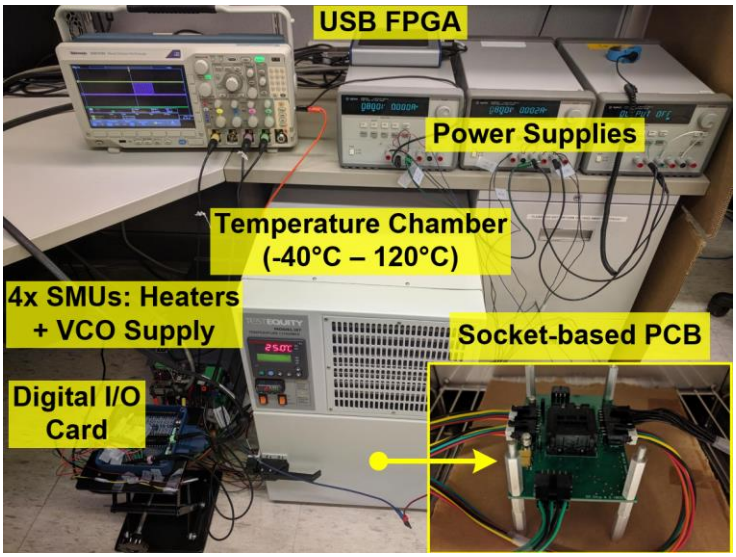


Figure 3.18: Measurement Setup.

Fig. 3.18 shows the details of our measurement setup. An FPGA board is used for digital signal generation and acquisition activities, for instance, defining the time window for error collection, scan-based digital control, on-chip VCO characterization, acquiring counter outputs etc. The on-chip heater regulation and VCO power supply are controlled using source-measure units. A benchtop temperature chamber is used for the heater temperature-coefficient-of-resistance (TCR) extraction and maintaining an ambient temperature of 20°C throughout the test.

Fig. 3.19 illustrates the measurement flow. To ensure confidence in the measured BER, the algorithm keeps updating the window size until the specified statistical targets are met. Fig. 3.19 (b) showcases the linearity in the sampled mean counts for a range of linearly increasing window sizes, with the measured BERs approaching near constant values for larger window sizes on account of the increasing statistical confidence. Fig. 3.19

(c) further illustrates the sampled counter outputs corresponding to two different windows, clearly showing reduced variation for larger measurement windows.

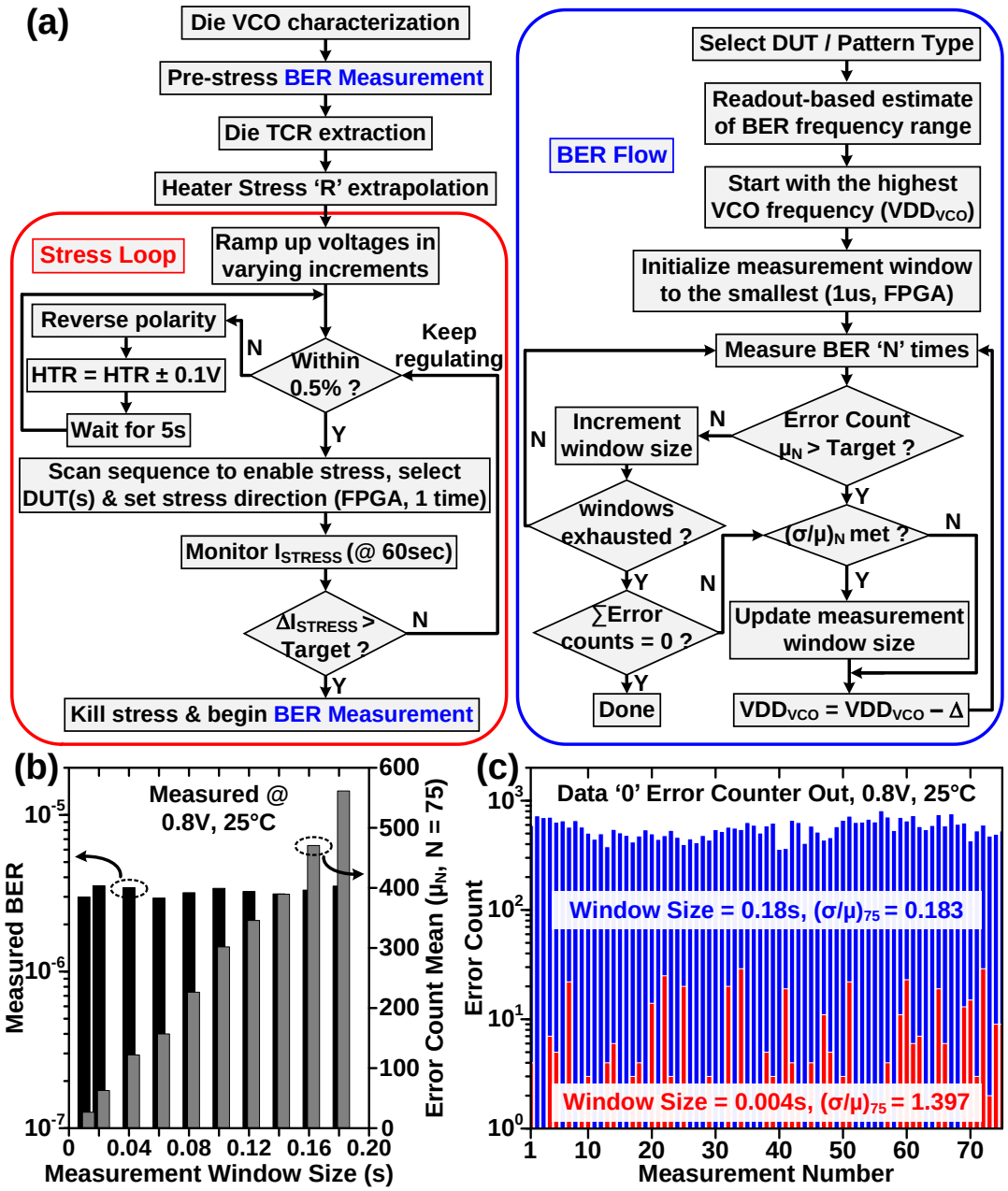


Figure 3.19: (a) Automated test-flow for die TCR extraction, on-chip stress regulation and datapath BER characterization. (b) Measured BER and linearity in error counts vs. the chosen window size. (c) Sampled counter outputs for distinct window sizes.

Fig. 3.20 presents the T-0 heater resistances and the extracted TCR of the on-chip heaters from 9 dies, showcasing the excellent linearity in the measured TCR, as well as the extrapolation to target stress temperature (360°C). With the per heater stress resistance targets known, the voltages across these are ramped up in varying increments to reach within 0.5% of the final temperature and continuously regulated thereafter. From here on, the shift in stress current is monitored every minute, until it falls below a set target, after which the stress is turned off and the BER for the selected datapaths are characterized again, with the process repeated for the entire duration of the experiment.

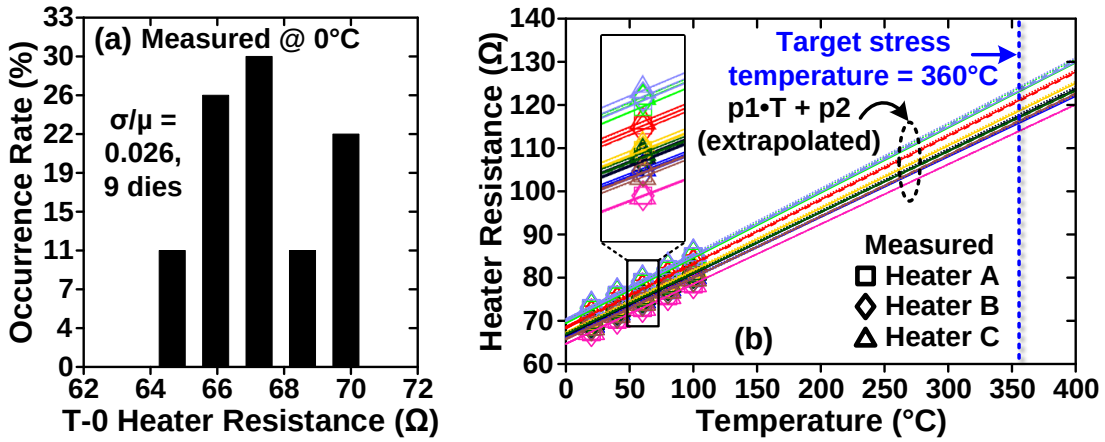


Figure 3.20: (a) T-0 heater resistances & (b) Heater TCR characterization and extrapolation to target stress temperature

Fig. 3.21 presents the heater power needed to reach different stress temperatures mined from multiple dies.

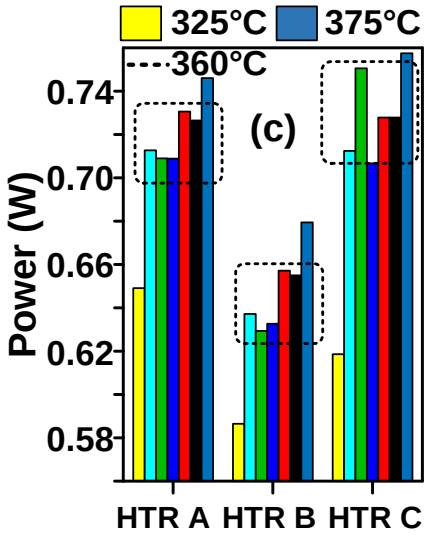


Figure 3.21: Measured heater power from multiple dies needed to reach different on-chip stress temperatures.

3.2.4 Measurement Results from 16nm Test-Chip

We first show test data verifying the basic functionality of the 16nm test chip. Fig. 3.22 (a) presents the T-0 BER measurements corresponding to a total of 24 datapaths, in varying flavors of interconnect lengths (50, 100, 200µm) and metal stacks (M4-M3-M4, M2-M3-M2), characterized using a pattern sequence with alternating '0's and '1's. As the VCO clock frequency is swept from a low frequency to a high frequency, the BER increases exponentially and then plateaus, giving rise to the typical BER curve shape shown in Fig. 3.21(a). As expected, BER curves saturate at lower VCO frequencies for longer interconnects. From Fig. 3.22 (a) & (b), it can also be observed that 200µm datapaths have a tighter distribution compare to the shorter datapaths, a consequence of the interconnect delay dominating the overall delay, offsetting any device variations.

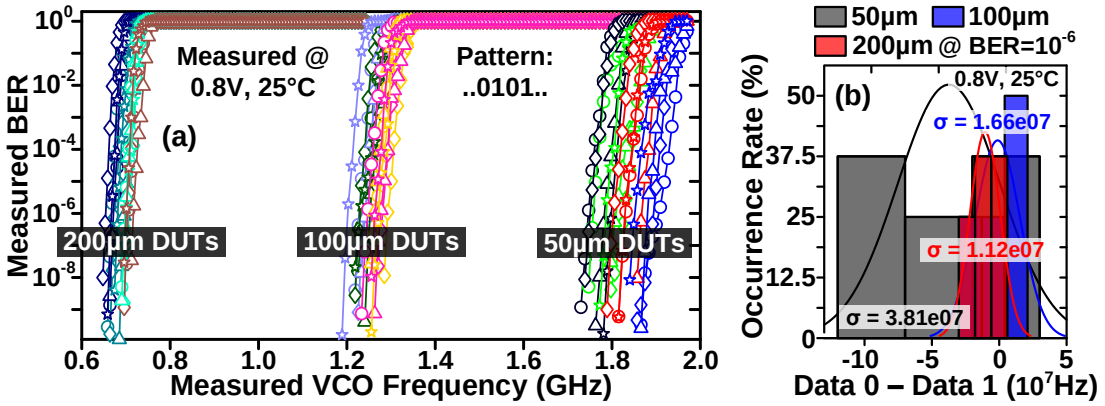


Figure 3.22: (a) Time-0 BER characterization of fresh datapaths comprising 50, 100, and 200µm wires & (b) Separation between the data '0' & '1' curves for a BER = 10^{-6}.

Fig. 3.23 presents the measured BER results corresponding to patterns with differing numbers of consecutive '0's / '1's, performed as a sanity check on the same 50µm M4-M3 datapath. As expected, the BER saturates at a level that corresponds to the ratio between the number of 0-to-1 or 1-to-0 transitions and the total number of bits transmitted.

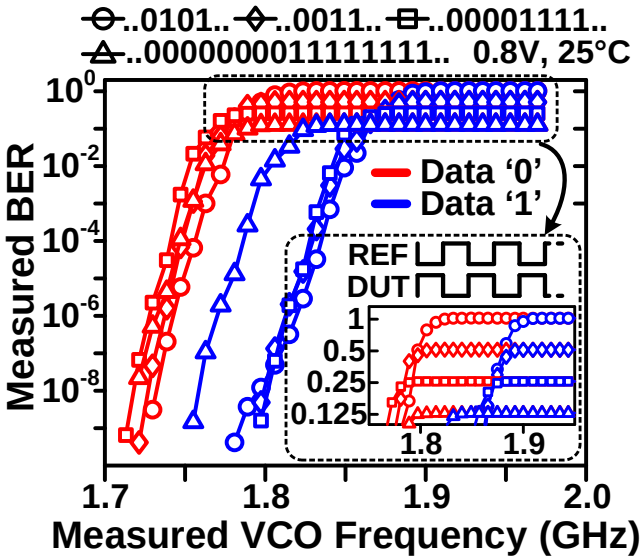


Figure 3.23: BER sanity check using different pattern types.

For the stress experiments, we have chosen the pattern with alternating '0's and '1's. A total of 8 datapaths, comprising only the 50µm M4-M3 interconnects, are stressed in parallel, with stress current flowing from the receiver end to the driver end of the wire. The 40 wire DUTs are stressed while elevating the heater temperature to 360°C. To avoid the possibility of degradation in the drivers themselves, we used a nominal 0.8V VDD for DC stress. Fig. 3.24 presents the degradation in stress current I_{STRESS} over the experiment duration, as a consequence of the EM induced DUT resistance degradation over time. Fine shifts in I_{STRESS}, with the threshold set between 1-1.5% are monitored per stress cycle (except for the final two), following which, the heaters and stress current are turned off, and the VCO and datapath BER are re-characterized. VCO characterization results in Fig. 3.25 confirm that the output clock frequency remains constant throughout the experiment, with no signs of device degradation.

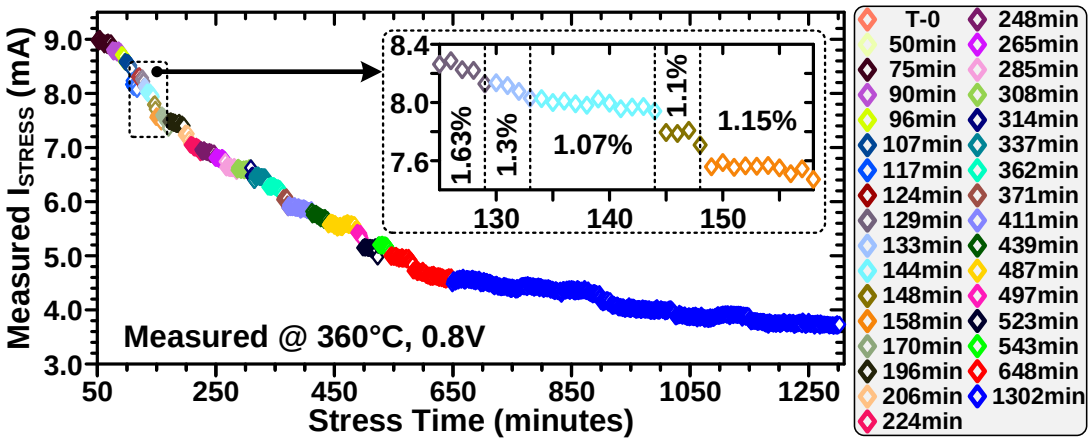


Figure 3.24: Measured stress current (I_{STRESS}) degradation.

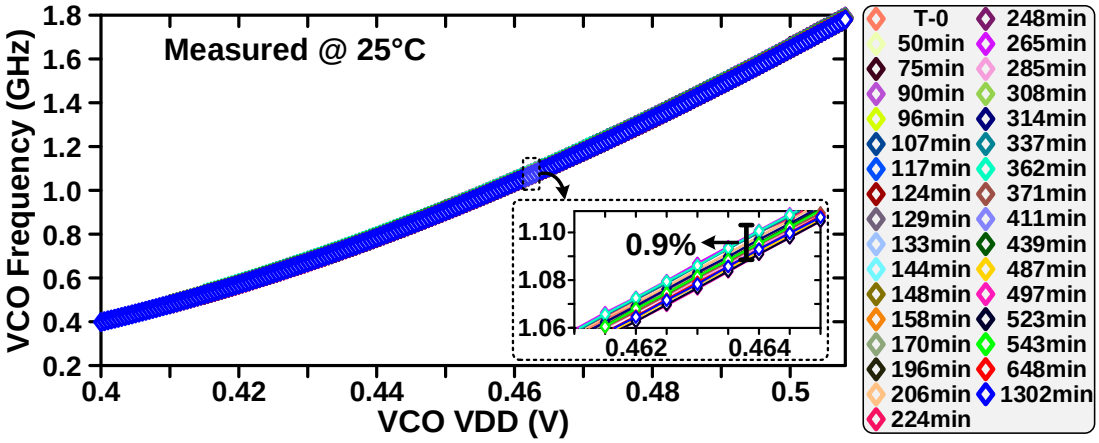


Figure 3.25: On-chip VCO characterization between multiple stress cycles.

Fig. 3.26 shows the BER characterization results from 4 independent datapaths, together with the degradation in path frequency for a BER = 10^{-6} as a function of the stress time. Due to the long BER test time, we only measured 4 out of the 8 stressed datapaths. The BER curves remain relatively constant with no appreciable shifts for the first 96 to 248 minutes, depending on the datapath number. This represents the initial phase of EM where vacancies are moving towards receiver end of the DUT wire. As the EM voids nucleate and grow, distinct abrupt and progressive failure signatures are observed over the experiment duration for the same stress time in different datapaths. For instance, datapath 1 exhibits a slow EM failure rate, with little to no degradation for the first 4 hours, followed by an abrupt shift and a slow progressive deterioration thereafter. In contrast, datapaths 2-4 exhibit a fast degradation over time, exhibiting distinctly evident fast abrupt shifts over the same stress time, followed finally by slow progressive degradation. We did not observe any noticeable difference between the data ‘1’ and data ‘0’ BER curve shifts. It’s also interesting to see that none of the datapaths completely failed even after stressing the DUTs

for 1,302 minutes, suggesting that low frequency functionality testing may not be effective in diagnosing EM-induced BER shifts. Statistical data from additional datapaths will be required to build BER models that can capture the EM failure characteristics. This work represents the first step towards understanding post-EM BER degradation trends of signal interconnects.

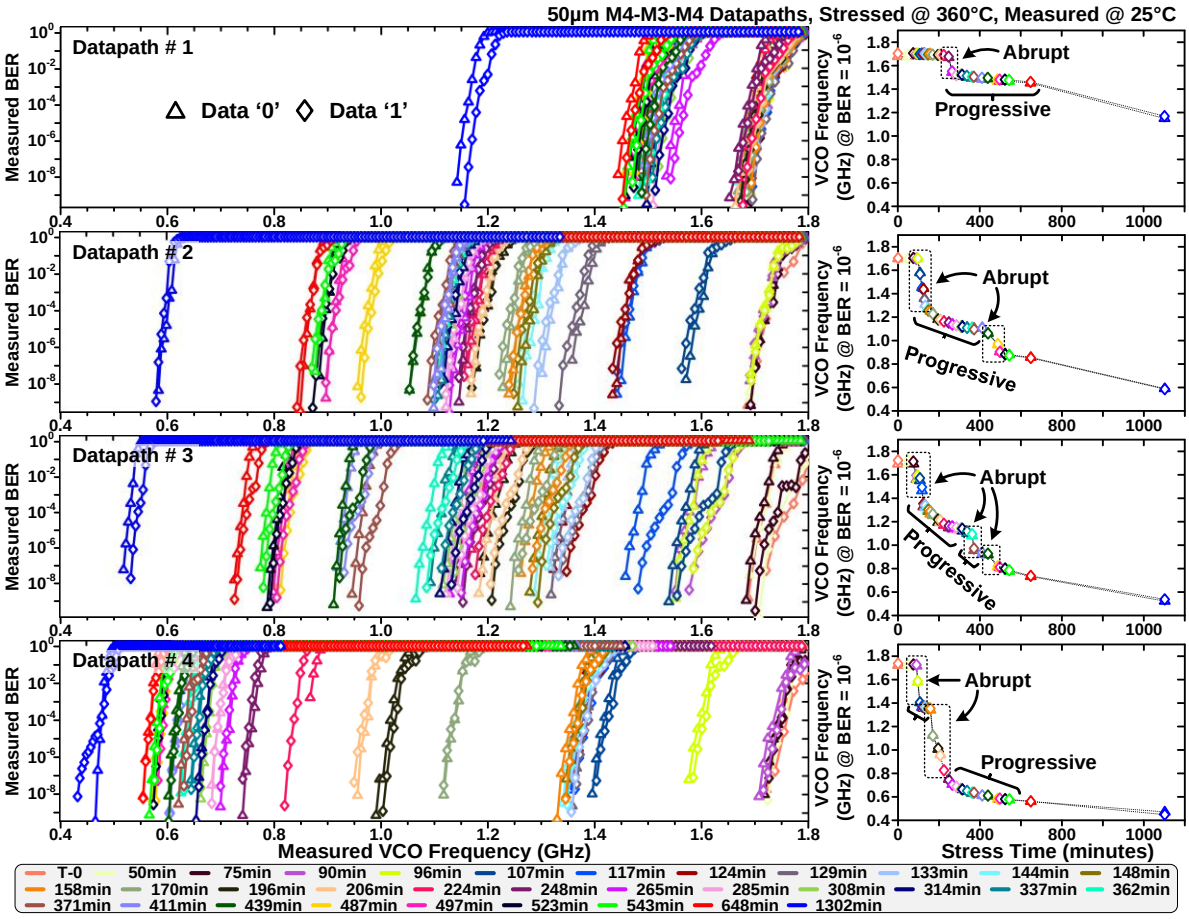


Figure 3.26: (a) Measured shift in BER captured at different stress intervals up to 1,302 minutes of total stress time and (b) Degradation in path frequency vs. stress time for a BER of 10^{-6}.

3.2.5 Conclusion

In this section, an array-based test-vehicle for tracking bit-error-rate (BER) degradation of signal interconnects subject to DC electromigration (EM) stress was proposed and implemented in a 16nm FinFET process. A unit interconnect path comprised of five identical interconnect stages where each wire is driven by inverter based buffers. Accelerated EM stress testing was achieved entirely on-chip using metal heaters located directly above the devices-under-test (DUTs) and separate stress circuits driving both ends of the wire. BER was introduced as a new metric for diagnosing EM induced resistance degradation. Finally, measurement strategy and results from four individual interconnect paths were presented and analyzed.

3.3 A 22nm Array-Based EM-Induced Bit-Error-Rate Degradation Monitor featuring DUT Resistance Sensing & Stress Capabilities with added flexibility in Mode-Control

3.3.1 Introduction

In section 3.1, we introduced an array-based characterization vehicle with on-chip stress and 4-wire Kelvin sensing capabilities specifically meant for gathering the DUT MTTF statistics efficiently from a large sample population of wires. Section 3.2 further introduced a new test-vehicle wherein monitoring the Bit-Error-Rate of an interconnect path was proposed as a new metric for quantifying the EM-induced resistance shifts

directly in terms of the datapath's signaling ability. In this section, we build upon the work presented in the previous two sections, by introducing a new EM characterization macro, which features a combination of capabilities borrowed from the previous two designs, with the primary motivation being the translation of a measured EM-induced resistance shift to an associated degradation in the measured BER of the interconnect path. The vehicle further features an additional flexibility in enabling the Kelvin-Sensing transmission gate pair, which is helpful in isolating the Joule heating solely due to the stress current in the wire or monitoring the local DUT temperature during the stress mode when the on-chip heaters are activated. The proposed test-vehicle designed in a 22nm Fully-Depleted-Silicon-On-Insulator (FDSOI) technology node is presented in the subsequent section, along with a description of the added improvements compared to the designs presented in the former two sections.

3.3.2 The 22nm EM Resistance Shift & BER Degradation Capture Monitor

Fig. 3.27 presents an overview of the proposed EM characterization macro. A single metal heater unit with embedded 4-wire resistance sensing capability, similar in design to the one presented in section 3.2, is employed for local stress temperature generation and regulation on-die. A key difference, however, is the separation between the metal heater and the core circuits, which is kept at 10µm in the new design (as opposed to 50µm in the former two) based on experimental studies from another test-vehicle designed in 65nm. This results in a more realistic DUT design, with the feeder resistance now being a smaller fraction (<10%, post-layout simulations) of the overall DUT resistance, causing most of the stress current to appear across the narrower segment, while also reducing the capacitive

load for the core-driver to that of the wire-segment of interest. The driver buffers (Fig. 3.27 (c)) are sized to 8 times the minimum based on post-layout simulations, to result in a realistic Fan-Out of 4 (FO4) delay, commonly seen in most digital systems. As before, the reference datapath comprises of a single minimum sized buffer.

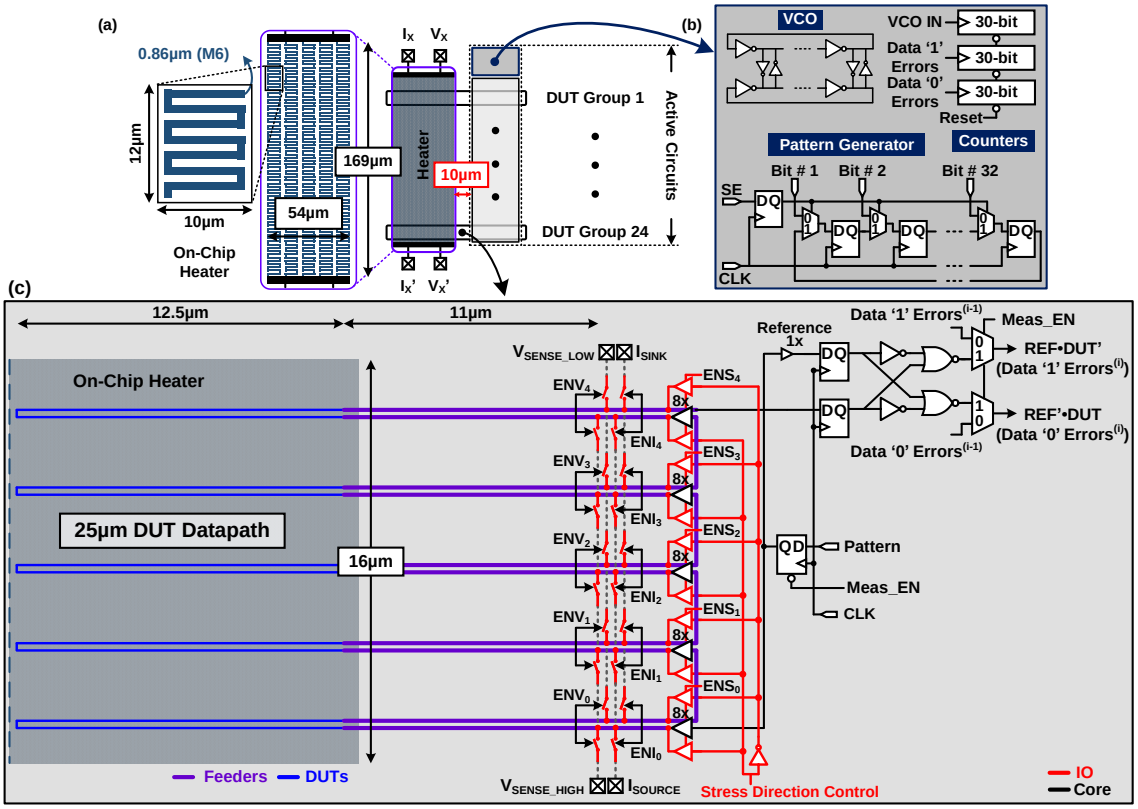


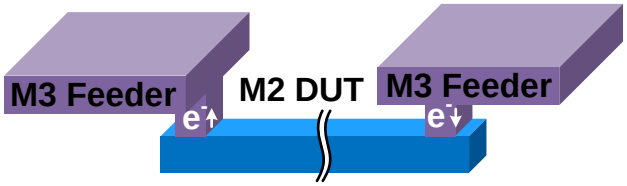
Figure 3.27: (a) Overview of the 22nm FDSOI EM-Induced Resistance Shift and the associated BER degradation capture macro, detailing the on-chip metal heater specifics, (b) Measurement top housing the VCO, 30-bit counters & the 32-bit pattern generator block & (c) Illustration of a single DUT group, detailing the layout of interconnects under the heater together with the core and IO circuits local to the group.

The stress drivers (Fig. 3.27 (c)) are wide tri-state, IO-based buffers, to allow for larger stress current densities without the possibility of Bias-Temperature-Instability (BTI) corruption at elevated temperatures during stress mode, when the on-chip heaters are activated. The stress drivers are enabled using a scan-based control individually per

interconnect (ENS_x), to allow for investigating the BER shift as a consequence of EM-induced degradation due to stressing one or multiple wires simultaneously in a datapath or select wires multiple datapaths. Furthermore, 4-wire / Kelvin sensing is implemented per interconnect, with a scan-based control for separately activating the voltage sensing (ENV_x) and / or the current sensing gates (ENI_x) independently, using a total of four shared pads at chip-level. This not only results in the basic 4-wire resistance sensing capability similar to that implemented in the EM-induced resistance degradation macro (section 3.1), but also helps isolating the Joule heating, as a consequence of a desired stress current, by using the stress drivers to force the target current with the sensing gates enabled at a desired ambient temperature. Furthermore, the same configuration can also be used to measure the local DUT temperature with the on-chip heater activated, to allow for experimentally monitoring possible vertical temperature gradients. During stress mode, all the core circuits, except the scan chains used for mode control are powered down to avoid any possibility of corruption at elevated temperatures.

Fig. 3.27 (b) presents details of the measurement top. A five-stage, cross coupled, ring-based VCO is used for generating the on-chip clock, whereas a 32-bit circuit shift register with a scan-based parallel load capability is used to generating the pattern used for measuring the BER of the datapath. A key difference compared to our previous design in section 3.2 is the larger counter size, 30-bits as opposed to 10-bits, to intrinsically allow for larger window sizes, which not only results in a greater statistical confidence, but also allows for the same window size to be used across a range of BER values, minimizing the iterations and resulting in an overall reduced measurement time.

(a)



(b)

Process	22nm FDOI
Core, I/O VDD	0.8V, 1.8V
DUT Type	M3 (Feeder)-M2
Wire Lengths	25µm, 50µm
Wire Widths	Minimum
Datapaths / Chip	10 (x5 = 50 total wires)

Figure 3.28: (a) Implemented metal test-structure detail (b) Chip feature summary.

Fig. 3.28 presents the test structure detail along with the implementation summary. The test-chip is designed in a 22nm FDSOI process, featuring core and IO voltages of 0.8V and 1.8V respectively. The interconnect is a standard EM downstream / upstream test-structure implemented using Metal-3 (M3) for the wide feeders and Metal-2 (M2) for the DUT. Two wire lengths, 25µm and 50µm are implemented per chip, for a total of 10 datapaths, comprising 5 wires each. As before, the feeders are connected using a single-via at the DUT end and multiple vias at driver ends to ensure redundancy. Fig. 3.29 presents the test-chip layout detail.

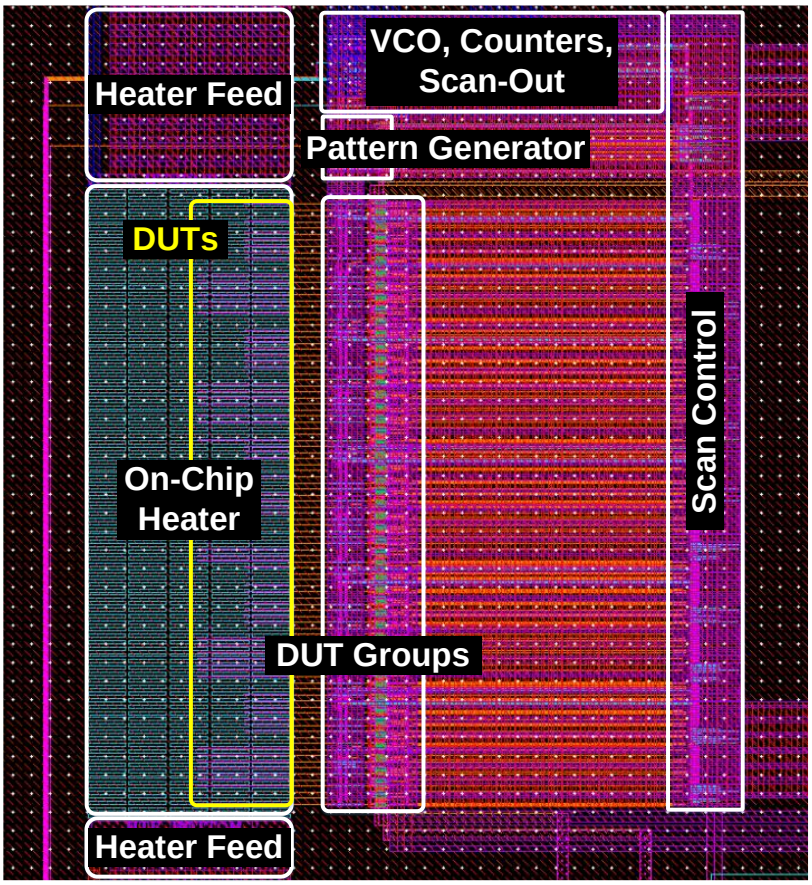


Figure 3.29: Implemented test-chip layout detail.

3.3.3 Conclusion

In this section, a new EM characterization macro featuring a combination of capabilities borrowed from our previous works, together with additional flexibility in mode-control to allow for detailed EM-induced resistance degradation and corresponding BER shift studies, as well as monitoring Joule heating and spatial temperature gradient effects, is proposed and implemented in a 22nm process. At the time of writing this body of research, the test-chip is yet to be taped-out. We hope that the measurement results from this new design will reveal deeper insights, advancing the current state of the art.

3.4 Summary

In this chapter, three array-based hardware monitors meant specifically for characterizing EM effects on-chip were proposed and implemented. The first design featured on-chip stress drivers and 4-wire sensing capabilities to track the wire resistance shift over time under high temperature stress generated using the on-chip metal heater, with the primary goal being the demonstration of a methodology for Median-Time-To-Failure (MTTF) extraction from a large sample population of wires in a limited amount of time in case of a realistic on-chip driver stress scenario, as opposed to conventional industry standard equipment with characterization constraints. The second design proposed monitoring the BER of an interconnect path as a new metric for capturing an EM-induced resistance-shift signature directly in terms of a measured degradation in the signaling capability of the datapath. Detailed measurement results capturing the shift in BER as a consequence of abrupt and progressive resistance changes validated the efficacy of the proposed metric. Finally, in section 3.3, yet another design was proposed, this time the motivation being the translation of a measured resistance change to an accompanying shift in the measured BER of the interconnect path. The new design featured additional flexibility in mode-control, which would allow for Kelvin measurements to be taken in multiple ways, helpful for monitoring effects such as Joule heating for a target stress current or tracking the local DUT temperature with the on-chip heaters activated, while also granting freedom in the selection of one or multiple wires in a datapath for stress further allowing for a detailed EM investigation.

Chapter 4. A High-Density, Array-Based, Soft-Error-Rate & Single-Event-Transient Pulse-Width Characterization Macro meant for Standard Combinational Logic

4.1 Introduction

With the miniaturization of feature sizes leading to increased on-chip device densities together with the reduction in supply voltages, radiation induced soft-errors continue to remain a critical reliability concern even in FinFET technologies where the per-device SER is known to be significantly lower than that of planar CMOS [3], [4], [5]. Single Event Transients (SETs) in combinational logic now become particularly unsettling, as the probability of a narrower pulse getting latched in a sequential element is amplified due to the improvement in circuit delays with scaling [21], [22]. Understanding the key parameter dependencies influencing the soft-error susceptibility of a logic block from a design perspective, is therefore a quintessential step towards engineering robust systems for mission critical applications.

To this end, in this chapter, we introduce a soft-error characterization vehicle featuring SER and SET pulse width measurement capabilities. The proposed design is readily scalable and employs a regular, unit cell based layout featuring standard logic gate variations, enabling circuit designers to directly utilize statistically meaningful test data pertinent to specific logic gate types / path depths, thus obviating the need for indirect

translations from other inaccurate representations ([23], [24]). The test-circuit overhead is small, with the closely embedded measurement circuitry employing local sampling to avoid any pulse-width distortion effects [25], along with the flexibility of fine tuning the pulse width measurement resolution post-silicon.

The next part of this work discusses the SER cross-section statistics mined using the proposed macro, from neutron irradiation experiments conducted at the Los Alamos Neutron Science Center (LANSCE). Multiple dependencies, existing at the device, individual gate and the logic block level, as well as their intricate interaction impacting SER are explored and analyzed. The impact of logic chain length induced electrical masking / pulse quenching effect [26], [27], [28], [29], [30], [31], [32], [33] is broken down in detail, revealing its severity on a chosen gate topology, relative drive strength and the operating voltage. The overall discussion is targeted towards developing an understanding of the factors contributing to SER from a circuit design perspective.

In the final part of this work, we present in detail, the corresponding sampled pulse width distribution data, highlighting the principal utility of the proposed test-vehicle. Several studies have reported pulse width distributions for transistor variants and VDDs at multiple process nodes, for instance [4,5]. This work provides further insight on the impact of key logic design parameters on the SET pulse width distributions in context to the different logic gate types.

4.2 Design Idea & Test Chip Implementation

Although prior works have reported a number of soft error rate (SER) and/or SET pulse width characterization circuits, most of these works suffer from a limitation of one kind or the other (Table 4.1). Several test vehicles either feature a limited device-under-test (DUT) count owing to an irregular design [25] or a large test overhead [34], [23], [24], cause significant pulse width modulation [25] or present analysis on incorrect representations of standard logic [23], [24], which require an indirect translation to be useful to the actual circuits of interest. In this section, we address these shortcomings by demonstrating a high-density soft error characterization vehicle featuring standard unbiased logic gate variations with a regular, tile-able unit cell-based layout, ps-order tunable resolution and an embedded measurement circuit employing local sampling to avoid pulse width modulation effects.

	Skewed Buffer Based [23]	TDC Based [25]	Back - Sampling Chain [24]	This Work
Unit Cell Layout	Regular, Scalable	Irregular, Not Scalable	Regular, Scalable	Regular, Scalable
Area Utilization*	12%	~ 90%	12%	75%
Standard Gates	NO	YES	NO	YES
Chain Length Study	NO	YES	NO	YES
Resolution	< 1ps	> 30ps	> 1.3ps	< 1ps
Tunable Resolution	NO	NO	YES	YES
Pulse Distortion	Immune	Susceptible	Immune	Immune

*Estimate considering identical standard cell layout

Table 4.1: Comparison with prior state of the art.

4.2.1 Proposed Concept

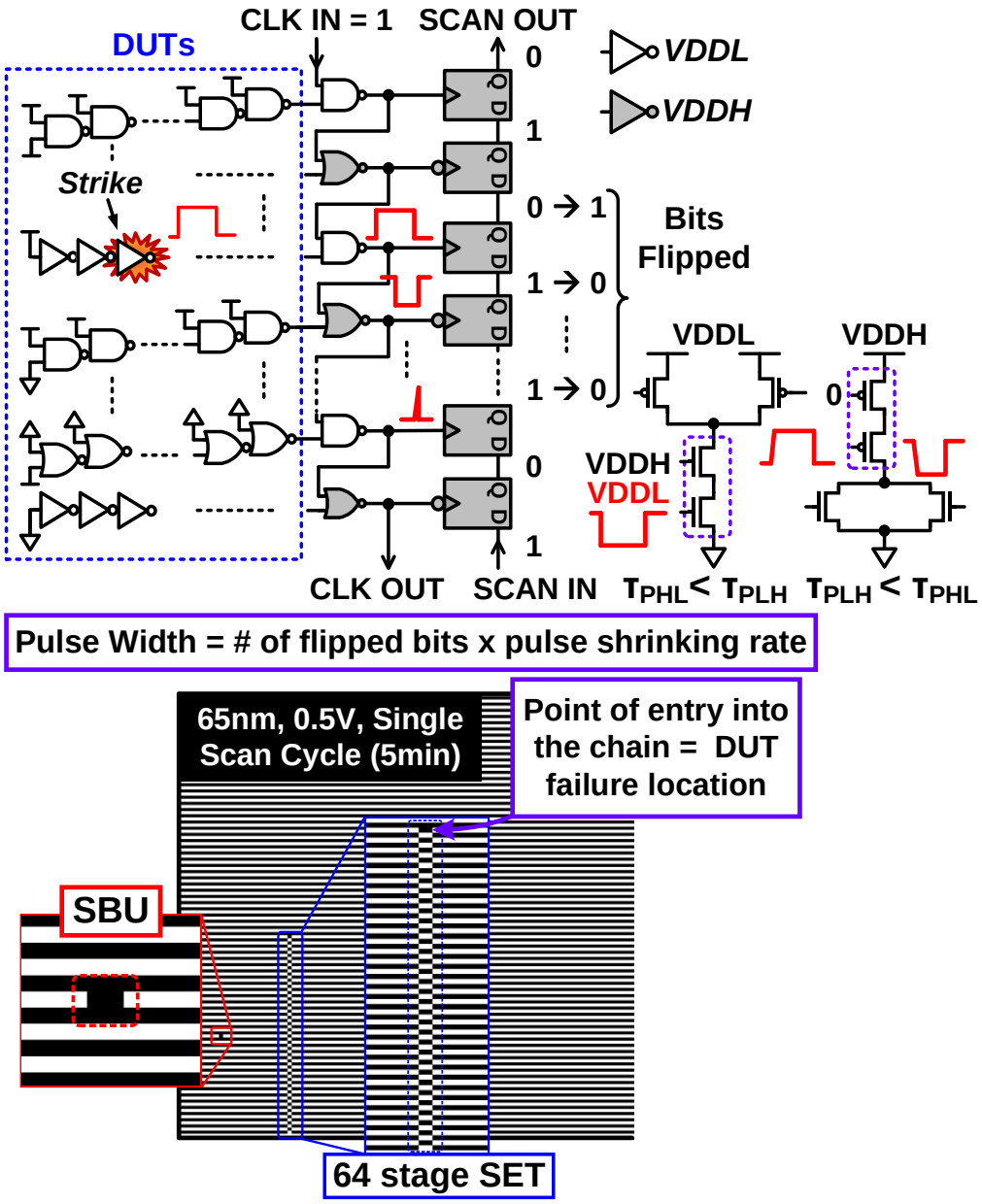


Figure 4.1: Graphical illustration of the proposed concept (top) and array bit-map (bottom) showing an SET and a Single Bit Upset (SBU). The shifted portion of flip-flop bits in the scanned out data reveals the SET injection point in the chain as well as the pulse width information.

Fig. 4.1 (top) conceptually illustrates the proposed logic soft-error characterization macro. The design consists of varying chain lengths of standard combinational logic gates,

essentially the Devices-Under-Test (DUTs) feeding a NAND-NOR chain functioning as clock to a string of serially connected D-Flipflops (DFFs). The NOR gates in the chain together with the DFFs have a slightly higher supply voltage (VDDH) as compared to the NAND gates in the chain and the DUTs (VDDL). This arrangement causes the drive strengths of the stacks comprising the NAND-NOR clock chain to increase, with the consequence that any transient errors entering the clock chain via the DUTs experience a controlled, consistent shrink rate as they propagate downwards along the chain, graphically illustrated in Fig. 4.1. The transient while travelling along the chain also clocks the string of serially connected DFFs in succession, which have been routed in the direction opposite to the clock to prevent any hold time violations. This causes the checkerboard sequence initially programmed into the flops to progressively shift up by one bit, until the transient pulse-width itself becomes less than the minimum detectable pulse width of the DFFs, following which the initially programmed bit sequence remains intact for the subsequent flops. Local sampling of the transient by the closely embedded DFFs prevents any possibility of pulse-width modulation in the test structure. Furthermore, the measurement circuit supply (VDDH) in conjunction with the DUT supply (VDDL) provides the designer with a differential, high resolution post-silicon tuning knob, independent of the technology used for implementation. Unlike the back sampling chain technique [24], a single on-chip matched path (Fig. 4.2) consisting of a MUX De-MUX unit provides for a robust, one-time calibration routine for determining the net pulse contraction for specific values of the two supply voltages, while also cancelling any unwanted (routing, I/O, on-board interconnect) delays. This calibration routine was performed prior to irradiation for the DUT voltages of

interest (VDDL), and the corresponding VDDH values were tuned such that the resulting stage delay values were appreciably higher compared to the measured standard deviation in the per stage delay values obtained from multiple boards. This ensured that any variation induced differences in the per stage shrink rate were insignificant compared to the stage delay itself and that the pulse shrinking was consistently controllable within tolerable margins for the transient error studies.

Fig. 4.1 (bottom) shows the experimentally obtained array bit map corresponding to a single scan run highlighting the SET, manifested as a trail of shifted DFF bits with the point of entry into the NAND-NOR chain and the end points demarcated by consecutive '1's or '0's and the single bit upset (SBU), manifested as an isolated bit flip. The collective failure location statistics obtained from thousands of such scanned bit maps at multiple supply voltages constitute the SER specifics for the different DUT types, which is the focus of the second part of this work. To summarize, the proposed characterization vehicle presents a highly scalable solution, offers pico-second order post silicon resolution trimming, employs local sampling to avoid any pulse width distortion, and facilitates statistically meaningful characterization of both SER and SET events for the relevant circuits of interest.

4.2.2 Implementation Specifics

Fig. 4.2 & 4.3 presents the test chip die micrographs and the implementation specifics. The circuit architecture is based on a unit tile-able cell, schematically illustrated for the 65nm implementation in Fig 4.3, consisting of the three standard logic gate types with both the PMOS and NMOS devices identically sized: either 1X or 2X and

implemented in two V_{TH} flavors: Regular V_{TH} (RVT) or High V_{TH} (HVT) leading to total12 unique chains in 65nm and 16nm, with unit cells of varying chain lengths: 16, 32 and 64 distributed uniformly to cover the complete die area for the 65nm implementation. Non-minimum sized Low V_{TH} (LVT) devices were used to improve the NAND-NOR chain and DFF resiliency to soft-errors. Presented later in Fig. 4.9, irradiation results on a dummy version comprising only the NAND-NOR chain and DFFs show negligible soft errors in the read out circuit.

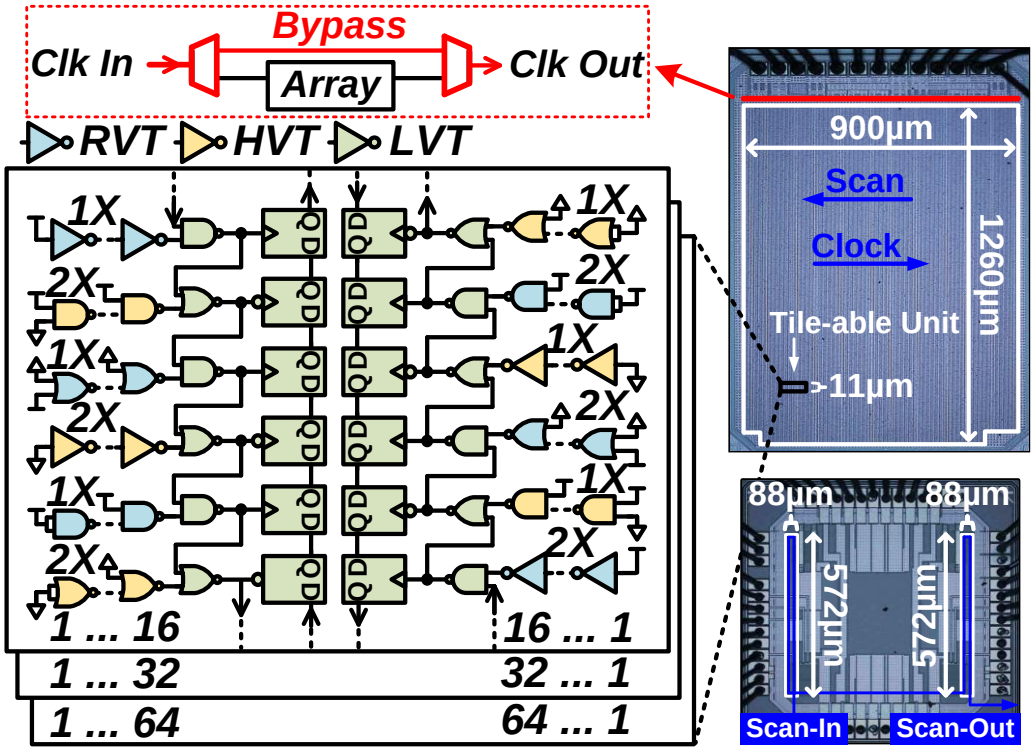


Figure 4.2: Die micrographs together with the schematic illustration of the unit tile-able cell.

The 16nm version consists of chains of fixed lengths: 128 for Inverters (INV) and 96 for NAND and NOR gates and is a demonstration of the easy process portability of the proposed design-concept. With a negligible test area overhead, the design can be easily

accommodated in constrained spaces making it ideal as a scribe line test-structure, for potential use as a technology characterization vehicle. Two isolated arrays can be tiled together simply by daisy chaining the scan and clock, as can be seen from the 16nm die micrograph. Furthermore, the I/O requirements for implementing the design are also modest, limited to a total of six pads.

Process	65nm GP CMOS
Core VDD	1.0V
Irradiation VDD	0.45V – 1.0V
Core + Test Area	1.26mm x 0.9mm
Gate Count	0.53M
Device Widths	120nm(1X), 240nm (2X)
Gate Flavors	RVT, HVT
Chain Lengths	16, 32, 64 Total = 16,272

Process	16nm FinFET
Core VDD	0.8V
Irradiation VDD	0.4V
Core + Test Area	572µm x 177µm
Gate Count	0.419M
Number of Fins	2 Fins (1X), 4 Fins (2X)
Gate Flavors	RVT, HVT
Chain Lengths	INV = 128, NAND/NOR = 96

Figure 4.3: 65nm and 16nm feature implementation summaries.

4.3 Neutron Irradiation Setup

As shown in Fig. 4.4 (a), nine 65nm boards and five 16nm boards were stacked vertically and subjected to spallation neutron beam irradiation. Each 65nm board consists of 5 chips (total 45 chips, 24M gates) placed within a beam diameter of 2” while each 16nm

board has one chip (2M gates). The chips on each board are serially connected through a single scan data and single scan clock signal. The chips were exposed to a neutron beam for five effective days. The beam energy ranged between 1.38 - 750MeV while the average beam flux was 2.03×10^6 neutrons/cm^2/sec.

Fig. 4.4 (b) shows the FPGA based control used for digital I/O (DIO). Following a one-time initialization procedure, a checkerboard map (752Kbits) was written to all 50 chips simultaneously while reading out the stored data in parallel. This was then followed by irradiation for 5 minutes, with the cycle continuously looped.

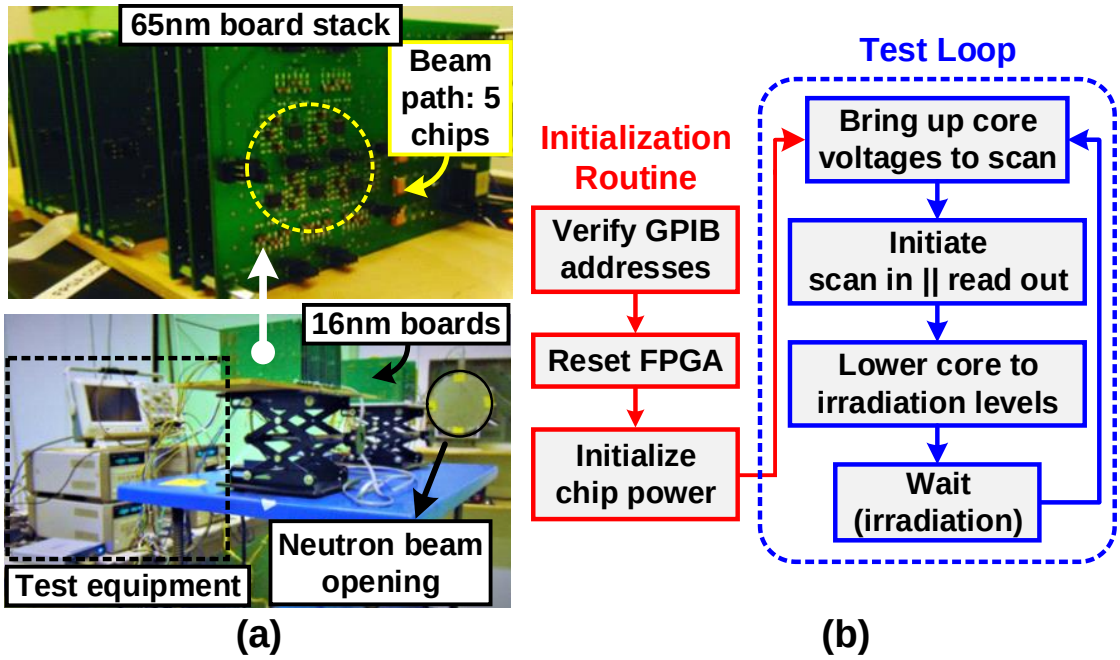


Figure 4.4: (a) Neutron testing details at LANSCE and (b) FPGA based test-flow.

4.4 Technology Analysis

In this section, we present simulation results of key 65nm circuit metrics, insightful in understanding the SER data shown in section 4.5. These simulations were performed on

the Typical-Typical (TT), Fast-Fast (FF) and the Slow-Slow (SS) process corners, however, the results presented in this section correspond to the TT process corner, as they were found to correlate well with the measured data trends.

4.4.1 Restore Current Simulations

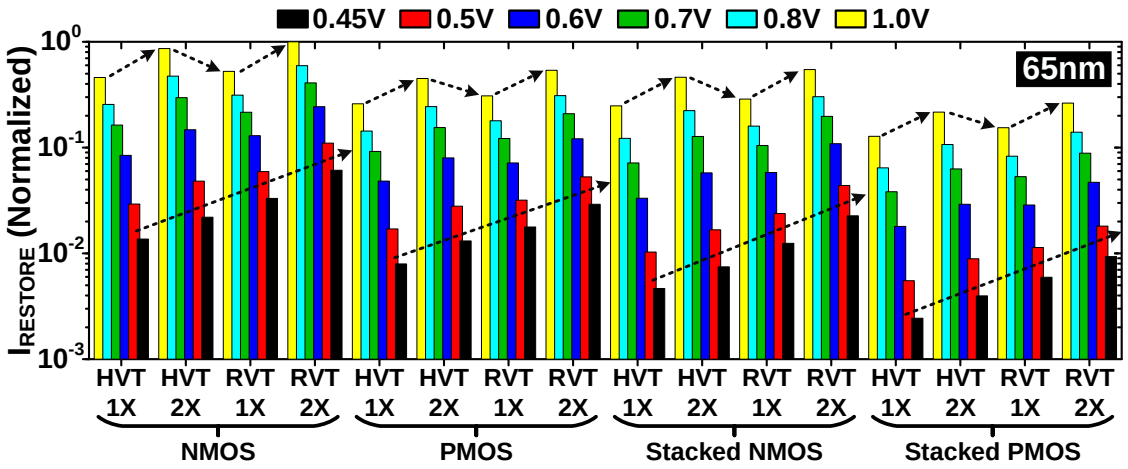


Figure 4.5: Simulated $I_{RESTORE}$ for PUN and PDNs for the different gate topologies and different supply voltages.

Fig. 4.5 presents the simulated transistor restore current $I_{RESTORE}$ for the different pull-up network (PUN) and pull down network (PDN) topologies corresponding to standard gate variants implemented in the 65nm test-chip. Due to the low mobility of p-channel devices in this process, the PUNs exhibit 0.41~0.59X lower drive current compared to their PDN counterparts. Stacked topologies exhibit lower $I_{RESTORE}$, with the effect compounded at low voltages (0.45V, 0.5V) and for PMOS devices. From Fig. 4.5, it can also be observed that the HVT devices are especially vulnerable at low voltages, with even the 2X sized devices displaying lower drive currents compared to the 1X RVT versions. As the supply voltage increases beyond half VDD, the increase in gate overdrive offsets this effect, with the 2X HVT flavors now exhibiting a higher drive current than 1X

RVT flavors as expected. The above observations will be useful in analyzing the experimentally obtained results, discussed in subsequent sections.

4.4.2 Critical Charge Analysis

Fig. 4.6 details the simulated critical charge (Q_{CRIT}) trends, insightful in revealing the collective SER susceptibility for the standard gate types. We define Q_{CRIT} as the minimum amount of charge required to generate a transient at the output node of the subsequent gate in a logic chain with a peak amplitude of 0.5 times the DUT supply voltage for the simulation. This definition accounts for the drive strengths of the PUN or the PDN of interest for the simulation, the gate output loading as well as the switching threshold of the subsequent stage. The SER sensitivity pertaining to the PUN and the PDN for a gate type is analyzed separately using a current source model in Verilog-A following the double exponential transient shape for modeling both the high-going (with τ_r = 10ps, τ_f =100ps [35], [36], [37], [38], [39], Fig. 4.6(a)) and the low-going (with the source polarity reversed, Fig. 4.6(b)) transients. The simulations were performed on the post-layout circuit netlists of the unit tile-able block, to carefully account for the layout introduced parasitics, which become crucial elements for determining the SER vulnerability of a node particularly in the low voltage regime. No impact on the relative Q_{CRIT} distributions was observed for sweeping the rise (5ps to 15ps) and fall (50ps to 150ps) time constant values [35]. From our perspective, the relative Q_{CRIT} distributions among the different gate types are more relevant to this work in comparison to accurately estimating the absolute Q_{CRIT} magnitudes or the corresponding SER cross-sections, since we are interested in understanding the SER susceptibility of a gate type relative to others from a circuit design standpoint. Although

the simple double exponential model used for this analysis systematically overestimates the Q_{CRIT} / SER cross sections by 10-15% [35], its use still results in a reasonably accurate estimation of the relative gate susceptibility, as evinced from a good match between the simulated trends and the measured data.

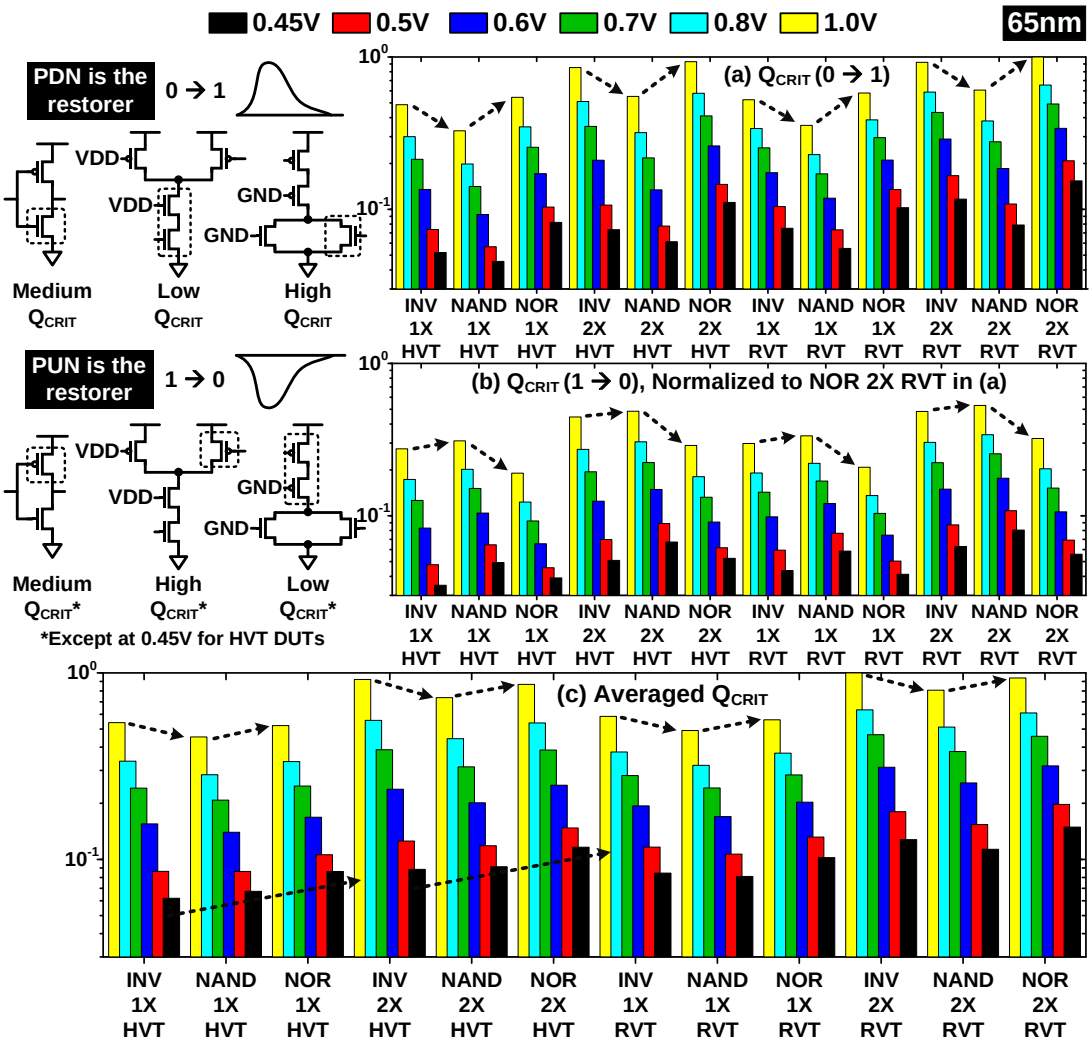


Figure 4.6: Simulated Q_{CRIT} trends: (a) 0→1 Q_{CRIT}, (b) 1→0 Q_{CRIT}, (c) averaged Q_{CRIT}, assuming a uniform strike rate for both '0' and '1' nodes.

Both the 0→1 and 1→0 simulations follow the expected trends when accounting for the $I_{RESTORE}$ and node capacitance as shown in the circuit schematics in Fig. 4.7. For

instance, networks featuring stacks would typically require lower Q_{CRIT}, due to the lower $I_{RESTORE}$. Among unstacked versions, gates having extra node capacitance (viz. the off transistor in NAND PUN or NOR PDN) would require higher Q_{CRIT} than others (INV). The intensity of this trend also varies with supply voltage and the device flavor, sometimes resulting in deviations from expected norm e.g. at 0.45V for PUNs of the HVT variants, with the node capacitance becoming a dominant factor at lower supply especially for these weak flavors.

We also introduce another metric, the averaged Q_{CRIT} of the 0→1 and 1→0 transitions in Fig. 4.6(c), useful for estimating the overall SER susceptibility of a gate relative to others. With the high and low nodes uniformly distributed throughout the test structure for all the logic chains of the standard gate variants, the averaged Q_{CRIT} in fact provides a reasonable first-order estimate for the SER sensitivity of a gate type relative to others, not accounting for effects such as electrical masking or pulse quenching, exposed diffusion areas, process variations etc. The simulated Q_{CRIT} trends presented in this section will be analyzed in further detail alongside corresponding experimental data in the subsequent section.

4.5 SER Results and Dependencies On Key Parameters

This section presents the measured SER cross section, defined as the number of errors for a given DUT type divided by the fluence, expressed in neutrons/cm^2, for the entire duration of exposure at the specific circuit setting. For the irradiation voltages presented in this work, the fluence values ranged between $1.134 \times 10^{11} – 4.189 \times 10^{11}$ neutrons/cm^2. Furthermore, to establish the uniformity of flux between individual boards,

the SER data corresponding to each of the dependences presented in the subsequent sections was also analyzed board-wise. No regular trend indicating a possible loss in the flux uniformity due to stacking of multiple boards was observed (Fig. 4.7).

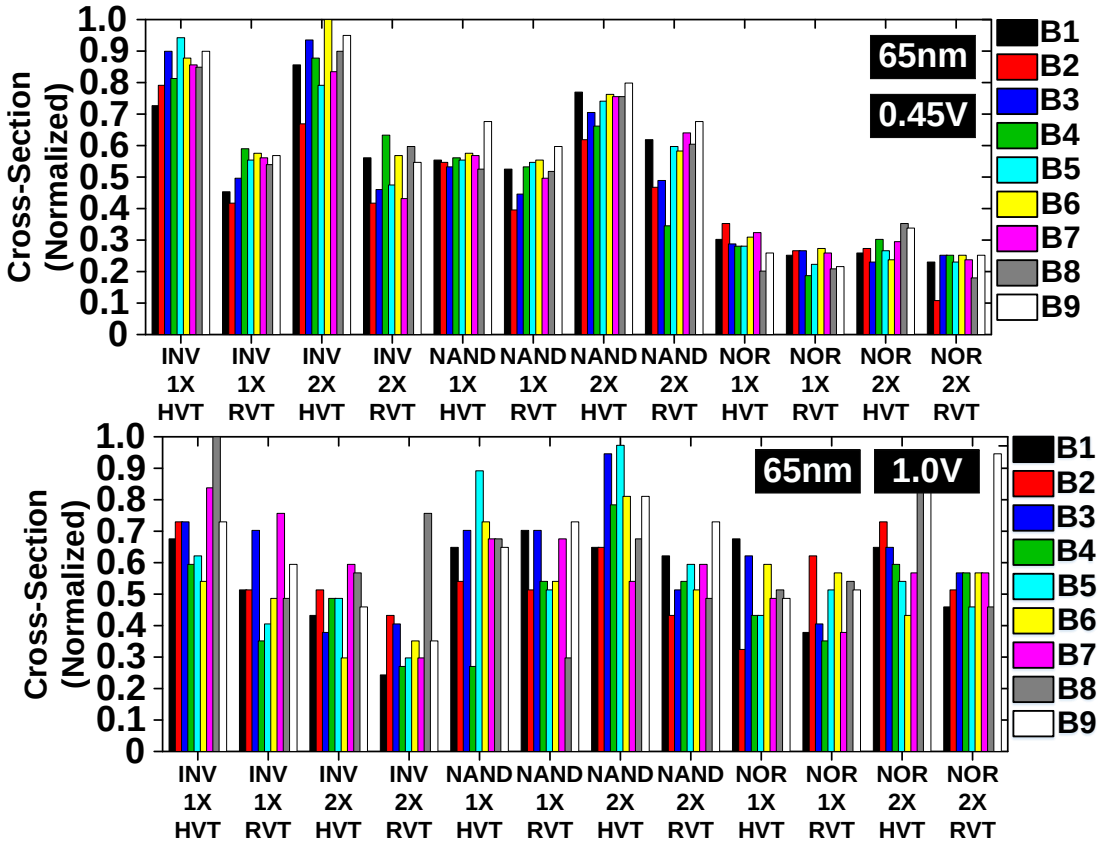


Figure 4.7: Logic gate data arranged board wise, corresponding to figure 4.10 (V_{TH} dependency) presented for two distinct voltages. Where, (B1 – closest to the beam to B9 – furthest away from the beam) in the beam path at two voltages (0.45V (near threshold voltage) and 1.0V (nominal voltage). No regular pattern trends between board-to-board data were observed.

4.5.1 Supply Voltage Dependence

Fig. 4.8 shows the measured cross-section for the different standard gate flavors considered together, along with SBUs in the embedded DFFs comprising the measurement circuit at multiple supply voltage levels. The failures go up exponentially as the supply voltage goes down [34], [40], [41], [42], witnessing the highest increase (i.e. 7.1X)

considering the case of INVs, with the errors sharply increasing as the supply voltage scales down to near-threshold voltages. This increase, compared to other gate types, is due primarily to the stronger node capacitance dependence as well as a heightened sensitivity to process variations, further pronouncing pulse quenching effects at low VDDs [27], both of which become deciding factors for soft error vulnerability in the low supply voltage regime, with the $I_{RESTORE}$ becoming a weaker factor. The pronounced impact of node capacitance on the SER vulnerability among the standard gate types is clearly reflected in the Q_{CRIT} simulation results in Fig. 4.6 at lower supply voltages, with the $0\rightarrow1$ results demonstrating a substantially higher Q_{CRIT} for all the NOR gate variants as compared to the INV types, despite both featuring logically equivalent pull down networks. The impact is also faintly seen in the $1\rightarrow0$ results, with the weaker $I_{RESTORE}$ flavors (1X HVT, 2X HVT) still exhibiting a higher Q_{CRIT} for the NOR gates compared to the INV types, contrary to what one would expect, given that the gate features the weakest pull-up network among all. For stronger variants (1X RVT, 2X RVT), the increase in $I_{RESTORE}$ offsets the node capacitance dominion, which is clearly demarcated by the higher Q_{CRIT} for the INV types. In addition to the node capacitance, digital gates become particularly sensitive to process variations in the low voltage regime, with the susceptibility dependent on the active transistor count directly influencing the propagation delays.

At higher voltages, $I_{RESTORE}$ becomes a dominant factor overcoming both node capacitance and process variation effects. This is evinced by the higher SER for NOR gates and the lowest SER for INV types in Fig. 4.8, even when considering all the gate types collectively. This is because the NOR gate features the weakest stack (PMOS) among all

types, a result again evident from the Q_{CRIT} 1→0 results, demonstrating substantially lower values for all variants of the NOR type at higher supply levels. It should also be noted that the 0→1 Q_{CRIT} results at nominal indicate only marginally higher values for the NOR gates, indicating the much weaker contribution from the node capacitance as compared to lower supply values. These dependences will be broken down in further detail in the subsequent sections.

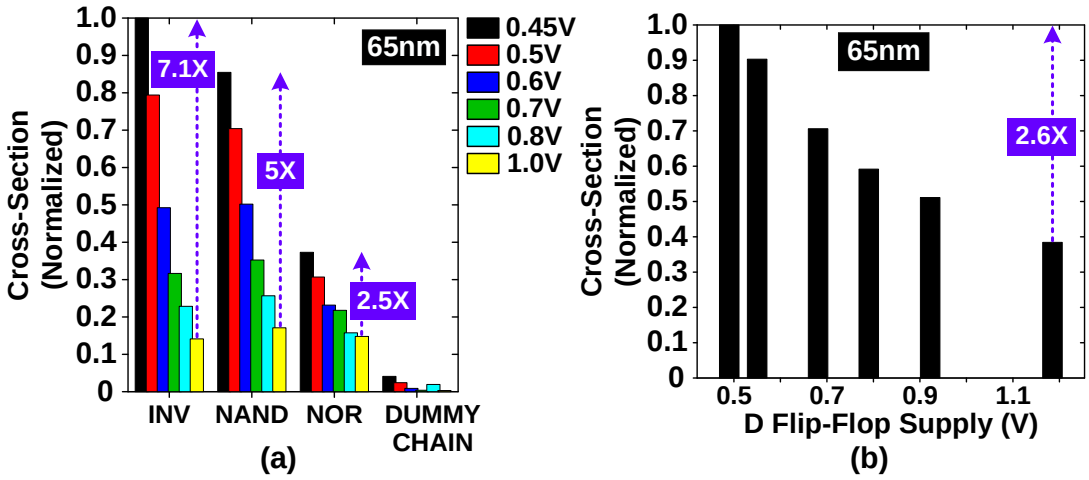


Figure 4.8: Measured cross-section at multiple supply voltages for (a) the different standard gate variants considered together and (b) SBUs in D-Flipflops.

For SBUs in DFFs, a 2.6X increase is observed for a 695mV difference (1.185V to 0.49V) in the supply voltage, which is attributed not only to the non-minimum sized LVT transistors used for designing these circuits, but also the interconnect load (total 8um, horizontal + vertical) at the output node for routing to the D node of the subsequent DFF in the vertical arrangement. No MBUs were observed even at lower VDDs, for instance, 0.45V for the 65nm technology. This can be attributed to the relatively large center-to-center (1.72µm, vertical between rows; 8.1µm horizontal between the two adjacent columns) distance between the flops (see [34] for a comparison in a similar bulk process

to SRAM bit cells, with MBU occurrences strongly correlated to the layout) in conjunction again, with the flop design itself.

One limitation of the circuit, however, is that it cannot distinguish between SETs and Single Event Multiple Transients (SEMTs). If multiple transients are induced in the logic chains that are both spatially and temporally adjacent, they might overlap and be counted by the circuit as one single transient. However, the probability of this occurrence would be lower as compared to the SETs themselves, especially in the 65nm process node.

4.5.2 Threshold Voltage (V_{TH}) Dependence

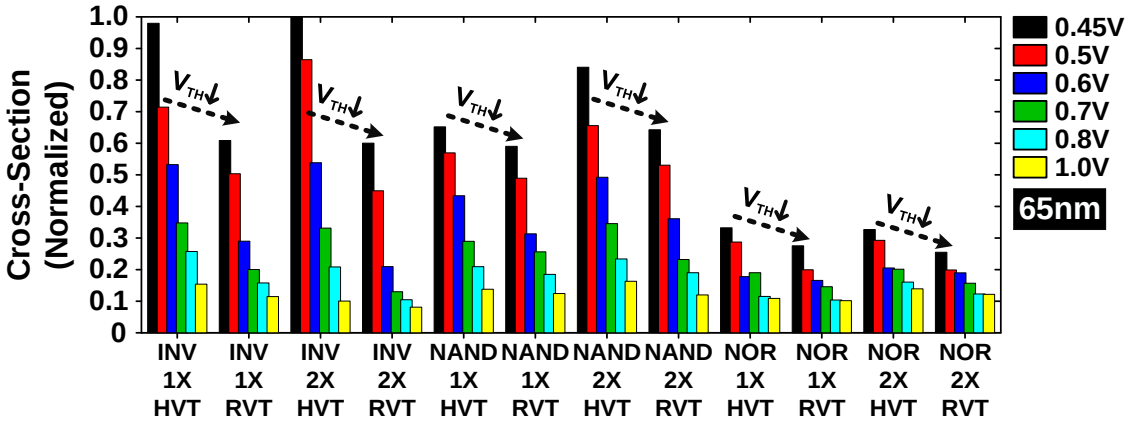


Figure 4.9: Measured cross-section results for HVT DUT types vs. RVT DUT types, with the V_{TH} dependence highlighted.

Fig. 4.9 showcases the measured SER cross-section results for HVT and RVT flavors of the standard gate types placed side by side. With other parameters (such as node capacitance, gate topology, supply voltage and device size) being identical, the HVT variants demonstrate a heightened SER sensitivity, due primarily to the lower $I_{RESTORE}$ [43], [44] with the trend being consistent for each gate type at all supply voltage levels. The impact of this dependence becomes more pronounced at low voltage levels (0.45V, 0.5V)

due to the relative difference in the overdrive voltage ($V_{OV} = V_{GS} - V_{th}$) between the two variants, which gradually builds up as the supply voltage is lowered. This trend is clearly reflected from the simulated results in Fig. 4.5. Simulation results in Fig. 4.6 also show a 1.3-1.5X higher Q_{CRIT} for the RVT variant, supporting the experimental observation.

While all gate types consistently demonstrate a heightened SER sensitivity for the HVT variants at all voltage levels, the relative change in the SER magnitude as a result of V_{TH} change among different gate types depends on the change in average $I_{RESTORE}$ (PUN and PDN combined) in going from a non-stacked gate type viz. INV to a stacked gate type (NAND featuring NMOS stack, NOR featuring the PMOS stack) when considering a device size (1X or 2X) for comparison at a particular supply voltage. The change in $I_{RESTORE}$ is not perfectly linear between unstacked and stacked versions (~2X resistance) for a change in variant type from HVT to RVT depending on supply voltage and is particularly difficult to analyze in the presented data, not only due to averaging effects resulting from a stacked PUN and an unstacked PDN (NOR) or vice versa (NAND) dictating the overall gate susceptibility, but also due to the contamination resulting from electrical masking effects (presented in the subsequent section).

4.5.3 Logic Chain Length Dependence

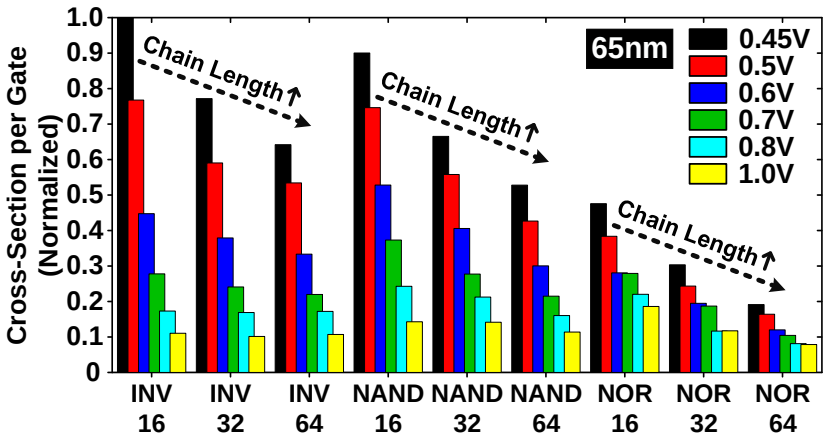


Figure 4.10: Collective cross-section results for multiple logic chain lengths.

Fig. 4.10 illustrates the impact of logic chain length on the measured SER cross-section (normalized to gate count) for the standard gate variants considered collectively. With the increase in logic depth, electrical masking or pulse quenching effects [26], [27], [28], [29], [30], [31], [32], [33] begin to dominate, causing the pulse to shrink and disappear within the chain. The trend becomes particularly noticeable at lower supply voltages [27], on account of (i) exponentially higher delays or delayed response time and (ii) increased sensitivity to process variations, which have a marked impact on the gate propagation delays. A transistor operating in the slow process corner for a gate in the logic chain results in further increase (and mismatch) in delays, causing significant pulse attenuation and impeding further propagation, the probability of which compounds with both the logic distance travelled as well as the active device count directly determining the delay for the gate topology. Especially evident at 0.45V for the three gate topologies, the effects gradually taper off with the increase in supply voltage. Among the standard gate topologies, INV DUTs show little to no logic depth dependence for higher supply voltages (0.8, 1.0V), whereas gates featuring stacked devices still show the dependence with the

trend being especially noticeable for the NOR gates. Owing to lower mobility of p-channel devices in 65nm and identical transistor sizing (PMOS:NMOS=1:1) in all gate types, the NOR gate PMOS stack is the weakest of all PUNs / PDNs among the different gate types even at nominal supply voltages and is thus, still susceptible to electrical masking effects especially in longer chains.

Fig. 4.11 presents the measured SER cross section results capturing the detailed impact of logic depth for the three gate topologies. Arranged in order of increasing drive strength at near-threshold voltages (NTV) (0.45V, 0.5V, see Fig. 4.5), measured data showcases the impact of $I_{RESTORE}$ on the susceptibility of a gate variant to logic chain length induced electrical masking effects, with the stronger $I_{RESTORE}$ variants exhibiting little to no dependence even at lower voltages and weaker $I_{RESTORE}$ variants demonstrating the dependence all the way up to 1.0V. For instance, for INVs, the effect persists from 0.45V to 0.7V for the weakest $I_{RESTORE}$ variant (1X HVT), while stronger variants such as 1X RVT and 2X RVT demonstrate the dependence at only very low supply levels (0.45V, 0.5V). NAND and NOR gates are particularly prone to these effects on account of stacked devices which not only lead to larger delays, but also mismatch in the τ_{PHL} (high-to-low delay) and τ_{PLH} (low-to-high delay) values, resulting in pulse attenuation in both height and width [27], [28] as the transient travels along the chain. This is clearly reflected in the measured data, with the stronger NAND gate variants (1X, 2X RVT) demonstrating the dependence upto 0.7V, whereas NOR gates featuring the weak PMOS stack exhibit this

dependence even at nominal supply level (1.0V) for both stronger and weaker flavors alike.

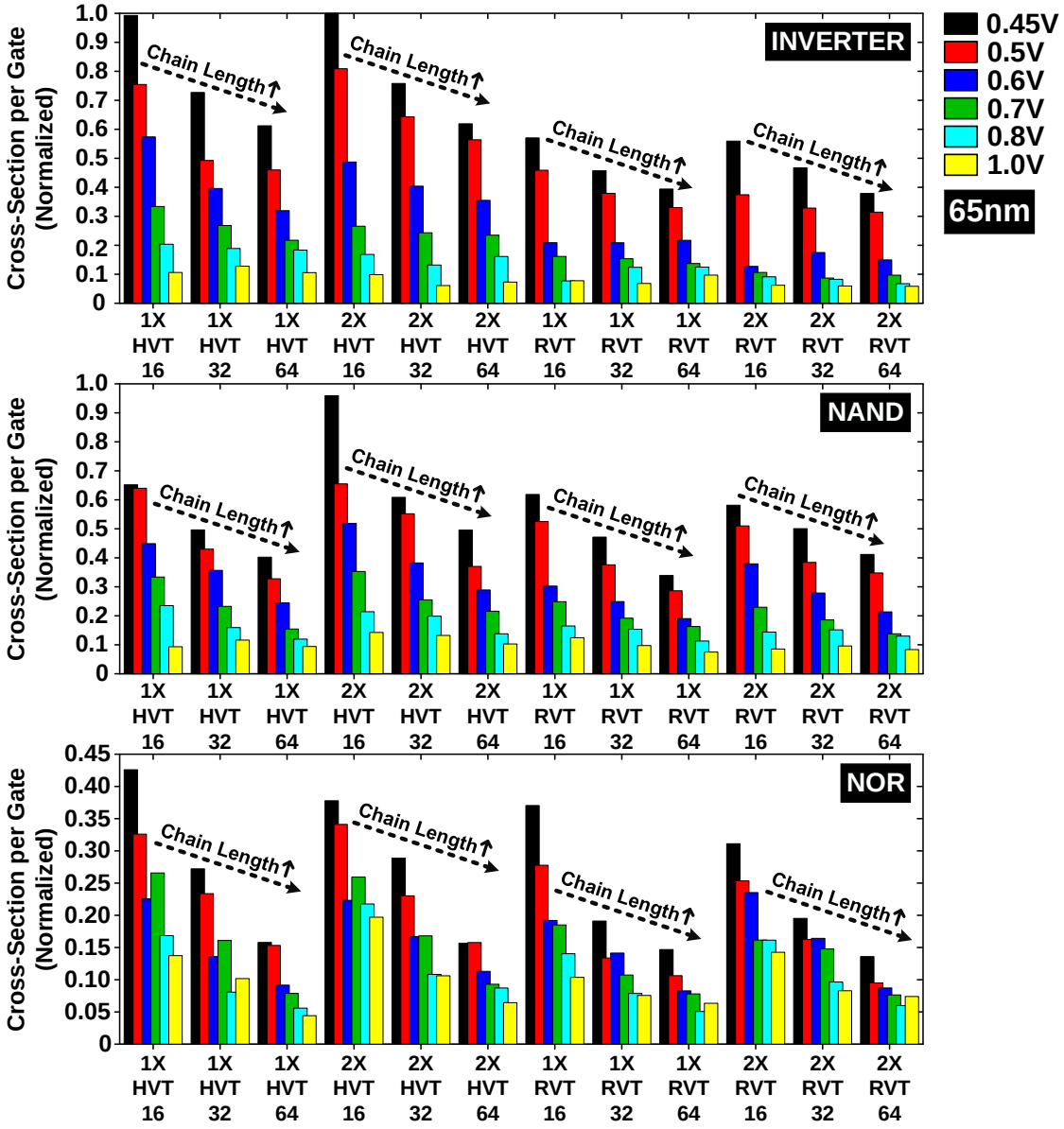


Figure 4.11: Measured cross-section results showcasing the detailed impact of logic depth for variants corresponding to the standard gate topologies.

4.5.4 Gate Type Dependence

Fig. 4.12 shows the measured cross section data for the different DUTs arranged gate-wise. The SER in the different gate types can be understood by analyzing the low

voltage (0.45V, 0.5V) and nominal voltage (1.0V) behaviors for the weakest and the strongest $I_{RESTORE}$ sets (Fig. 4.12). As discussed in previous sections, among gates having the lowest $I_{RESTORE}$ (1X HVT), the dominant factors contributing to SER are the node

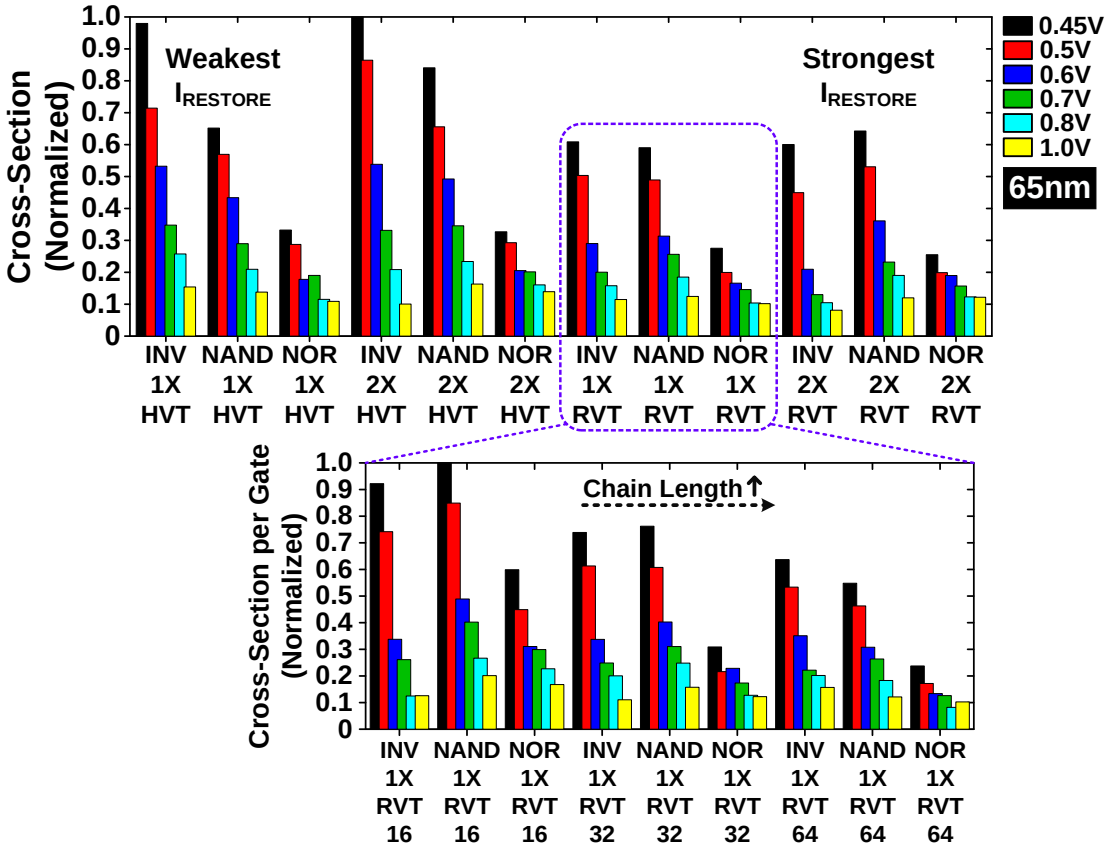


Figure 4.12: Measured cross-section data indicating the logic gate type dependence.

capacitance and the susceptibility to process variations, with both dependences amplified at low voltages, where the contribution from $I_{RESTORE}$ is the weakest (Fig. 4.5, 4.6). Therefore, in the low voltage regime, the gate with the lowest capacitance (INV) and the one least susceptible to process variations (again, INV) should exhibit higher failures compared to other types, especially for the weaker $I_{RESTORE}$ flavors. This trend is clearly reflected in both the measured data, Fig. 4.12, as well as the mean Q_{CRIT} results, presented

in Fig. 4.6(c) corresponding to the HVT variants (both 1X and 2X). NOR gates are especially robust at lower voltages, attributed in part to the strong pull down network featuring an extra capacitance (the tendency reflected in Fig. 4.6(a), as the substantially higher Q_{CRIT} needed for flipping the node state at 0.45V, 0.5V compared with other gates types) and in part to their higher susceptibility to electrical masking effects (Fig. 4.10, 4.11). The increase in $I_{RESTORE}$ for the stronger RVT variants at NTV offsets the impact of node capacitance, an effect evinced in both measured data (increased failure rate for NAND gates as compared to INV due to lower $I_{RESTORE}$ in the stacks) as well as simulation results (Fig. 4.6(b), where INVs exhibit a higher Q_{CRIT} relative to NOR gates for RVT flavors, indicating the dominance of $I_{RESTORE}$ over the node capacitance). At higher voltages (0.8V, 1.0V), INVs show lower SER cross-section, which can be especially seen for the case of stronger $I_{RESTORE}$ variants (2X HVT, 2X RVT, see Fig. 4.5 for the $I_{RESTORE}$ trends at these voltages). This is an expected result (confirmed also by the mean Q_{CRIT} trends in Fig. 4.6(c)), since $I_{RESTORE}$ now becomes the single most dominant factor determining the gate robustness at nominal levels, largely offsetting both process variations as well as capacitive loading effects and rendering the gates employing stacks more vulnerable to errors. Although the mean Q_{CRIT} data predicts this tendency for both weaker and stronger gate variants (with pronounced prediction for stronger types), it should be kept in mind that these simple post-layout simulations do not account for the other effects such as pulse quenching in longer chains, which prevent the generated SET pulses in weaker gate variants from reaching the measurement circuitry. To illustrate this point, we present the detailed breakdown for the 1X RVT variants, arranged gate-wise for the different chain

lengths, clearly showing the expected trend obtained from Q_{CRIT} analysis for the shorter chain length (lowest SER for INVs), as well as the contamination at both NTV and nominal levels, progressively creeping in as the chain length increases to 4X. It should also be noted that the un-normalized counts from longer chain lengths would also be higher and hence, would contribute fairly in shaping the overall collective trend. This discussion also explains why the measured results from the weakest $I_{RESTORE}$ (1X HVT) variants show higher SER for INV DUTs even at nominal, contrary to the Q_{CRIT} predictions. In general, a complex interplay between supply voltage, node capacitance, $I_{RESTORE}$ (affected by device V_{TH}, width and stacking) as well as logic distance travelled to the sequential element determine the relative SER susceptibility among the standard logic gate types.

4.5.5 Width Type Dependence

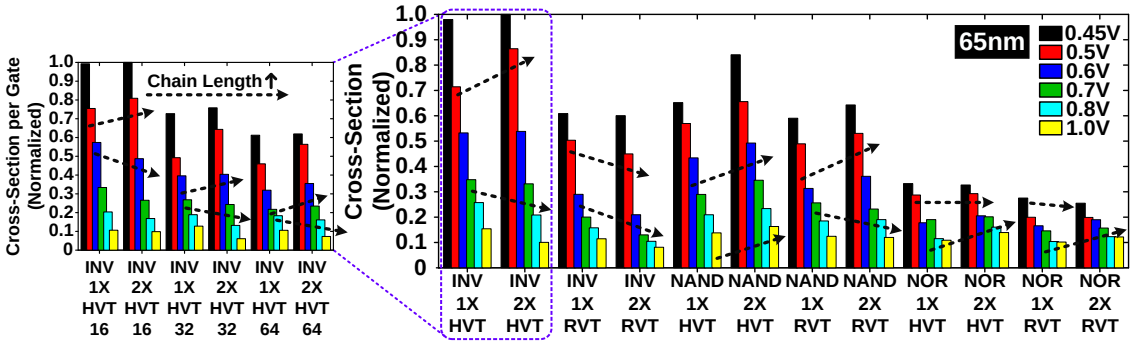


Figure 4.13: Cross-section results highlighting the detailed device width dependence.

Fig. 4.13 presents the measured SER for the different DUT types arranged width-wise. On account of having a lower $I_{RESTORE}$ (Fig. 4.5), a higher SER would typically be expected for the 1X variants. However, the comparatively small feature sizes used for implementing both 1X as well as 2X flavors render them susceptible to process variations, particularly at low supply voltages [27], leading to increased electrical masking, as seen

for the 1X variants. As discussed, this effect is magnified for longer chains, for weaker $I_{RESTORE}$ (HVT) variants and low VDDs. The trend is clearly reflected in the test data for the INV DUTs, where slightly higher SER values are observed for the 2X HVT flavor at NTV levels, with detailed chain length breakdown showing the progressive increase in masking effects for longer chains, shaping the overall SER. With the increase in supply voltage, the increased gate overdrive offsets such effects, resulting in the expected trend beyond 0.6V. RVT flavors show the expected trend due to the stronger $I_{RESTORE}$. However, as discussed previously, on account of having stacked devices, the NAND/NOR gates are far more vulnerable to electrical masking effects. This can be seen for HVT variants of the NAND gate, with the 2X flavors exhibiting a higher SER all the way up to nominal. The RVT variants show the expected trend for the 2X DUTs, albeit for supply voltages of 0.7V and beyond, again attributed to increased drive current. NOR gate types being particularly susceptible due to the weak PMOS stack, show slightly higher error rate for even the 2X RVT variants at nominal voltages, with the NTV behavior exhibiting similar to slightly higher SER for the 1X flavors of both V_{TH} variants due to slight improvement in the weak $I_{RESTORE}$ for 2X devices coupled with similar susceptibility to masking effects at low voltages. In addition to these effects, the exposed diffusion area for larger widths might also play a role in their susceptibility to SER [45].

4.5.5 Technology Comparison: 65nm Planar Vs. 16nm FinFETs

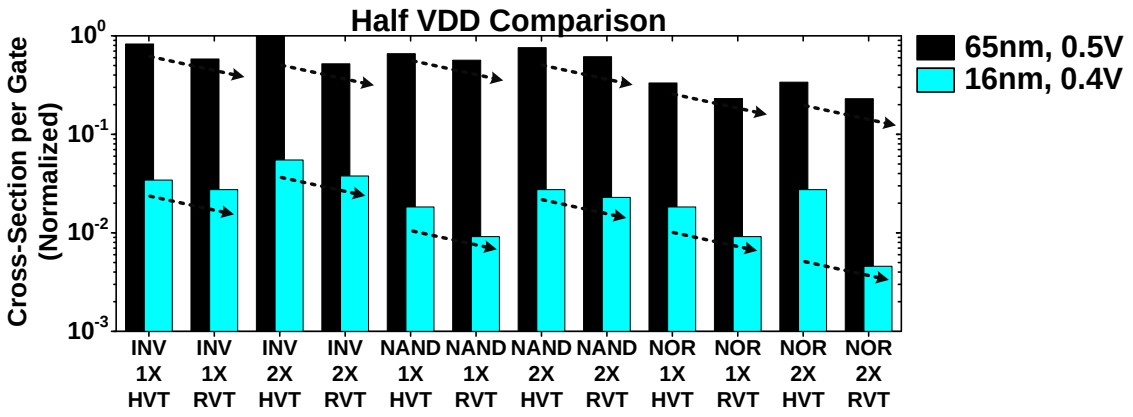


Figure 4.14: Measured cross-section results for 65nm planar CMOS and 16nm FinFET DUTs at technologically similar bias points.

Fig. 4.14 presents the measured SER cross-section (normalized to gate count), for the standard gate types implemented in 65nm planar bulk CMOS and 16nm FinFET processes, with supply voltages in both cases at half the nominal VDD. Previously reported in literature [3], [4], [5], the marked resiliency of FinFETs is clearly evident in our results with standard gate variants exhibiting nearly a 15-60X lower SER compared to their planar equivalents. Although an exact comparison is difficult due to second order effects caused by implementation differences such as logic chain lengths and input connections, measured data still clearly reflects some of dependences discussed in previous subsections.

4.6 Analysis on the Neutron-Induced SET Pulse Width Distribution of Logic Gates

As discussed in section 4.2, an SET transient in the logic gate DUTs, enters the NAND-NOR clock chain, experiencing a picosecond-order bias controlled, uniform shrink rate as it propagates downwards along the chain, while also clocking the string of serially

connected D-Flipflops (DFFs) routed oppositely. This causes the initially programmed checkerboard sequence in the array bit map (Fig. 4.2) to shift up by one bit until the SET pulse width falls below the capture resolution of the DFF (Fig. 4.15). Although the DFF resolution can be improved to beyond the limit imposed by the process technology, for instance, by skewing the clock pulse to expand within the DFF, it would not be representative of a real test-case and construe unnecessary overdesign. Hence, the proposed test vehicle takes the corresponding sequential element resolution into account, with the DFF acting as a high pass filter, rejecting any SETs below its capture resolution, as in a realistic scenario. The sampled pulse widths in the following sections are obtained by

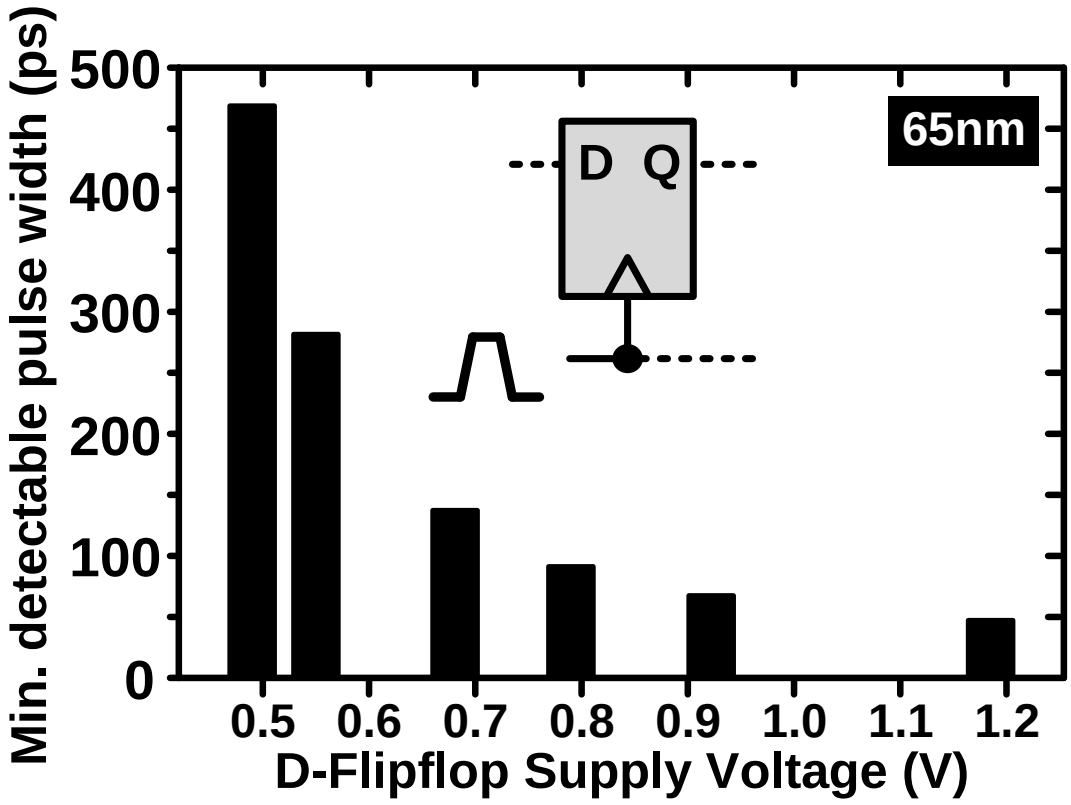


Figure 4.15: Post-layout simulations indicating the minimum detectable pulse width of a D-flip-flop at various VDDs.

multiplying the calibrated stage delay corresponding to a specific VDD setting with the number of flipped DFF stages in the chain. We now present an analysis on the impact of the key logic design parameters impacting the pulse-width distributions in the subsequent subsections similar to that presented for the SER trends previously.

4.6.1 Supply Voltage and Gate Type Impact

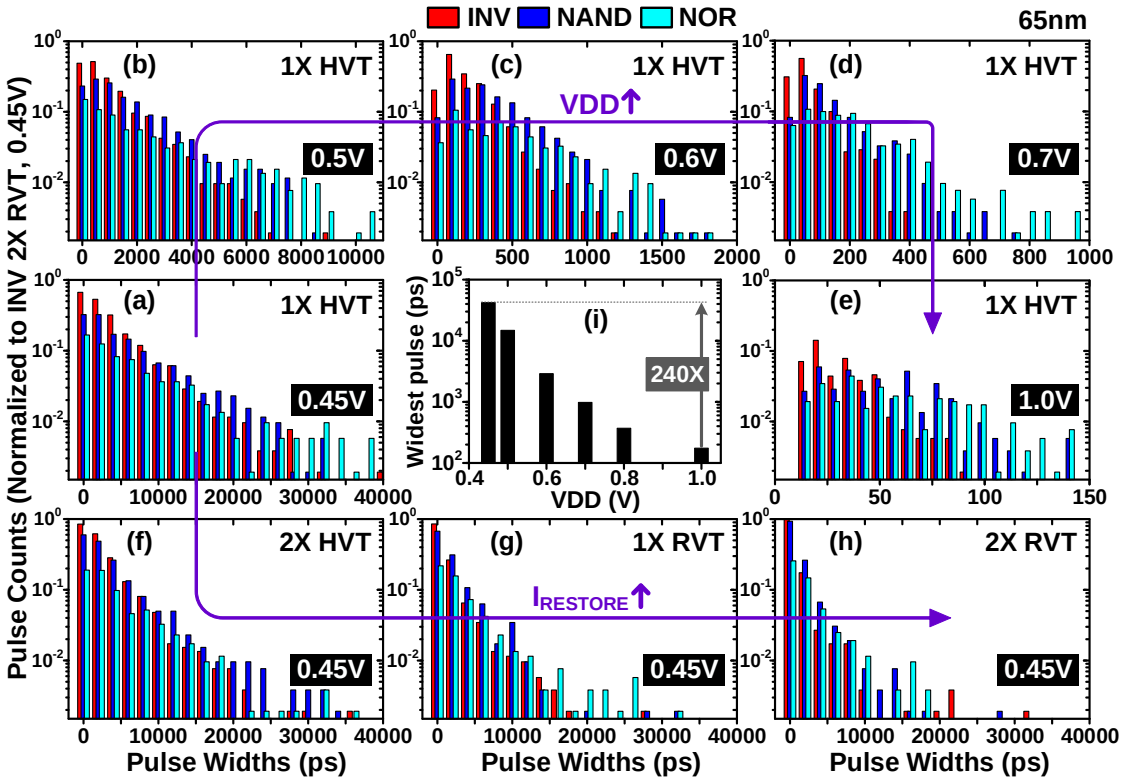


Figure 4.16: Sampled pulse width distributions corresponding to the standard logic gate types presented at multiple VDDs and for distinct transistor drive current variants.

Fig. 4.16 showcases the sampled pulse width distributions corresponding to the three standard logic gate types at multiple supply voltages (a)-(e) and for distinct V_{TH}: High V_{TH} - HVT, Regular V_{TH} - RVT and width variants: 1X, 2X (a)-(h). The increase in supply voltage increases the $I_{RESTORE}$ of the gate, making it more immune to soft-errors as well as

improving its response time, thus allowing rapid restoration of the affected node state and resulting in narrower pulse width distributions with fewer pulse counts sampled. Evinced from measured data, sampled pulse width distributions span a range of ~40ns at a Near Threshold Voltage (NTV) level of 0.45V (Fig. 4.16(a)) to <150ps at the nominal voltage (1.0V, Fig. 4.16(e)), exhibiting nearly a 240X increase in the widest sampled pulse, as seen in Fig. 4.16(i).

Among the different gate types, gates featuring stacks (i.e. NAND, NOR) exhibit broader distributions with distinctly higher pulse counts at wider pulse widths, due to (1) the lower $I_{RESTORE}$ for a stacked Pull-Up / Pull-Down network (PUN/PDN) which results in poorer immunity and slower response time to an SET event and (2) electrical masking effects, causing narrower SET pulses to shrink within the logic chains, the severity of which strongly depends on the supply voltage, process variations, drive current and the gate-type, with stacked gates particularly susceptible (as explained previously in Section 4.5, [8]). The above discussion is especially true for the NOR gates in the 65nm process on account of lower PMOS mobility, leading to further lowered $I_{RESTORE}$, which results in broader sampled distributions for this gate-type. Although, the distinction between stacked and unstacked gate-types is not clearly apparent at NTV levels (i.e. 0.45V, 0.5V) owing to similarly weaker $I_{RESTORE}$ in all gate-types, it becomes apparent as the supply voltage is increased, with higher counts for wider pulse widths primarily sampled from stacked gate types.

Fig. 4.16(a)-(h) present the impact of increasing the transistor drive strength or $I_{RESTORE}$. Arranged in the order of increasing drive current strengths at NTV (Fig. 4.5), an

increase in the gate drive results in the distributions to shift leftward to lower pulse widths, with more narrower pulses being sampled for the stronger variant of the same gate-type, also attributed partially to the reduction in electrical masking effects.

4.6.2 V_{TH} Impact

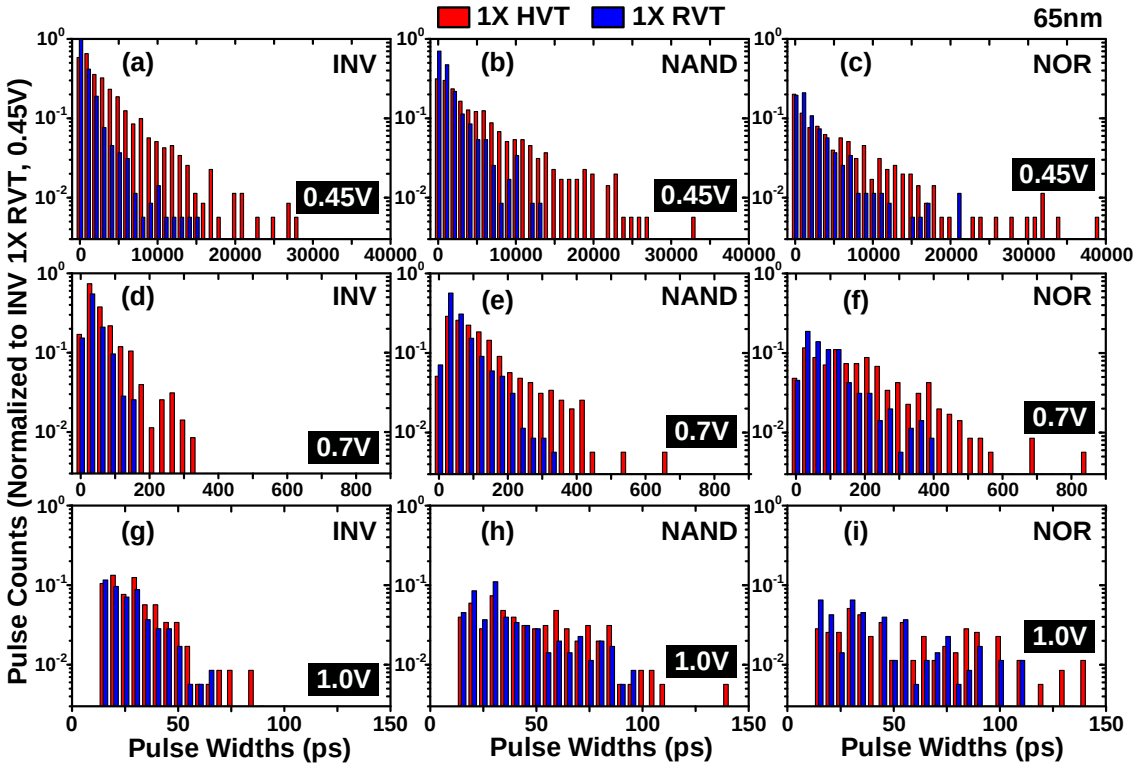


Figure 4.17: Sampled pulse width distributions comparing the V_{TH} variants for the 1X device widths at multiple VDDs.

Although prior art [43], [44] has studied the impact of V_{TH} variants on pulse width, further investigation is required to understand its effect in different gate types. Fig. 4.17 presents the sampled pulse width distributions for the V_{TH} variants corresponding to 1X device width flavor of the three logic gate-types at multiple VDDs. Evinced from measured distributions at 0.45V, a change in the V_{TH} variant from HVT to RVT impacts both the SER sensitivity and the sampled pulse width at NTV levels, owing to the increased $I_{RESTORE}$. The increase in VDD, however, negates this effect due to relative difference in

the overdrive voltage $V_{GS} - V_{TH}$ between the two variants leveling with the increase in VDD, observed from measured results at 0.7V, where the pulse width distributions for the two variants begin to approach each other. Finally, at 1.0V, nearly identical pulse width distributions with similar pulse counts are observed for all gate types. Similar trends were observed for 2X device widths.

4.6.3 Width Impact

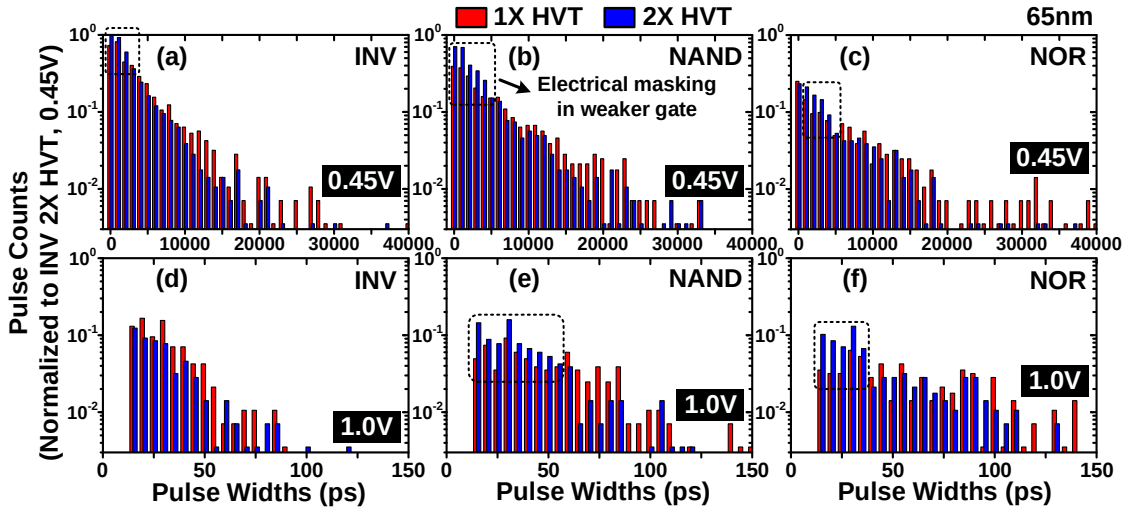


Figure 4.18: Pulse width distributions corresponding to the device width variants of the HVT gate flavors.

Fig. 4.18 presents the measured pulse width distributions corresponding to the width variants of the HVT flavor for standard gate-types. The increase in device width by a factor of two, does not impact the pulse width distributions noticeably at both nominal VDD as well as at NTV levels. In section 4.5, we discussed the impact of device width on the measured SER cross-section results, where it was observed that electrical masking effects progressively dominate as chain length increases from 16 to 64, resulting in a collective

higher SER for the 2X versions, especially for the weaker HVT flavors, stacked gate-types and at lower VDDs. Highlighted in Fig. 4.18, the increase in SER discussed previously for the 2X flavors, comes about from the loss of several narrower transients in the weaker 1X gates, which attenuate significantly before reaching the detection circuitry, particularly in longer chains. This effect will be isolated further in section (d). The impact is lower for the stronger RVT gates (not shown).

4.6.4 Logic Chain Length Impact

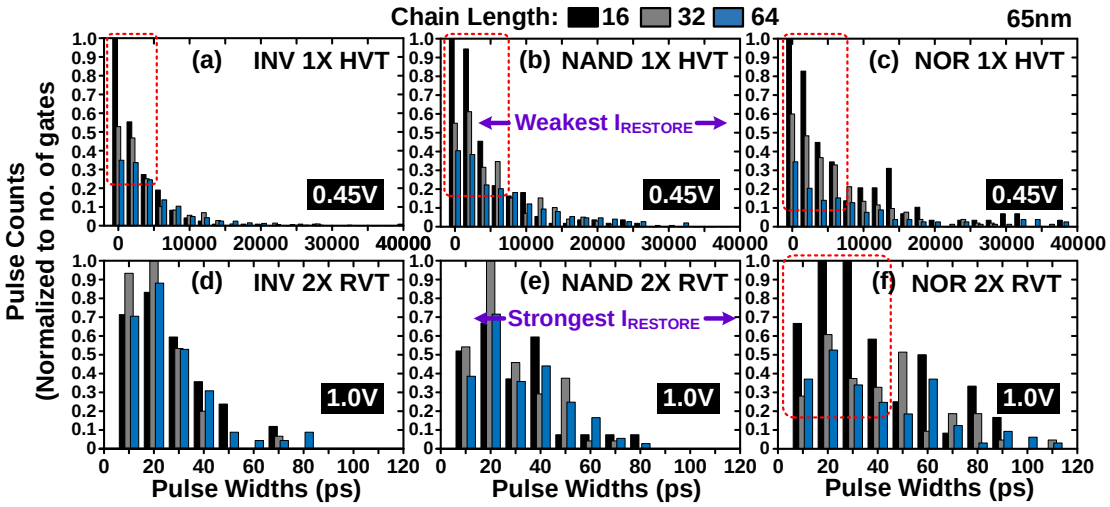


Figure 4.19: Sampled pulse distributions highlighting the impact of logic chain length

Fig. 4.19 showcases the impact of logic chain length with the sampled pulse width distributions normalized to the gate count per chain in each of the subfigures. To illustrate the severity of electrical masking and/or pulse quenching effects in longer logic chains, we present results from two corner cases: (1) weakest $I_{RESTORE}$, 1X HVT gates at NTV and (2) strongest $I_{RESTORE}$, 2X RVT gates operating at nominal VDD. At 0.45V, the joint impact of weaker $I_{RESTORE}$ and process variations (causing delay uncertainty and mismatch

between stages) has a pronounced effect on the gate response time, the probability of which compounds with the logic distance travelled, causing several narrower SETs to disappear in-chain. The impact is seen in terms of distinctly higher pulse counts sampled for narrower pulses corresponding to a shorter chain length of the same gate variant. At 1.0V, the increased $I_{RESTORE}$ results in an improved gate response time, largely offsetting process variations and enabling narrower pulses to reach the detection circuitry, evinced from similar sampled pulse counts for the 2X RVT variants. It should be noted, however, that NOR gates are still more susceptible even at voltages as high as 1.0V, due to the weak PMOS stack.

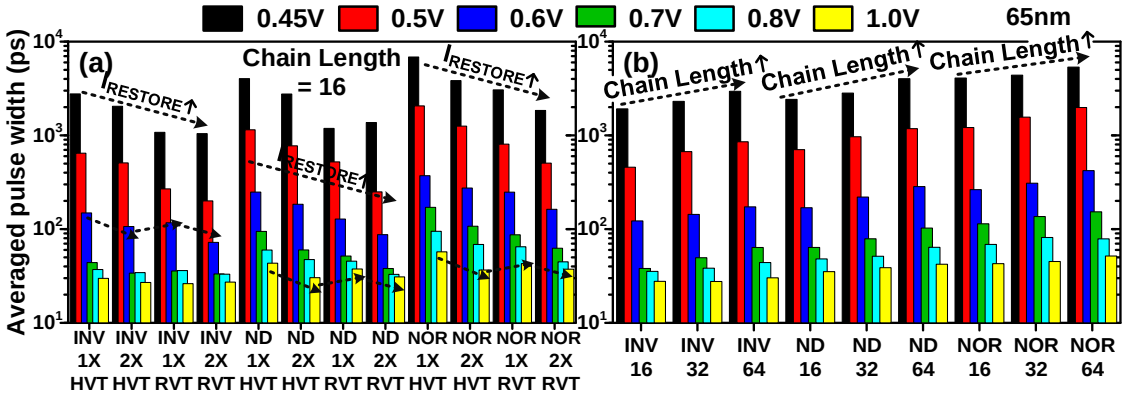


Figure 4.20: Averaged pulse widths: (a) indicating the $I_{RESTORE}$ dependence which agrees well with the simulations presented in section 4.4 and (b) impact of increasing chain length on the averaged pulse width.

Fig. 4.20(a) presents the averaged pulse widths for the distinct gate variants corresponding to a chain length of 16 at multiple VDDs. The results agree well with the $I_{RESTORE}$ simulations we presented in section 4.4, especially for INVs, due to a reduced susceptibility to masking effects as compared to stacked gates. Fig. 4.20(b) presents the averaged pulse width results for logic gates variants considered collectively, indicating the movement of the distribution towards a higher average value with increase in chain length, again a consequence of electrical masking effects.

4.7 Summary

This chapter presented neutron radiation induced Soft-Error-Rate (SER) and Single-Event-Transient SET pulse-width distribution statistics and detailed analysis thereof, revealing a multitude of circuit parameters impacting the soft-error susceptibility of standard combinational logic in advanced CMOS nodes. A high density array-based soft-error characterization vehicle is presented, featuring standard logic gate chains of varying lengths. Neutron irradiation data obtained from gate variants employing devices with distinct channel widths and threshold voltage flavors is analyzed at multiple supply voltages, ranging from nominal down to near-threshold. Supplemented with first-order simulations, measured SER cross-section results obtained from test structures implemented in a 65nm planar CMOS technology node reveal the complex interplay between factors such as supply voltage, node capacitance, restore current ($I_{RESTORE}$), gate topology and logic chain length responsible in contributing towards the collective soft error susceptibility of a standard gate type, which constitutes the main focus of this work. In addition, the easy process portability of the proposed macro is demonstrated through implementation in a 16nm FinFET process.

Chapter 5. Conclusion

The uniting theme and motivation behind each of the works presented in this thesis has been the conceptualization and demonstration of a novel circuit-based technique for efficient on-chip reliability characterization, given a problem. Furthermore, in each of the presented works, the approach has been to come up with an array-based test-vehicle, occupying a small footprint, which can be easily accommodated as an expandable scribe-line test-structure, given an available area, for use as a technology characterization vehicle. As discussed in the preceding chapters, the on-chip circuit-based approach has numerous advantages in comparison to the conventional probe-based characterization methodology, such as, improved test-efficiency on account of parallel characterization of a number of DUTs, better die-area utilization due to fewer shared pads, reduction in equipment complexity, fast I/O and easily automated setup on account of scan-based mode control, the most important being that the on-chip circuit-based characterization approach closely mimics the realistic scenario.

In chapter 2, an array-based test structure, with the specific aim of quantifying the impact of feedback on the aging dynamics of diode-connected MOSFETs was designed and implemented in a 65nm LP process. Diode-connected MOSFETs are ubiquitous in almost all Analog / Mixed signal circuits and systems such as current-mirrors, amplifiers and voltage or current controlled oscillators, to name a few. HCI, in conjunction with the feedback inherent in this circuit configuration results in a positive feedback leading to accelerated transistor aging overtime. Our aim was to obtain comprehensive test-data for the purpose of building an understanding of the actual aging dynamics of this circuit

configuration, using a dedicated test-vehicle which would allow for parallelly stressing the DUTs as well as individual characterizing them using only a few shared I/Os. Using the proposed vehicle, we characterized the impact of feedback on the aging rate and compared it to the no-feedback case and, evaluated the efficacy of iterative simulations for lifetime projection in comparison to the method based on the universality of hot carrier degradation extended to the case featuring feedback. Accelerated aging rate in this configuration led to an faster onset of saturation, thereafter which the circuit was found to age at a much slower rate. For high reliability applications, this effect would need to be accounted in, for an accurate estimation of device lifetime.

In chapter 3, we proposed three circuit-based test vehicles for on-chip EM characterization, implemented in the state of the art 16nm FinFET node. The presented test-vehicles featured metal heaters for achieving temperatures in excess of 300°C, to facilitate accelerated stress testing entirely on-chip, along with standard EM upstream / downstream test-structures as the DUTs. The first test-chip was essentially a resistance sensing monitor with parallel stress capability, geared towards efficient collection of Median-Time-To-Failure statistics from a large sample population of wires, using digital I/Os, SMUs and a simple temperature chamber in lieu of a dedicated oven-based complicated setup with limited wire slots. Interconnects with distinct metal stacks and DUT geometries were characterized using this test-vehicle and the results were analyzed in detail. The second test-chip was a fully digital EM monitor, which introduced BER tracking of an interconnect path as a new metric for both diagnosing as well as quantifying the impact of EM induced resistance shifts on the signaling capability of the datapath. We

introduced a novel test-vehicle, wherein the data ‘0’ and ‘1’ BERs were accounted for separately, the related test-methodology, and the detailed characterization results from four individual datapaths stressed at a temperature of 360°C. Both abrupt as well as progressive resistance shift signatures were distinctly captured in the measured BER over the total stress time, validating the efficacy of the technique in both capturing the EM effects, as well as translating them to the signaling capability of the datapath. In the last section of this chapter, we introduced yet another EM characterization macro, featuring a combination of capabilities borrowed from the former designs, the primary motivation being the translation of a measured EM-induced resistance change to an accompanying shift in the BER of a datapath. The vehicle further featured flexibility in mode-control allowing for wires comprising a datapath to be stressed independently for detailed EM studies, as well as introducing multiple methods for polling the DUT resistance, to monitor additional effects such as Joule heating or the local DUT temperature, when the on-chip heaters are activated.

Finally, in chapter 4, we introduced an array-based, high density, soft-error test vehicle meant for standard combinational logic characterization, including both Soft-Error-Rate (SER) as well as Single-Event-Transient pulse-width distribution characterization capability. The proposed macro concept was demonstrated in 65nm planar CMOS and 16nm bulk FinFET nodes and the corresponding test-chips were irradiated under a spallation neutron beam at Los Alamos Neutron Science Center. Detailed analysis and comparison on the measured SER statistics as well as the SET pulse-width distributions was presented for each of the standard combinational logic gates, outlining the key

parameter dependencies influencing the soft-error susceptibility from a circuit design perspective. Since standard gates themselves were used as the DUTs, no indirect translations are further required for utilizing the obtained statistics.

Bibliography

[1] D. Jang, E. Bury, R. Ritzenthaler, M. G. Bardon, T. Chiarella, K. Miyaguchi, P. Raghavan, A. Mocuta, G. Groeseneken, A. Mercha, D. Verkest and A. Thean, "Self-heating on bulk FinFET from 14nm down to 7nm node," in *International Electron Devices Meeting*, Washington, DC, 2015, pp. 11.6.1-11.6.4.

[2] T. Takahashi, N. Beppu, K. Chen, S. Oda and K. Uchida, "Thermal-aware device design of nanoscale bulk/SOI FinFETs: Suppression of operation temperature and its variability," in *International Electron Devices Meeting*, Washington, DC, 2011, pp. 34.6.1-34.6.4.

[3] S. Lee, I. Kim, S. Ha, C. Yu, J. Noh, S. Pae and J. Park, "Radiation-induced soft error rate analyses for 14 nm FinFET SRAM devices," in *IEEE Internation Reliability Physics Symposium*, Monterey, CA, 2015, pp. 4B.1.1-4B.1.4.

[4] T. Uemura, S. Lee, J. Park, S. Pae and H. Lee, "Investigation of logic circuit soft error rate (SER) in 14nm FinFET technology," in *IEEE International Reliability Physics Symposium*, Pasadena, CA, 2016, 3B-4-1-3B-4-4.

[5] T. Uemura, S. Lee, D. Min, I. Moon, J. Lim, S. Lee, H. Sagong and S. Pae, "Investigation of alpha-induced single event transient (SET) in 10nm FinFET logic circuit," in *IEEE International Reliability Physics Symposium (IRPS)*, Burlingame, CA, 2018, pp. P-SE.1-1-P-SE.1-4..

[6] N. Pande, G. Park, C. H. Kim, S. Krishnan and V. Reddy, "Investigating the Aging Dynamics of Diode-Connected MOS Devices Using an Array-Based Characterization Vehicle in a 65nm Process," in *IEEE International Physics Symposium (IRPS)*, Monterey, CA, 2019, pp. 1-6.

[7] N. Pande, C. Zhou, M. Lin, R. Fung, R. Wong and C. H. Kim, "Characterizing Electromigration Effects in a 16nm FinFET Process Using a Circuit Based Test Vehicle," in *IEEE International Electron Device Meeting (IEDM)*, San Franciso, CA, 2019, pp. 5.3.1-5.3.4.

[8] N. Pande, S. Kumar, L. R. Everson and C. H. Kim, "Understanding the Key Parameter Dependences Influencing the Soft-Error Susceptibility of Standard Combinational Logic," *IEEE Transactions on Nuclear Science (TNS),* vol. 67, pp. pp. 116-125, 2020.

[9] K. B. Sutaria, A. Mohanty, R. Wang, R. Huang and Y. Cao, "Accelerated Aging in Analog and Digital Circuits With Feedback," *IEEE Transactions on Device and Material Reliability (TDMR),* vol. 15, no. 3, pp. 384-393, Sept. 2015.

[10] K. B. Sutaria, P. Ren, A. RamKumar, R. Zhu, X. Feng, R. Wang, R. Huang and Y. Cao, "Diagnosing bias runaway in analog/mixed signal circuits," in *IEEE International Reliability Symposium (IRPS)*, Waikoloa, HI, 2014.

[11] D. Patra, J. Zhang, R. Wang, M. Katoozi, E. H. Cannon, R. Huang and Y. Cao, "Compact modeling and simulation of accelerated circuit aging," in *IEEE Custom Integrated Circuits Conference (CICC)*, San Diego, CA, 2018, pp 1-4.

[12] V. Huard, F. Cacho, X. Federspiel, W. Arfaosui, M. Saliva and D. Angot, "Technology scaling and reliability: Challenges and opportunities," in *IEEE International Electron Devices Meeting (IEDM)*, Washington, DC, 2015, pp. 20.5.1-20.5.6.

[13] K. Hess, A. Haggag, W. McMohan, B. Fischer, K. Cheng, J. Lee and J. Lyding, "Simulation of Si-SiO2 defect generation in CMOS chips:From atomistic structure to chip failure rates," in *International ELectron Device Meeting (IEDM)*, San Francisco, CA, 2000, pp 93-96.

[14] D. Varghese, M. A. Alam and B. Weir, "A generalized, IB-independent, physical HCI lifetime projection methodology based on universality of hot-carrier degradation," in *IEEE International Reliability Physics (IRPS)*, Anaheim, CA, 2010, pp 1091-1094.

[15] S. Mahapatra and R. Saikia, "On the Universality of Hot Carrier Degradation: Multiple Probes, Various Operating Regimes, and Different MOSFET Architectures," *IEEE Transactions on Electron Devices*, vol. 65, no. 8, pp. 3088-3094, Aug. 2018.

[16] E. Takada and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," *IEEE Electron Device Letters,* vol. 4, no. 4, pp. 111-113, April 1983.

[17] D. Varghese, H. Kufluoglu, V. Reddy, H. Shichijo, S. Krishnan and M. A. Alam, "Universality of Off-State Degradation in Drain Extended NMOS Transistors," in *International Electron Device Meeting (IEDM),* San Franciso, CA, 2006, pp. 1-4.

[18] C. Zhou, X. Wang, R. Fung, S. Wen, R. Wong and C. H. Kim, "A Circuit-Based Approach for Characterizing High Frequency Electromigration Effects," *IEEE Transactions on Device and Materials Reliability (TDMR),* vol. 17, no. 4, pp. 763-772, Dec. 2017.

[19] S. Lee and A. S. Oates, "Identification and Analysis of Dominant Electromigration Failure Modes in Copper/Low-K Dual Damascene Interconnects," in *IEEE International Reliability Physics Symposium Proceedings (IRPS),* San Jose, CA, 2006, pp 107-114.

[20] A. S. Oates and M. H. Lin, "Electromigration Failure Distributions of Cu/Low-k Dual-Damascene Vias: Impact of the Critical Current Density and a New Reliability Extrapolation Methodology," *IEEE Transactions on Device and Materials Reliability (TDMR),* vol. 9, no. 2, pp. 244-254, 2009.

[21] S. Buchner, M. Baze, D. Brown, D. McMorrow and J. Melinger, "Comparison of error rates in combinational and sequential logic," *IEEE Transactions on Nuclear Science (TNS)*, vol. 44, no. 6, pp. 2209-2216, Dec. 1997.

[22] N. N. Mahatme, S. Jagannathan, T. D. Loveless, L. W. Massengil, B. L. Bhuva, S. J. Wen and R. Wong, "Comparison of Combinational and Sequential Error Rates for a Deep Submicron Process," *IEEE Transactions on Nuclear Science (TNS)*, vol. 58, no. 6, pp. 2719-2725, Dec. 2011.

[23] J. Furuta, C. Hamanaka, K. Kobayashi and H. Onodera, "Measurement of neutron-induced SET pulse width using propagation-induced pulse shrinking," in *International Reliability Physics Symposium (IRPS)*, Monterey, CA, 2011, pp. 5B2.1-5B2.5.

[24] S. Kumar, M. Cho, L. Everson, H. Kim, Q. Tang, P. Mazanec, P. Meinerzhagen, A. Malavasi, D. Lake, C. Tokunaga, H. Quinn, M. Khellah, J. Tschanz, S. Borkar, V. De and C. H. Kim, "Statistical characterization of radiation-induced pulse waveforms and flip-flop soft errors in 14nm tri-gate CMOS using a back-sampling chain (BSC) technique," in *Symposium on VLSI Technology*, Kyoto, 2017, pp C114-115.

[25] T. D. Loveless, J. S. Kaupilla, S. Jagannathan, D. R. Ball, J. D. Rowe, N. J. Gaspard, N. M. Atkinson, R. W. Blaine, T. R. Reece, J. R. Ahlbin, T. D. Haeffner, M. L. Alles, W. T. Holman, B. L. Bhuva and L. W. Massengill, "On-Chip Measurement of Single-

Event Transients in a 45 nm Silicon-on-Insulator Technology," *IEEE Transaction on Nuclear Science (TNS),* vol. 59, no. 6, pp. 2748-2755, Dec. 2012.

[26] J. R. Ahlbin, L. W. Massengill, B. L. Bhuva, B. Narasimham, M. J. Gadlage and P. H. Eaton, "Single-Event Transient Pulse Quenching in Advanced CMOS Logic Circuits," *IEEE Transactions on Nuclear Science (TNS),* vol. 56, no. 6, pp. 3050-3056, Dec. 2009.

[27] M. Mitrovic, M. Hofbouer, K. Voss and H. Zimmerman, "Experimental Investigation of the Joint Influence of Reduced Supply Voltage and Charge Sharing on Single-Event Transient Waveforms in 65-nm Triple-Well CMOS," *IEEE Transactions on Nuclear Science (TNS),* vol. 65, no. 8, pp. 1908-1913, Aug. 2018.

[28] M. Mitrovic, M. Hofbauer, B. Goll , K. Schneider-Hornstien, R. Swoboda, B. Steindl, K. Voss and H. Zimmermann, "Experimental Investigation of Single-Event Transient Waveforms Depending on Transistor Spacing and Charge Sharing in 65-nm CMOS," *IEEE Transactions on Nuclear Science (TNS),* vol. 64, no. 8, pp. 2136-2143, Aug. 2017.

[29] M. Mitrovic, M. Hofbauer, K. Schnieder-Hornstien, B. Goll, K. Voss and H. Zimmermann, "Evidence of Pulse Quenching in AND and OR Gates by Experimental Probing of Full Single-Event Transient Waveforms," *IEEE Transactions on Nuclear Science (TNS),* vol. 65, no. 1, pp. 382-290, Jan 2018.

[30] S. Kumar, M. Cho, L. Everson, H. Kim, Q. Tang, P. Mazanec, P. Meinerzhagen, A. Malavasi, D. Lake, C. Tokunaga, H. Quinn, M. Khellah, J. Tschanz, S. Borkar, V. De and C. H. Kim, "An ultra-dense irradiation test structure with a NAND/NOR readout chain for characterizing soft error rates of 14nm combinational logic circuits," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, 2017, pp 39.3.1-39.3.4.

[31] J. Qin, S. Chen, B. Liang, Z. Ge, Y. He, Y. Du, B. Liu, J. Chen and D. Li, "Voltage Dependency of Propagating Single-Event Transient Pulsewidths in 90-nm CMOS Technology," *IEEE Transactions on Device and Materials Reliability (TDMR)*, vol. 14, no. 1, pp. 139-145, March 2014.

[32] J. R. Ahlbin, M. J. Gadlage, D. R. Ball, A. W. Witulski, B. L. Bhuva, R. A. Reed, G. Vizkelethy and L. W. Massengill, "The Effect of Layout Topology on Single-Event Transient Pulse Quenching in a 65 nm Bulk CMOS Process," *IEEE Transaction on Nuclear Science (TNS)*, vol. 57, no. 6, pp. 3380-3385, Dec. 2010.

[33] M. Glorieux, A. Evans, V. Ferlet-Cavrois, C. Boatella-Polo, D. Alexandrescu, S. Clerc, G. Gasiot and P. Roche, "Detailed SET Measurement and Characterization of a 65 nm Bulk Technology," *IEEE Transactions on Nuclear Science (TNS)*, vol. 64, no. 1, pp. 81-88, Jan. 2017.

[34] R. Powlowski, J. Crop, M. Cho, J. Tschanz, V. De, T. Fairbanks, H. Quinn, S. Borkar and P. Y. Chiang, "Characterization of radiation-induced SRAM and logic soft errors

from 0.33V to 1.0V in 65nm CMOS," in *IEEE Customs Integrated Circuit Conference (CICC)*, San Jose, CA, 2014.

[35] F. Wrobel, L. Dilillo, A. D. Touboul, V. Pouget and F. Saigne, "Determining Realistic Parameters for the Double Exponential Law that Models Transient Current Pulses," *IEEE Transactions on Nuclear Science (TNS)*, vol. 61, no. 4, pp. 1813-1818, Aug. 2014.

[36] G. C. Messenger, "Collection of Charge on Junction Nodes from Ion Tracks," *IEEE Transaction on Nuclear Science (TNS)*, vol. 29, no. 6, pp. 2024-2031, Dec. 1982.

[37] B. Gill, M. Nicolaidis, F. Wolff, C. Papachristou and S. Garverick, "An efficient BICS design for SEUs detection and correction in semiconductor memories," in *Design, Automation and Test in Europe*, Munich, Germany, 2005, pp. 592-597.

[38] T. Heijman, D. Giot and P. Roche, "Factors that impact the critical charge of memory elements," in *IEEE International On-Line Testing Symposium (IOLTS'06)*, Lake Como, 2006, pp 57-62.

[39] E. Farjallah, V. Gherman, J. Armani and L. Dilillo, "Evaluation of the temperature influence on SEU vulnerability of DICE and 6T-SRAM cells," in *International Conference on Design & Technology of Integrated Systems In Nanoscale Era (DTIS)*, Taormina, 2018.

[40] J. A. Maharrey, J. S. Kaupilla, R. C. Quinn, T. D. Loveless , E. X. Zhang, W. T. Holman, B. L. Bhuva and L. W. Massengill, "Heavy-Ion Induced SETs in 32nm SOI Inverter Chains," in *IEEE Radiation Effects Data Workshop (REDW)*, Boston, MA, 2015, pp 10-14..

[41] J. A. Maharrey, J. S. Kaupilla, R. C. Harrington, P. Nsengiyumva, D. R. Ball, T. D. Haeffner, E. X. Zhang, B. L. Bhuva, W. T. Holma and L. W. Massengill, "Impact of Single-Event Transient Duration and Electrical Delay at Reduced Supply Voltages on SET Mitigation Techniques," *IEEE Transactions on Nuclear Science,* vol. 65, no. 1, pp. 362-368, Jan 2018.

[42] J. S. Kaupilla , J. A. Maharrey, R. C. Harrington, T. D. Haeffner, P. Nsengiyumva, D. R. Ball, A. L. Sternberg, E. X. Zhang, B. L. Bhuva and L. W. Massengill, "Exploiting Parallelism and Heterogeneity in a Radiation Effects Test Vehicle for Efficient Single-Event Characterization of Nanoscale Circuits," *IEEE Transactions on Nuclear Science,* vol. 65, no. 1, pp. 486-494, Jan 2018.

[43] R. C. Harrington, J. A. Maharrey, J. S. Kaupilla, P. Nsengiyumva, D. R. Ball, T. D. Haeffner, E. X. Zhang, B. L. Bhuva and L. W. Massengill, "Effect of Transistor Variants on Single-Event Transients at the 14-/16-nm Bulk FinFET Technology Generation," *IEEE Transactions on Nuclear Science (TNS),* vol. 65, no. 8, pp. 1807-1813, Aug. 2018.

[44] J. A. Maharrey, R. C. Quinn, T. D. Loveless, J. S. Kaupilla, S. Jagannathan , N. M. Atkinson, N. J. Gaspard, E. X. Zhang, M. L. Alles, B. L. Bhuva, W. T. Holman and L. W. Massengill, "Effect of Device Variants in 32 nm and 45 nm SOI on SET Pulse Distributions," *Transactions on Nuclear Science (TNS),* vol. 60, no. 6, pp. 4399-4404, Dec. 2013.

[45] H. Nakamura, T. Uemura, K. Takeuchi, T. Fukuda, S. Kumashiro and T. Mogami, "Scaling effect and circuit type dependence of neutron induced single event transient," in *IEEE International Reliability Physics Symposium (IRPS)*, Anaheim, CA, 2012, pp 3C.3.1-3C.3.7.