Test Chip Design for Process Variation Characterization in 3D Integrated Circuits

by

Conor T. O'Sullivan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A test chip design is presented for the characterization of process variations and Through Silicon Via (TSV) induced mechanical stress in 3D integrated circuits. The chip was designed, layed-out, and taped-out for fabrication in a 130nm Tezzaron/GlobalFoundries process through CMC microsystems. The test chip takes advantage of the architecture of 3D ICs to split its test structure onto the two tiers of the 3D IC, achieving a device array density of $40.94\mu m^2$ per device. The design also has a high spatial resolution and measurement fidelity compared to similar 2D variation characterization test structures.

Background leakage subtraction and radial filtering are two techniques that are applied to the chip's measurements to reduce its error further for subthreshold device current measurements and stress-induced mobility measurements, respectively. Experimental measurements are be taken from the chip using a custom PCB measurement setup once the chip has returned from fabrication.

Acknowledgements

I would like to thank my supervisors, Siddharth Garg and Peter M. Levine, for their support and assistance towards my research. I would like to thank CMC for the opportunity to create an IC.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Three dimensional integrated circuit (3D IC) technologies create 3D CMOS ICs from multiple 2D CMOS dies by stacking and connecting them together vertically. As CMOS process nodes have decreased in size, an increasing concern has been the trend of increasing delay along the wires of an IC [1]. 3D ICs offer a potential solution to this problem. By stacking the design of an IC onto separate tiers in a vertical stack, the average net length of the IC decreases, and delay decreases along with it [15]. However, 3D ICs have several reliability issues, including thermal dissipation, clocktree imbalances, and substrate noise injection [23, 12, 24].

Two 3D IC issues are of particular interest to this work, as little measured data has been published to characterize them. The first is the effects of process variation on a 3D IC, specifically the characterization the performance mismatch caused by inter-die variations between the tiers of a 3D IC. Another issue of secondary concern is the stress introduced by the vertical interconnects of the 3D IC, and the effects that this mechanical stress has on the performance of nearby devices. In this work, the design of a test chip used for the characterization of these issues is proposed and detailed. Additionally, a comparison is made to see the performance impact on implementing a characterization test structure in 3D, with respect to an established 2D design.

1.1 Outline

This thesis is structured as follows: chapter 2 provides background on 3D IC technology and describes the structure and features of the 3D IC technology used to design and fabricate the test chip. It also describes the effects of process variation and stress within 3D ICs. Chapter 3 provides a description of the design of the test chip and compares its design to prior, similar test structures. Chapter 4 describes techniques used to reduce the error of measurements taken from the test chip. Chapter 5 provides simulation results used to predict the accuracy of the test chip's measurements and its performance relative to that of a previous test structure design. It determines the measurement error of the test chip for Monte Carlo simulations and stress simulations, and determines the effectiveness of the error reduction techniques proposed in chapter 4. Chapter 6 briefly describes the measurement setup that will be used to operate the test chip upon its return from fabrication. It also describes extraction techniques that will be used to extract device parameters from the chip for process variation characterization. Chapter 7 concludes the thesis and discusses future work that can be made using the test chip, as well as other potential 3D IC designs that can be based on the chip's current design.

Chapter 2

Background

Three dimensional integrated circuit (3D IC) technology is an alternate manufacturing process to conventional planar 2D ICs. While a 2D IC consists of only a single layer of active devices, a 3D IC consists of multiple layers of vertically-stacked active devices, with each layer communicating to the others through vertical connections. For example, in a traditional 2D CMOS process die, the lowest layers would define the transistors of the IC, and the layers above these house the metal and via connections that define the wiring of the circuit. A 3D CMOS IC would feature several 2D CMOS dies (or tiers) stacked on top of each other with additional vertical connections made to ensure communication between each die.

2.1 Motivation for Using 3D ICs

3D ICs offer certain advantages over 2D ICs. For example, they allow for easier heterogeneous integration. Creating a system-on-a-chip (SoC) requires the integration of several disparate circuits: usually analog, digital, and memory (DRAM, SRAM, flash, etc.). Placing these all within a single chip requires a very complicated manufacturing process. 3D IC technology can simplify this, as different parts of a system can be manufactured on

separate dies with separate, simpler processes, and then combined into a single chip as one vertical stack [15].

3D ICs also provide a significant reduction of interconnect delay compared to 2D ICs. All routed metal wires in an IC have significant parasitic resistances and capacitances associated with them. A typical IC interconnect can be approximated using the lumped RC model, depicted in Fig. 2.1.

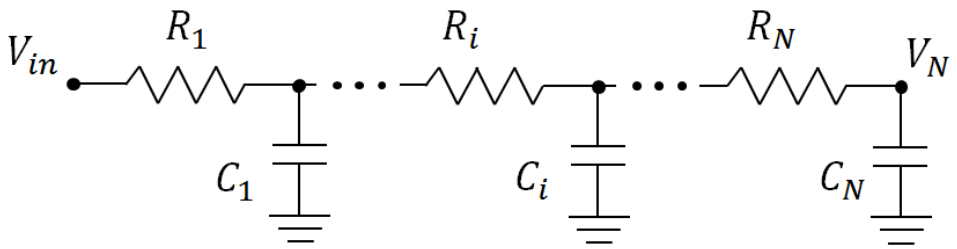

Figure 2.1: Lumped RC model of an interconnect.

This model depicts an interconnect as a series of discrete segments, with each segment R_i and C_i representing the sum of resistances and capacitances along each segment of the wire. The propagation delay along this wire is dependent on its time constant, τ, given as

$$\tau = RC$$

Where R and C are the sum of the resistances and capacitances of the wire, which are proportional to the length (L_{wire}) and width (W_{wire}) of the wire.

$$R \propto \frac{L_{wire}}{W_{wire}}$$
$$C \propto W_{wire} L_{wire}$$

It follows that τ is proportional to the square of the length of the wire.

$$\tau \propto L_{wire}^2$$

It is therefore evident that the approximate delay of an interconnect is proportional to the square of the wire's length and that the most effective way to lower the propagation delay of an interconnect is to keep it as short as possible [18]. This is perhaps the most significant advantage 3D IC technology offers, as a 3D implementation of a circuit can have significantly shorter interconnect lengths versus a 2D implementation of the same circuit. This is especially true for SoCs. In 2D, long global nets are required to connect the different parts of the system together. In 3D, the different parts of a system are allocated to different layers of the vertical stack, and components of the SoC that would be horizontally distant from each other are now connected vertically at a comparatively smaller distance. Generally, the approximate relationship between the total net length of a 3D IC and the total net length of an equivalent 2D IC can be approximated as

$$l_{tot,3D} \approx \frac{l_{tot,2D}}{\sqrt{T}} [9]$$

where T is the number of tiers that compose the vertical stack of the 3D IC [9]. For example, a two-tiered 3D IC would have a total net length smaller than its 2D counterpart's by a factor of 1.41, a three-tiered 3D IC would have it reduced by a factor of 1.73, etc. As CMOS process nodes continue to shrink to smaller sizes, interconnect delay has increased and is expected to increase for the foreseeable future as the sizes of CMOS gates shrink [1]. 3D technology is one potential solution to the problem of interconnect delay in modern ICs.

2.2 Related 3D IC Technologies

There are several different implementations of 3D technology, each defined by their approach to creating the vertical connections between the 2D tiers in their vertical stack. Some non-conventional approaches include the method of wire-bonding the tiers of the 3D IC together at the periphery of each tier [3], and using contactless connections by transmitting signals between the tiers using capacitive or inductive AC coupling circuits [14, 7].

A more common implementation of 3D interconnects is through the use of microbumps. Two dies can be thinned down mechanically to expose metal pads on their surfaces, and solder bumps are then used to connect these pads together. The result is a 3D design with lower interconnect length than an equivalent 2D design, but there are issues with such an approach. The connections between metal pads on each tier can only be made on the top layer of their respective dies, and so the microbump bonding must fuse the dies together using face-to-face bonding, a 3D IC stacking technique where the top sides of the dies are fused together. Face-to-face microbump bonding can only be performed once, and limits the maximum layers of the 3D IC to two. The solder microbumps can also be quite large (50μm to 500μm) which limits the interconnect density between the two tiers [3].

Another common approach to connecting 3D tiers is through the use of Through Silicon Vias (TSVs). TSVs are vertical metal connections made between the metal layers of two tiers of a 3D IC, allowing for direct communication between active devices from both tiers. TSVs are created during the assembly of a 3D IC, when one or more tiers of the IC are stacked face-to-face (or face-to-back, depending on the specific 3D technology process utilizing TSVs). Holes are etched through one tier to the other, lined with an insulator, then filled with a metal (usually copper or tungsten) to create the TSV connection between the tiers [3]. This process can be repeated for multiple tiers, which allows TSV processes to create 3D ICs with larger vertical stacks than microbump processes can. However, a 3D IC with many tiers also has its drawbacks. Every tier adds extra complexity to the IC assembly process, and heat dissipation in a multi-tier 3D IC can also be an issue, and is one of the primary limiting factors for the stack size of a 3D IC [23]. Yield is another issue, as each tier in the stack cannot be tested for faults, as the IC is not functional until the stack is completely assembled. This can negatively impact the yield of multi-tier 3D ICs, since a fault in only a single tier of the IC will usually cause a failure in the entire IC, and a large stack increases the probability of such an occurrence [23]. Despite these disadvantages, TSV processes are capable of creating TSVs with pitches of less than 50μm, giving them better interconnect density than most microbump processes, allowing them to take better advantage of 3D IC interconnect delay reduction.

2.3 Test Chip 3D IC Technology

A TSV-based 3D technology was used to create the test chip described in this thesis. It is a Tezzaron/GlobalFoundries technology using a 130nm CMOS process node. This technology is capable of creating a two-tiered 3D IC, but uses a unique method for connecting its two tiers. Fig. 2.2 shows a cross section example of the Two-tiered 3D IC. The tiers are bonded together face-to-face. Although this is a TSV-based 3D process, TSVs are not used to used to make connections between the two tiers. Instead, the highest metal layer (metal 6) of the dies of both tiers are reserved for the placement of copper bondpoints. The bondpoints of the two tiers are joined together using thermal compression to create the IC's interconnections. Since the bondpoints have a minimum spacing of 4μm, the Tezzaron process is capable of very high interconnect density.

TSVs are still required for the functionality of the two-tiered Tezzaron 3D IC. Since the two tiers have been bonded face to face, the circuitry of the chip is locked in the core of the IC. Tungsten TSVs are etched into the tiers before face-to-face bonding, and one tier of the IC is thinned down to expose the TSVs to I/O pads, allowing access to signals to and from the IC. Since the TSVs have a very short height of 6μm, the thinned tier is reduced to a total height of 750μm to approximately 15μm. This same thinning procedure cannot be performed on the other tier without damaging the IC. This limits the I/O of the chip, as an equivalent 2D chip would have about twice as much space for I/O pads for the same amount of core circuitry.

2.4 3D IC Reliability Issues

There are several major design and reliability issues associated with 3D ICs. Some are caused by the vertical stack that is characteristic of all ICs. These include thermal density issues caused by high power dissipation in intermediate tiers of a 3D stack [23], and issues balancing the clock tree of a circuit across multiple tiers of a chip [12]. Certain features unique to 3D ICs can cause problems as well. For example, TSVs carrying high-speed

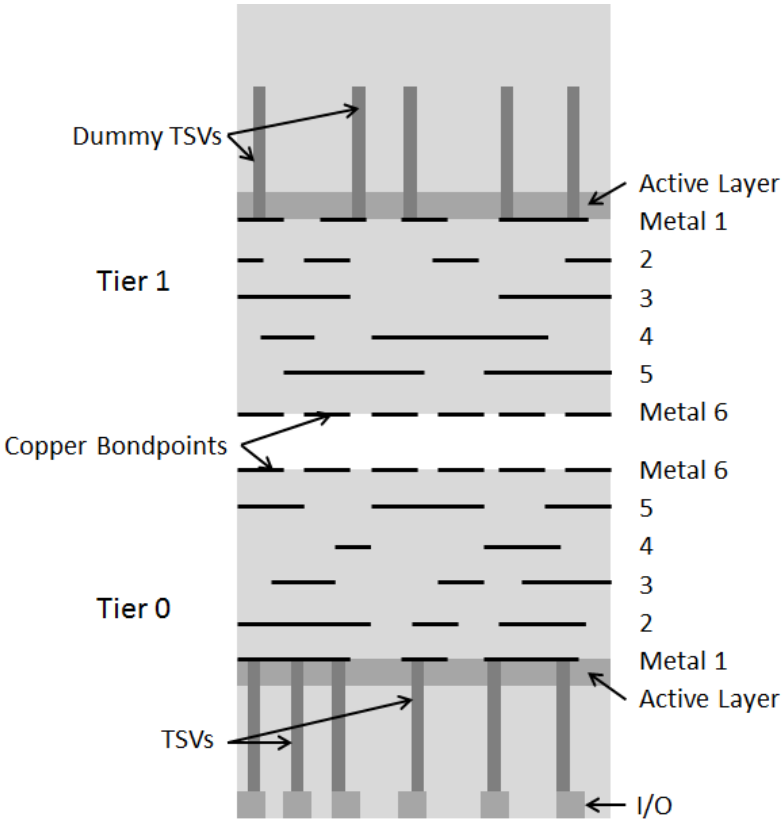

Figure 2.2: Cross section of a two-tier Tezzaron/GlobalFoundries 3D IC.

signals through the substrates of a 3D IC have the potential to introduce significant noise into the surrounding substrate, affecting the performance of nearby analog devices [11, 24].

The 3D reliability issue that the test chip has been primarily designed to characterize is process variation, as there has been very little published work examining experimentally measured data for process variations. A secondary issue that the test chip is also designed to characterize is TSV-induced mechanical stress.

2.4.1 Process Variation

Process variation is an issue that significantly impacts the performance of short-channel devices. Process variations have always been present in CMOS devices, but their effects have only become significant as CMOS processes scaled and device parameters have shrunk. In modern CMOS processes, a single set of fixed device parameters is no longer enough to accurately predict device performance and behaviour. In reality, when a wafer of CMOS dies is fabricated, the performances of any two transistors, whether they are located on two different dies or on the same die, are slightly different from one another. These performance mismatches are primarily due to variations in the device parameters caused by non-uniform conditions during the manufacturing of the wafer. Small errors in the photolithographic step of the manufacturing process might cause variations in the channel width (W) channel length (L), and source and drain regions of two devices, while non-uniform conditions from the deposition of oxide or impurities into the wafer can cause variations in device oxide thickness (t_{ox}) or the doping concentrations of the devices [18]. Some of the most important sources of process variations include variations in t_{ox} and device threshold voltage (V_{th}), as well as random dopant fluctuations, non-uniformity in implanting processes, and line-edge roughness affecting the dimensions of the device [4].

Models can be made to predict or characterize the behaviour of process variations. Accurately representing process variations with a model is a very involved process however, and there are many different mathematical models used to make this characterization, most of which involve statistical analysis. Generally speaking, each varying parameter (V_{th}, t_{ox}, L, etc.) can be represented as a random variable X. Each random variable can be represented with a probability distribution function with a certain mean (μ) and standard deviation (σ) [4]. Representing each device parameter as a random variable with a normal distribution is the simplest way to characterize their behaviour under process variations, and this is how they are represented when process variations are simulated using Monte Carlo simulations. However, a normal distribution is not always the most accurate way to simulate process variations. Device parameters can often fluctuate in ways that a normal distribution cannot predict. In this case, other methods must be used

to accurately characterize the behaviour of these device parameters [4].

Another complication to the modeling of process variations is that they can be classified as either intra-die or inter-die variations. Inter-die variations refer to variations in the parameters of each die on a wafer. Intra-die variations refer to the variations that occur within a specific die on the wafer, and are variable within that die. Inter-die variations are constant inside a die, but vary from die to die across the wafer [19]. Intra-die variations tend to be location invariant random parameter fluctuations, while inter-die variations tend to be represented as shifts in the mean values (μ) of the parameters [4]. A device parameter, of a specific die i, represented as a random variable X_i, can be described in terms of intra and inter-die variations as

$$X_i = X_{nom} + \Delta X_{inter} + \Delta X_{intra,i}[4]$$

where the device parameter is now determined by its nominal value (X_{nom}), plus a random inter-die variation ΔX_{inter} which is a fixed constant for all devices of that die, and an additional random intra-die variance ($\Delta X_{intra,i}$) for that die. Both ΔX_{inter} and $\Delta X_{intra,i}$ can be represented as some kind of probability distribution function [19, 4].

While the above representation can accurately represent the different process variations present in a 2D IC, which consists of only 1 die, it is inadequate for the characterization of process variations of 3D ICs, which can consist of many interconnected dies. Take the 130nm Tezzaron/GlobalFoundries 3D process described earlier in Fig. 2.2. Its ICs consist of two dies, distinguished as tier 0 and tier 1. A random variable X_{tier0}, representing the device parameter of all devices across tier 0 of a two-tiered Tezzaron 3D IC, can be modeled as

$$X_{tier0} = X_{nom} + \Delta X_{inter,tier0} + \Delta X_{intra,tier0}$$

While a random variable X_{tier1}, representing the device parameter of all devices across tier 0 of a two-tiered Tezzaron 3D IC, can be modeled as

$$X_{tier1} = X_{nom} + \Delta X_{inter,tier1} + \Delta X_{intra,tier1}$$

The implication of these equations is that the device parameters will not match between the two tiers, and that the total variance of these parameters is potentially larger for a Tezzaron 3D IC than a normal 2D IC [6]. The primary purpose of the test chip described in this paper is to measure and characterize the process variations of the Tezzaron/Globalfoundries 3D technology and confirm if 3D IC process variations have a significantly high variance.

2.4.2 TSV-Induced Mechanical Stress

One of several issues associated with embedding TSVs in silicon is mechanical stress. 3D ICs that utilize TSVs must tunnel them vertically through the tiers of the die, crossing through layers of dies normally reserved for active devices. These TSVs can impact the performance of nearby devices by exerting mechanical stress on them. This stress is primarily caused during the fabrication of an IC. The metal being deposited in the silicon to form the TSV is typically made from copper, which has a coefficient of thermal expansion (CTE) of $17{\times}10^{-6}{\times}K^{-1}$, compared to silicon's CTE of $3{\times}10^{-6}{\times}K^{-1}$. Because of this CTE mismatch, and the high temperatures required to fabricate the IC, tensile stress will be exerted onto the silicon from the TSV after the IC has returned to room temperature after fabrication [25].

When this stress exerts itself on the gates of nearby NMOS and PMOS transistors, it will induce a change in their electron and hole mobilities, respectively, which will alter the performances of these devices [25]. The direction of the tensile stress vector is important as well, as it can reduce or increase the mobility of the device depending on the alignment of the gate with respect to the direction of the applied stress [10, 25]. If stress in this scenario is considered to be acting nearby devices along a 2D plane as radial stress, then the stress can be modeled as

$$\sigma = \frac{B\Delta\alpha\Delta T}{2}(\frac{R}{r})^2 [25]$$

where σ is the applied stress, B is the biaxial modulus, $\Delta\alpha$ is the CTE difference between silicon and the material of the TSV, ΔT is the temperature difference between the manufacturing temperature of the TSV and the operating temperature of the IC, R is the radius of the TSV, and r is the distance from the centre of the TSV [25]. This model assumes that the TSV is cylindrically shaped (the Tezzaron/GlobalFoundries TSVs are shaped octagonally, but can be approximated as cylinders). The change in mobility can be derived from the stress as

$$\frac{\Delta\mu}{\mu(\theta)} = \Pi \times \sigma \times \alpha(\theta)[25]$$

where $\frac{\Delta\mu}{\mu(\theta)}$ is the change in mobility, x is the piezo-resistive coefficient of the gate relative to its orientation to the TSV, σ is the applied stress, and $\alpha(\theta)$ is the orientation factor, a function dependent on θ, the angle of orientation of the device's gate relative to the TSV [25]. Simulations made to verify these models typically find the change in mobility of nearby devices to be between -10% and 10%, depending on whether they were N-type or P-type devices [10, 25]. The test structure described in this paper will attempt to verify this model, as there is little published data for real measurements of TSV stress. However, the importance of the stress measurements of this test structure is minor compared to the characterization of process variations due to the design of TSVs in the Tezzaron/GlobalFoundries technology. These TSVs are relatively small, and the above equation shows that the applied stress of a TSV is proportional to its radius (R), i.e. its size. Furthermore, the Tezzaron TSVs are constructed from tungsten, which has a CTE of only $4.6\times10^{-6}\times K^{-1}$. The equation for applied stress shows that applied stress is proportional to $\Delta\alpha$. Since previous works only reported signifcant mobility variation using simulations of the applied stress model assuming the TSV material was copper (which has a CTE of $17\times10^{-6}\times K^{-1}$), it is likely that the impact tungsten TSVs have on $\frac{\Delta\mu}{\mu(\theta)}$ will be smaller, and that this variance may be negligible compared to the impact of process variations.

Chapter 3

Test Chip Design

The test chip was created using a 130nm Tezzaron/GlobalFoundries 3D process. It contains several copies of a test structure, designed for the dual purposes of measuring and characterizing process variations and TSV-induced stress. Following the precedent set by other test structures intended for the characterization of process variations, the test structure consists of a large array of CMOS devices, and peripheral circuitry used to measure those devices [21, 8]. Each device-under-test (DUT) consists of a single transistor: either an NMOS transistor or PMOS transistor. Data points from the I_{DS} vs V_{DS} and I_{DS} vs V_{GS} curves (or current-voltage characteristics) can be measured from these DUTs, and their device parameters, along with the effects of TSV stress, can be extracted from these measurements.

3.1 Design Requirements

The NMOS DUT and PMOS DUT test structures were designed under the consideration of three requirements. These requirements were established to maximize the usefulness of the test chip for its intended purpose and to conform to limitations of the available design area. The first requirement is to design the DUT arrays to have high densities and spatial resolutions, which will allow them to provide more measured data, which will give higher

statistical confidence to the measurement of the DUT's process variations [21]. Having a high spatial resolution is also critical since the effects of stress and process variation are distance dependent. The second requirement is to create a design that uses minimal I/O. This requirement is driven by the limitations of the fabrication of the IC described in this paper, as this IC's fabrication run allowed a maximum design area of only 1mm $\times$ 2mm. This greatly limits the number of I/O pads that can be placed on the IC's periphery, and subsequently limits the number of measurements that the IC can make. The third requirement is that the test structures should have a high fidelity; current-voltage characteristic measurements should have low noise and negligible sources of error, so that the measurements of variability and stress are as error-free as possible. Note that the measurements that will be made are DC measurements, and not AC or transient measurements.

3.2 Related Test Structure Designs

Because of these three design requirements, the most straightforward design for the DUT arrays, shown in Fig. 3.1, cannot be feasibly implemented. In the NMOS version of this design, the source terminals of each DUT are connected to a common ground, their gate terminals are connected to an off-chip voltage, and each drain terminal is routed to a separate I/O pad to measure their current-voltage characteristics. In a dense array with many DUTs, the number of I/O pads required would far exceed the limited area allocated for the pad, and the routing required to connect each device to its pad would lower the overall density of the DUT array. The additional routing would also result in lithographic differences in the areas surrounding each DUT. This is undesirable as these differences could affect the current measurements taken from each DUT, which would makes it more difficult to characterize the variations across the devices.

For the purpose of meeting the design requirements outlined in section 3.1, an improved design would be the structure proposed by Saxena *et al* [21], shown in Fig. 3.2. In this design, all DUTs in the array have their drain terminals connected together and routed out

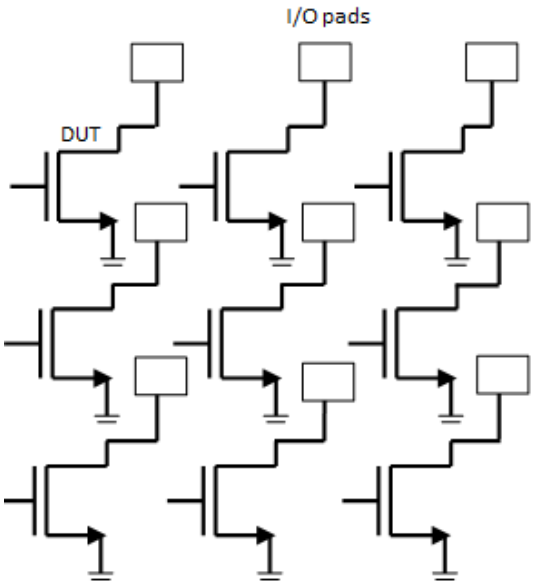

Figure 3.1: A simple DUT array: one I/O pad per DUT.

to an I/O pad. Peripheral circuitry is used to select a row for measurement by supplying each DUT of that row with a gate voltage. Each column I/O pad can then be used to measure the current of each DUT in that row. Rows that have not been selected by the row-selection logic instead receive a signal of 0V, keeping them off during the measurement of the other DUTs. This design can potentially use only a few I/O pads if the dimensions of the DUT array are adjusted to allow for only a few columns and many rows. However, such a design is infeasible under the design requirements for the 3D test structure. Such a design would only have a high spatial resolution along the dimension of the columns. The DUT array should be designed so that its dimensions are relatively square to give it a high spatial resolution along the plane of the IC.

The design proposed by Drego *et al* [5], shown in Fig. 3.3, offers an improvement on the indexed-column DUT array by reducing the number of required I/O pads to one. The design uses essentially the same peripheral selection circuitry and column-wiring, but all column wires are now routed into an analog multiplexer (MUX). The multiplexer consists of a number of pass-gates controlled by a few off-chip signals which allow it to pass a

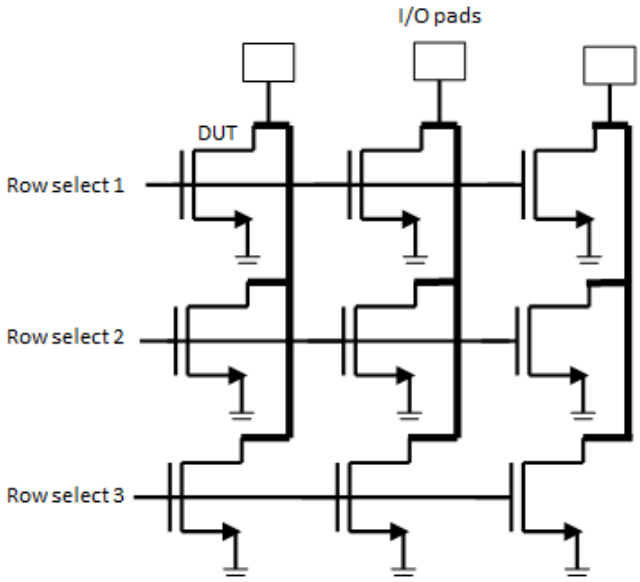

Figure 3.2: An indexed-column DUT array. [21]

selected column signal of the DUT array to a single I/O pad.

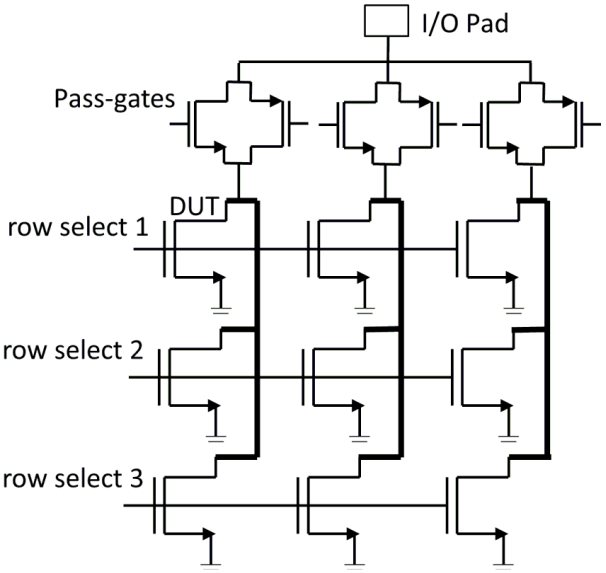

Figure 3.3: An analog MUX DUT array. [5]

While the inclusion of an analog MUX improves the I/O pin count of the DUT array, it also lowers its measurement accuracy. When a pass-gate of the MUX is active, the NMOS and PMOS transistors will each have a drain-to-source resistance r_{DS} that is dependent on the voltage applied to their gates (V_{GS}). When a measurement of a DUT is being made from an I/O pad, r_{DS} will cause a voltage drop across the pass-gate, and so the voltage seen at the I/O pad will differ from the voltage at the drain of the measured DUT, as illustrated in Fig. 3.4. This voltage drop will add a significant error to the current measurements taken from the DUT array. This error can be mitigated by reducing the voltage drop of the pass-gate. This can be accomplished by increasing the $\frac{W}{L}$ ratio of the pass-gate's transistors, since r_{DS} of a transistor is inversely proportional to its $\frac{W}{L}$ ratio. However, increasing the ratio to a point where the voltage drop is negligible would result in a large increase in the area of the MUX. Since the design of the 3D test structure is limited by the maximum design area, this increase would result in a reduction of the size of the DUT array, reducing its spatial resolution along with the total number of unique measurements that could be taken from it.

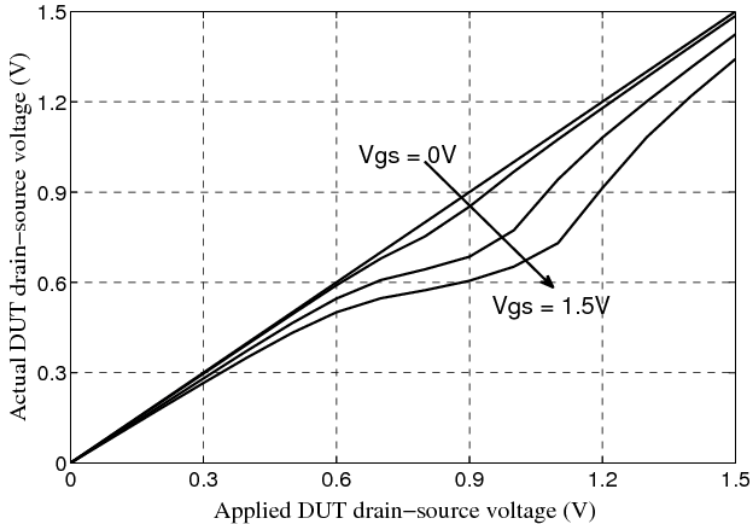

Figure 3.4: Voltage drop across an analog MUX.

Hess *et al* [8] proposed another test structure that that also reduces the I/O pad count

to one while avoiding the error introduced by an analog MUX. In this design, shown in Fig. 3.5, the control signals for the gate of each DUT are generated on-chip by complex peripheral circuitry. The drain terminals of every device are routed together to a single I/O pad, and the control signals are used to measure each DUT from this I/O pad, one at a time. However, the routing of these control signals is a complicated design problem, and the area needed to route a separate signal to every DUT would lower the spatial density of the array. The layout of the signal wires would also cause non-uniformity in the lithographic areas surrounding each device.

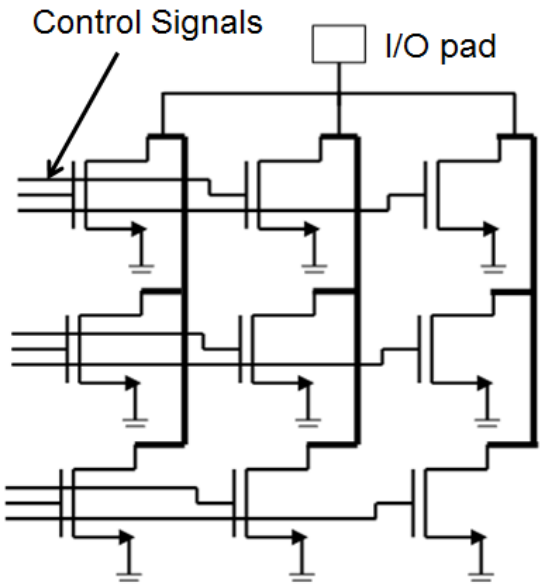

Figure 3.5: DUT array using on-chip control signal generation [8].

Other test structures have also been proposed for the characterization of stress on the performance of NMOS and PMOS transistors [17, 13]. These test structures also propose using single-transistor-device arrays to measure TSV stress by embedding TSVs within the arrays and measuring the devices surrounding them. This orients the TSVs to be aligned in longitudinal and tranverse directions of many transistors, allowing for a high spatial resolution of measurement [13]. Yu proposes a TSV stress characterization array, shown in Fig. 3.6, which connects its DUTs to I/O pads for measurement via an array of

switches, similar to the previously introduced analog MUX design [26, 5]. In this design, the voltage drop error across the switches of the I/O selection array is reduced using a kelvin measurement technique. While useful for mitigating voltage drop error, a kelvin measurement technique requires 4 I/O pads, and is not feasible given the limited I/O area of the test chip's design.

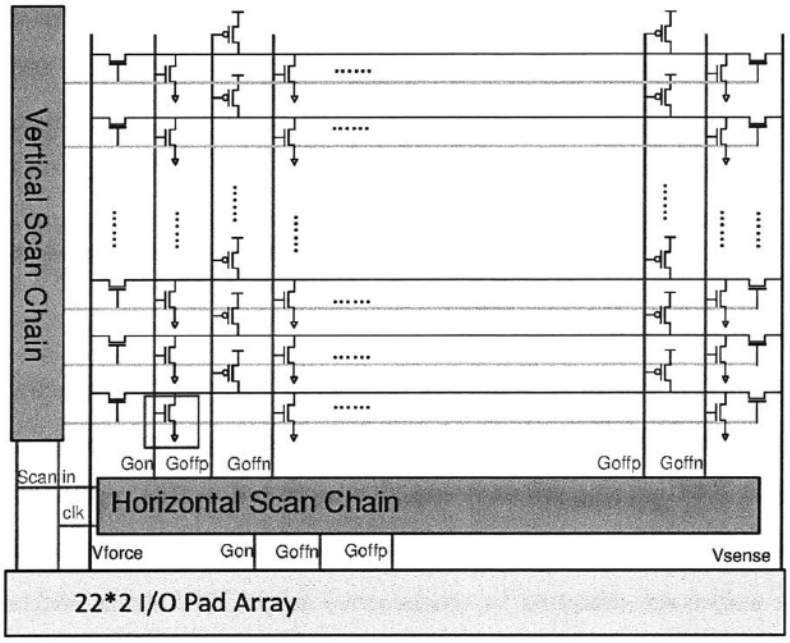

Figure 3.6: TSV variation test structure [26].

3.3 Test Structure Design

The test structure proposed in this thesis bases its design off of the previous test structures of section 3.2 by using a large DUT array of single NMOS and PMOS transistors to measure the device parameters needed to characterize process variations and TSV-induced stress. Additionally, it was designed to meet the three requirements described in section 3.1: the test structure is capable of taking a large number of measurements with a reasonable fidelity

from a single I/O pad. The proposed design also takes advantage of the 3D architecture of the IC to increase the density of the DUT array.

The proposed test structure is shown in Fig. 3.7. To ensure that the DUT array can be designed with the highest density possible, the circuitry of the test structure is divided between the two tiers of the 3D IC. All of the selection logic required to address and control the DUTs is placed on one tier of the IC, while the DUT array is placed on the other tier. The gate terminal of each DUT is addressed by its own gate select cell, and this connection is made vertically across the two tiers of the 3D IC via inter-tier copper bondpoints.

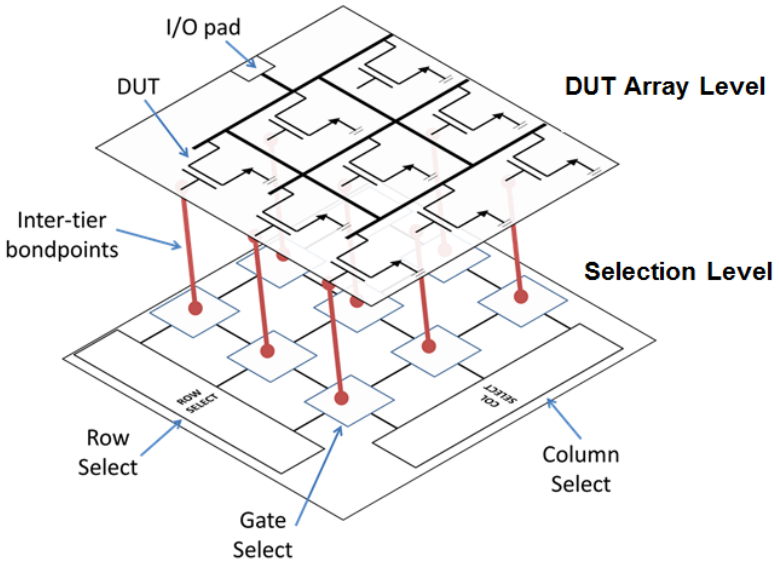

Figure 3.7: 3D test chip measurement structure.

3.3.1 Selection Level Design

The selection level of the test structure contains all of the circuitry necessary for selecting and controlling each device in the DUT array. Each DUT in the array is controlled by a gate select cell, and each gate select cell is connected together in the selection array. These gate select cells are controlled by the row select and column select scanchains located at

the periphery of the selection array. The scanchain design is shown in Fig. 3.8. These scanchains generate the row select and column select signals that turn the gate select cells on. Using a combination of CLK, Enable, Set, and Reset signals controlled off-chip, the scanchains can select a gate cell at a particular row and column location to enable the measurement of specific DUT.

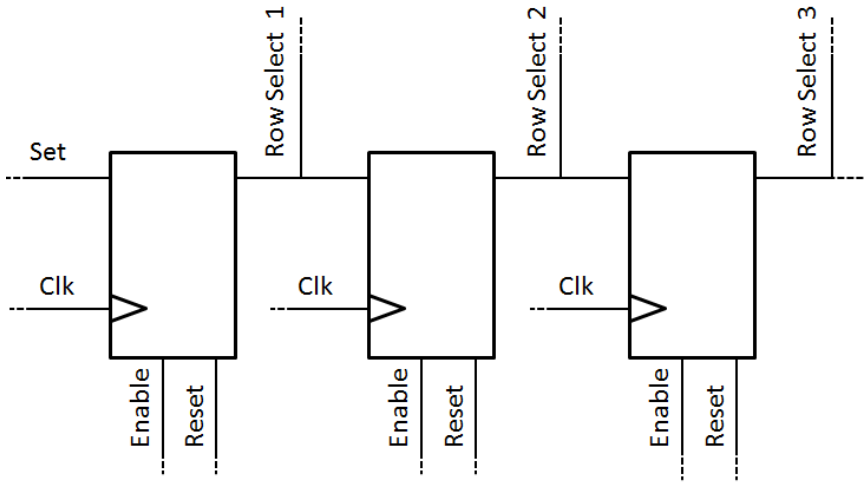

Figure 3.8: Row select scanchain for the selection logic array.

Each gate select cell has two functions: when it has been selected via signals from the row select and column select scanchains, it passes an adjustable, off-chip signal to the gate of its corresponding device in the DUT array. When it is not being selected to do this, it sends a voltage to the gate of its corresponding DUT that turns that device off.

The logic for a gate select cell for an NMOS DUT and PMOS DUT is shown in Fig. 3.9. Both feature a pass-gate along with simple logic that controls its operation. When the gate select cell receives "on" signals from both the column and row select scanchains, the pass-gate turns on and passes the off-chip signal V_{GS} to the gate of a device in the DUT array, located in the other tier of the IC. When the gate select cell is not receiving an "on" signal from both the column and row select scanchains, another transistor is turned on and passes a voltage signal to the DUT that will keep it off, so that the voltage at the gate of the DUT will not be left floating, and will not interfere with the measurements of other DUTs.

The NMOS gate select cell routes this off connection from the gate of the DUT to ground, while the PMOS gate select cell routes an off connection from the gate of the DUT to V_{DD} (nominally 1.5V in the Tezzaron/GlobalFoundries 3D technology). While these circuits and the connections routing them to the other tier have the potential to introduce unwanted parasitic capacitances and inductances to the measurement scheme of the test chip, said parasitics can be circumvented by taking only DC measurements from the test chip. This comes at a cost of slowing down the speed of the test chip's measurement procedure.

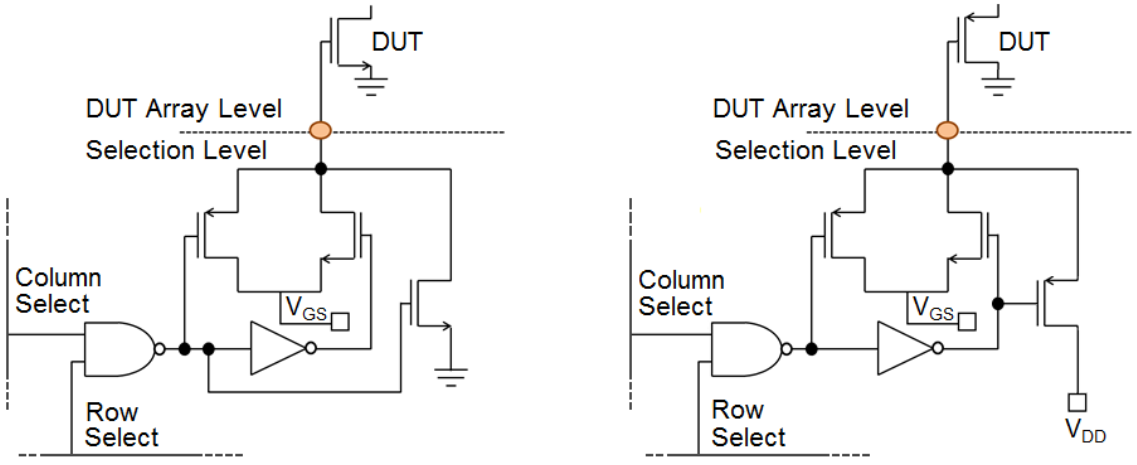

Figure 3.9: Gate Select Cell for an NMOS DUT (Left) and for a PMOS DUT (Right).

The gate select array contains 40 rows and 48 columns of gate select cells, one for each DUT in the corresponding DUT array on the other tier of the 3D IC. The layout of the gate select cell is shown in Fig. 3.10. The area of each cell is 4.41μm $\times$ 4.41μm. To simplify routing between the selection level and DUT level of the test structure, the size of each DUT cell in the device array is fixed to 4.41μm $\times$ 4.41μm as well, to ensure that the density and pattern of the copper bondpoints remains constant throughout the array.

3.3.2 DUT Array Level Design

The DUT array level hosts the DUT array of the test structure. Like the previous designs described in section 3.2, every DUT in the array is either a single NMOS or PMOS tran-

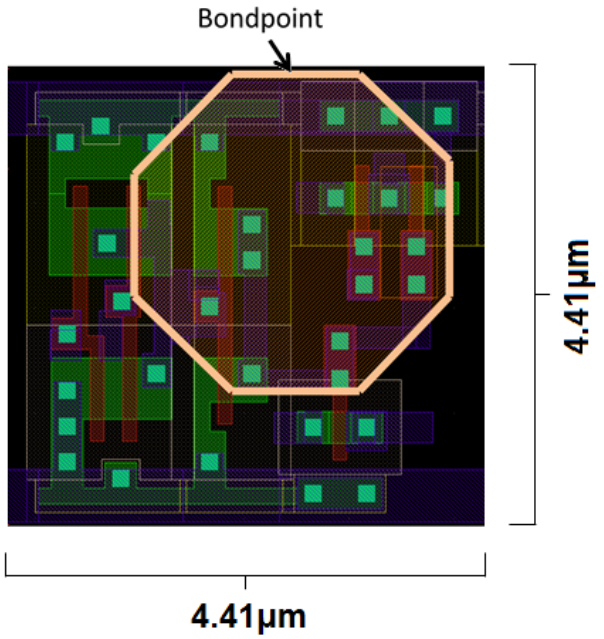

Figure 3.10: Gate Selection Cell Layout.

sistor. The designs of these DUT arrays are shown in Fig. 3.11. The drain terminals of each DUT are routed together in a mesh and connected to a single I/O pad. The I/O pad measures the current of each transistor as they are turned on one at a time. Each DUT is measured by modifying the voltage at its drain (from off-chip, through the I/O pad) and the voltage at its gate. An adjustable, off-chip gate voltage is supplied to it by its corresponding gate select cell, located on the other tier of the 3D IC, on the selection level of the transistor.

Each NMOS and PMOS DUT array consists of 40 rows and 48 columns of DUT cells. The area of each DUT cell is 4.41μm x 4.41μm, ensuring that the DUT array's total area will match the area of its gate select array on the other tier of the 3D IC. Additionally, the transistors of each DUT cell come in a set of different sizes to provide a larger sample of results for the characterization of process variations. The device parameters that are to be measured from the DUTs of the test structure for this purpose should have their standard deviations vary as $\frac{1}{\sqrt{WL}}$ [16]. Additionally, devices with larger areas, separated

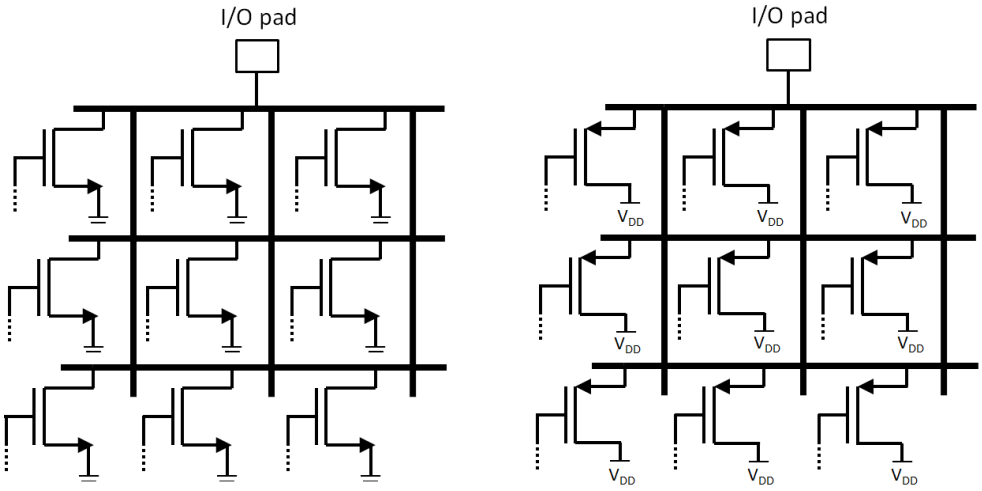

Figure 3.11: NMOS DUT Array (left) and PMOS DUT Array (right) of the Test Structure.

by longer distances, are expected to have greater device parameter mismatch. To confirm these effects, the gate lengths and widths of the devices of the DUT array are sized at different factors (i.e. $\times 1$, $\times 2$, $\times 4$, $\times 8$, $\times 16$, and $\times 32$) of the 3D technology's minimum transistor gate width (300nm). The $\times 16$ and $\times 32$ transistors are larger than the default DUT cell size of 4.41μm $\times$ 4.41μm, and therefore occupy the equivalent area of 4 and 16 DUT cells, respectively. Four columns in the DUT array are also oriented 90 degrees from the others. This gives them a different alignment along the crystal lattice orientation of the silicon wafer, which could create a measurable difference in its TSV-stress-induced mobility variation, and provides process variation measurements taken from devices angled at a different orientations along the silicon crystal lattice. Fig. 3.12 shows the size and orientations of transistors in the test structure's DUT array. hlThere are a total of 912 DUTs in each array on the test chip, and the density of each device array is $40.94\mu m^2$ per device.

10 TSVs are also embedded in the DUT array to measure the mobility variation of nearby devices. They are positioned among the minimum-sized devices in the array, since these are the devices that should experience the largest shift in mobility from TSV-induced mechanical stress. To fit the TSVs into the array, a single DUT is removed and replaced

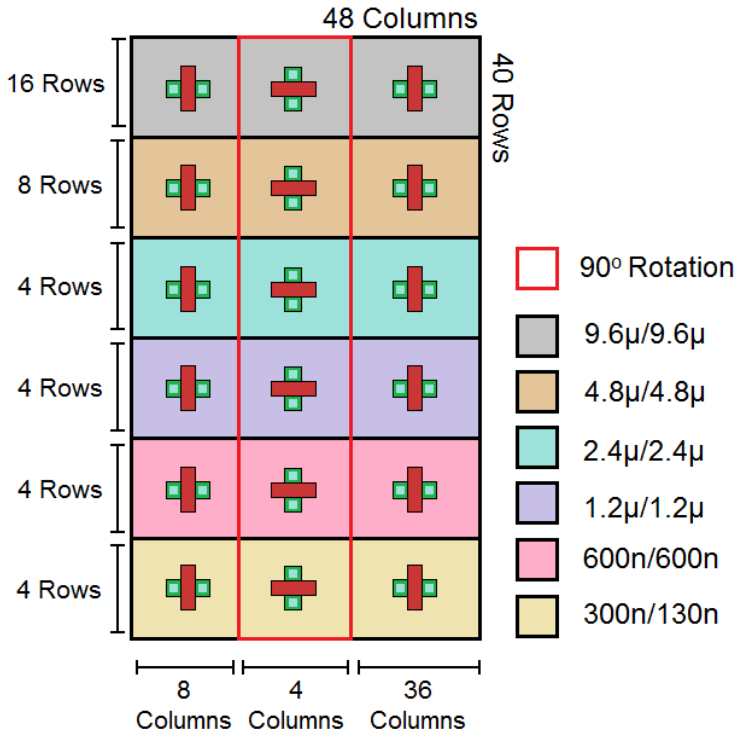

Figure 3.12: Size and Orientation of transistors in the test chip's DUT arrays.

with an unconnected TSV. The DUTs that surround the TSV can then be measured to provide TSV-induced mobility variations at different orientations with respect to the TSV.

3.4 Test Chip Architecture

The full layout of the Test chip is shown in Fig. 3.14. This layout design has been submitted for tape-out and is currently undergoing fabrication. In Tezzaron's 3D IC process, only the top tier of the 3D IC is thinned down to allow for the exposure of I/O pads, so only one tier has I/O pads at its periphery. There are 36 pads on the top tier of the test chip. Approximately half of the I/O pads deliver power (3.3V and 1.5V) and ground lines to both tiers of the chip, while the rest provide outputs for the DUT arrays and inputs for

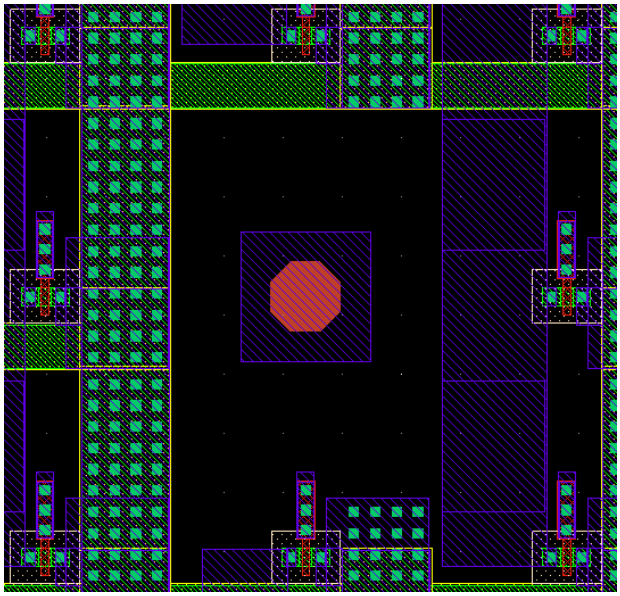

Figure 3.13: Layout of a TSV embedded in the DUT array.

the control signals for the test structures (CLK, Enable, Reset, etc). The remaining area of the chip hosts eight test structures. Four of these test structures have NMOS DUT arrays (N1, N2, N3, an N4) and the other four control PMOS DUT arrays (P1, P2, P3, and P4). In total, there are roughly 3600 NMOS devices and 3600 PMOS devices available for the measurement and characterization of process variations. Two NMOS DUT arrays (N1, N3) and two PMOS DUT arrays (P1, P3) are located on the top tier of the IC while the remaining DUT arrays (N2, N4, P2, P4) are located on the bottom tier of the IC. This allows for measurements to be taken on both tiers of the IC to discern if there is noticeable inter-tier mismatch across the two tiers of the test chip.

A comparison of the physical parameters of the 3D test chip to the other designs is shown in Table 3.1. The 3D test chip has the fewest total number of devices per chip due to the limited design area of the tape-out. The low number is also due to the large sizes of some of the devices in the array. If only minimum-sized devices were used in the test chip's device arrays, the number of total devices that could be housed on the chip would be 15360. Although any conclusions drawn from these comparisons are not

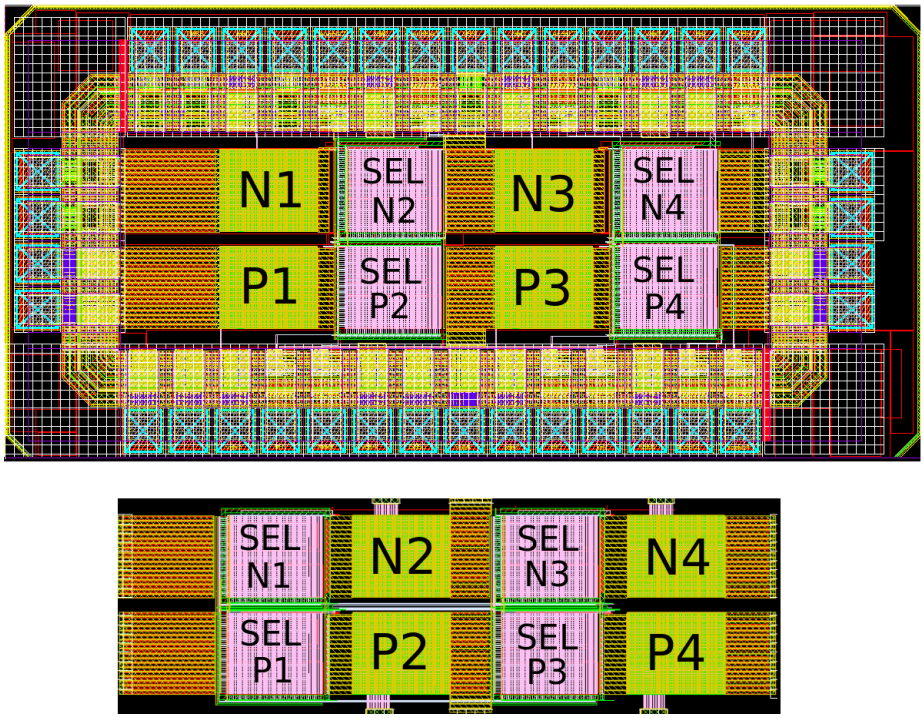

Figure 3.14: Full test chip layout showing the top and bottom tiers of the 3D IC. 'N' refers to a DUT array of NMOS transistors, 'P' refers to a DUT array of PMOS transistors, 'SEL' refers to gate selection logic.

neccesarily equitable (since each chip is made using a different CMOS technology and likely have different design requirements and limitations, such as design area), the 3D test chip contains the second highest devices per array, dwarfed only by the MUX-based chip. However, for this to be a meaningful comparison, array density would have to be considered as well, and this data is not available for the other chips.

Table 3.1: Physical parameters of the test chip compared to the chips described in [21, 8, 5]

	3D test chip	MUX-based chip [5]	Column-addressed chip [21]	Scribe CV chip [8]
process node	130nm	0.18μm	130nm	65nm
# of devices	7296	137160	24576	$\sim$25000
devices per array	912 (48×40)	11430 (127×90)	128(32×4)	32(8×4)
area	2mm×1mm	3.1mm×2.6mm	-	-

Chapter 4

Variability and Stress Measurements

Accurate measurements are required to characterize the effects of process variations and TSV induced stress on device performance. Two techniques can be used to mitigate error in current measurements and discern process variation from variation in stress. These techniques will be applied to the measurements taken from the test chip upon its return from fabrication.

4.1 Mitigating Current Error

Leakage current contributes significant error to measurements of the test chip's DUT currents (this is demonstrated in section 5.1). Due to the small area of the chip, only a limited number of I/O pads are available for the measurement of DUT current. To compensate for this restriction, the test structures were designed to have only one output pin from which to measure the current of their DUTs. Each DUT can be turned on and measured one at a time while the others have their gate voltages set so that they will be off and cannot interfere with this measurement. However, even while in an off state the transistors have a small leakage current, and this "background" current affects every current measurement of the test structure. While the background current is negligible for DUT measurements taken while the transistor is generating a large current in linear or saturation operation modes,

other modes of operation (such as $V_{GS} < V_{th}$ or low V_{DS}) will generate smaller currents, and the background current will contribute a significant error to the measurement.

For an NMOS DUT array, the current measured from the array when attempting to measure the current of DUT i can be expressed as

$$I^i_{meas} = I^i(V^{test}_{GS}, V^{test}_{DS}) + \sum_{j \neq i} I^j(0, V^{test}_{DS})$$

Where I^i_{meas} is the current measured from the DUT array, $I^i(V^{test}_{GS}, V^{test}_{DS})$ is the current of DUT i with an applied gate voltage V^{test}_{GS} and an applied drain voltage V^{test}_{DS}, and the final term is the summation the leakage current of all other DUTs in the array other than DUT i, which experience gate voltages of 0V and a drain voltages of V^{test}_{DS}. This last term is the sum of the leakage current and introduces error to the measurement of DUT i's current.

This measurement error can be greatly reduced using background subtraction. Before taking measurements of individual DUTs, all DUTs have their gate voltages set so that they are all off. A measurement of the combined background leakage current, I_{bg}, is then measured from the array. For an NMOS DUT array, I_{bg} can be expressed as

$$I_{bg} = \sum_i I^j(0, V^{test}_{DS})$$

This background current is then subtracted from the current measurements of the DUTs to yield the final current measurement $\hat{I}^i_{meas}$, which is expressed as

$$\hat{I}^i_{meas} = I^i_{meas} - I_{bg} = I^i(V^{test}_{GS}, V^{test}_{DS}) - I^i(0, V^{test}_{DS})$$

As a result of background subtraction, the current measurement's error is now only as large as the leakage current of a single DUT in the array. Because of this reduction in error, current measurements can be taken from the test chip for lower drain voltages and subthreshold gate voltages.

4.2 Distinguishing Stress and Process Variation

The test chip must be able to characterize process variations and TSV-induced stress, and to do so requires accurate measurements of both. However, due to the nature of process variations, a measurement taken from of the chip meant to quantify TSV stress will be affected by process variations as well. This issue is a result of the test chip's measurement technique, where all measurements are taken from the drain currents of transistors distributed over the area of the chip. When measuring transistors located near TSVs, the measurements would be expected to reflect a pattern of mobility variation like those predicted in Yang, *et al* [25]. Instead, the currents measured are affected not only by a change in mobility, but by variations in the transistors' device parameters caused by process variation.

An example of this is shown in Fig. 4.1. The spatial plot on the left shows the normalized current measured from NMOS DUTs surrounding a single TSV. With only the effects of TSV stress taken into account, the current shows a predicted radial, $\propto \frac{1}{r^2}$ mobility variation pattern [25]. The spatial plot on the right shows the same measurements with process variation effects added. The addition of these variations makes the pattern of mobility more difficult to discern.

To distinguish mobility from process variations, a radial filtering technique is used to diminish the impact of process variations from the current measurements. Based on an averaging technique from Chang *et al* [2], radial filtering reduces the error introduced by process variations by taking advantage of the symmetry of the mobility variation pattern of a TSV. Assume that the TSV is centered at location (x_{TSV}, y_{TSV}) and DUT i is at location (x_i, y_i). In polar coordinates, DUT i is at location (r_i, θ_i), with respect to the TSV where

$$r_i = \sqrt{(x_i - x_{TSV})^2 + (y_i - y_{TSV})^2}$$

And

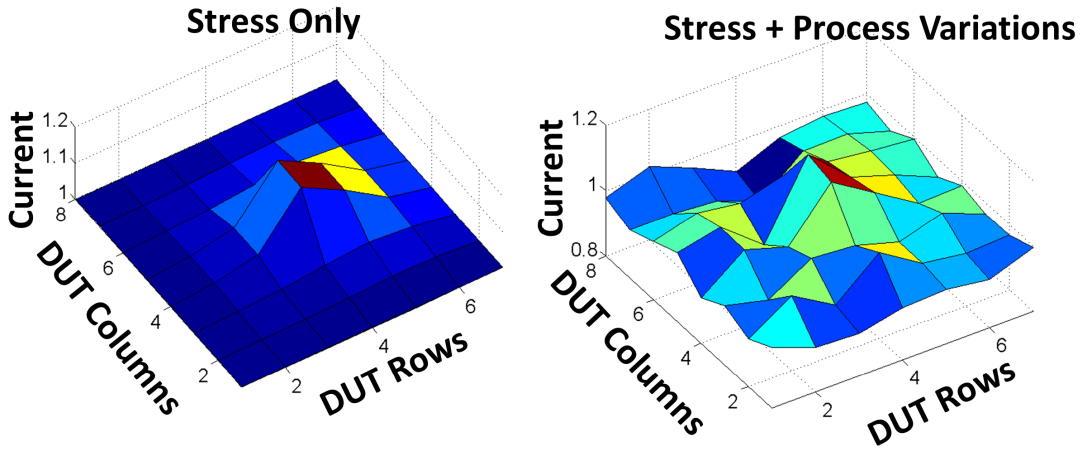

Figure 4.1: Impact of TSV stress on DUT current (left), and the impact of TSV stress and process variations on DUT current (right).

$$\theta_i = tan^{-1}(\frac{y_i - y_{TSV}}{x_i - x_{TSV}})$$

Based on these definitions, the current of DUT i, accounting for only TSV-induced mobility variation ($\hat{I}^i_{stress}$) can be estimated as

$$\hat{I}^i_{stress} = \frac{1}{4} \sum_{\{j: r_i = r_j, \theta_j = \pm\theta_i, \pi \pm \theta_i\}} \hat{I}^i_{meas}$$

The above estimation is valid because the four currents that are averaged are taken from DUTs that are equidistant from the TSV, and are expected to have the same stress-induced mobility variation. This averaging can potentially reduce the process variation error from the measured current.

Chapter 5

Simulation Results

This chapter describes the simulations used to verify the functionality and measurement accuracy of the test chip compared to a multiplexer-based design. It provides a comparison of the test chip's single I/O pad measurement setup to that of a similar multiplexer-based measurement approach, and demonstrates the error reduction of the background leakage technique. It also compares the test chip's ability to accurately measure process variations in drain current compared to a multiplexer-based test structure. Finally, it provides simulations of mobility variation from a TSV embedded in the DUT array of the test chip and the techniques used to discern mobility effects and process variation effects from simulated DC drain currents.

5.1 Measurement Accuracy

Simulations were made to compare the accuracy of the test chip's simulated measurements to that of similar designs. The test structure design used for these comparisons was reviewed in section 3.2, and is shown in fig. 3.3. It is an analog multiplexer-based measurement scheme based on multiplexer test structures described in prior works [5]. As described in section 3.2, This test structure uses a multiplexer to select devices from separately connected rows of a DUT array. Compared to the design of the 3D test structure, the

multiplexer design should have a smaller leakage current error associated with each simulated measurement. However, implementing this design under the same low I/O design constraint of the 3D IC test structure requires the multiplexer design to use only a single I/O pad for its measurements. Without a multi-I/O measurement scheme to calibrate the voltage drop error introduced by the multiplexer, this design has worse accuracy than the 3D test structure.

To simulate this error, a multiplexer-based test structure was created and simulated using Cadence/Spectre simulation tools. The multiplexer test structure was designed to be similar to the 3D test structure; it was also created with 40 rows and 48 columns of devices, and the devices were sized in the same manner as the 3d test structure's are, using channel width and channel length ratios of 300nm/130nm, 600nm/600nm, 1.2μm/1.2μm, etc., as described in section 3.3.2. Fig. 5.1 shows the error of the simulated DC drain current of one of the devices of the multiplexer-based test structure. This device's channel length to width ratio is 2.4μm/2.4μm. The error shown in the figure is based on the difference between the DC current simulated from the multiplexer-based design and the DC current of an ideal transistor with the same dimensions and applied V_{DS} and V_{GS}. The figure shows that the error is negligible for smaller values of V_{DS} and V_{GS}, but for large values of V_{GS} and mid-range (0.8V - 1V) values of V_{DS} the voltage drop across the multiplexer becomes significant and the error becomes as high as 16%. There is also a smaller error of 4% for very low V_{GS} simulated DC current taken from the device. This error is due to the combined leakage current from the rest of the devices located in the same column as the simulated device. This leakage current is small and only becomes significant when the simulated current reduces to very small values.

To complete the comparison, a simulation was made measuring the DC current of a 2.4μm/2.4μm sized DUT from the 3D test structure for sweeping values of V_{DS} and V_{GS}. This simulation was made using an *RC* parasitic netlist, extracted from the completed Cadence layout of the test chip. The error of these simulated DC currents is shown in Fig. 5.2. Once again, the error was measured as the difference between the simulated DUT current and the simulated current of an ideal transistor with the same dimensions and applied V_{DS} and V_{GS}. The error is very small for all simulated measurements taken

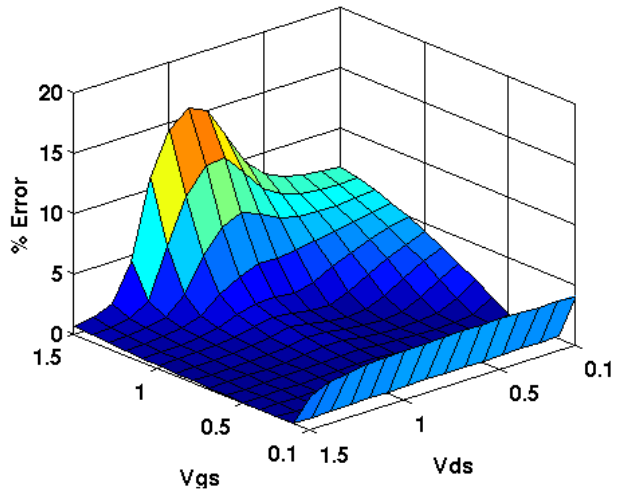

Figure 5.1: Simulated DC current error of a single device in a multiplexer-based test structure.

with a high applied gate voltage, but becomes overwhelmingly large for values of $V_{GS} <$ 0.1V. This error is caused by the combined leakage current of every other device in the test structure, and contributes a total current of approximately 10nA to each measurement.

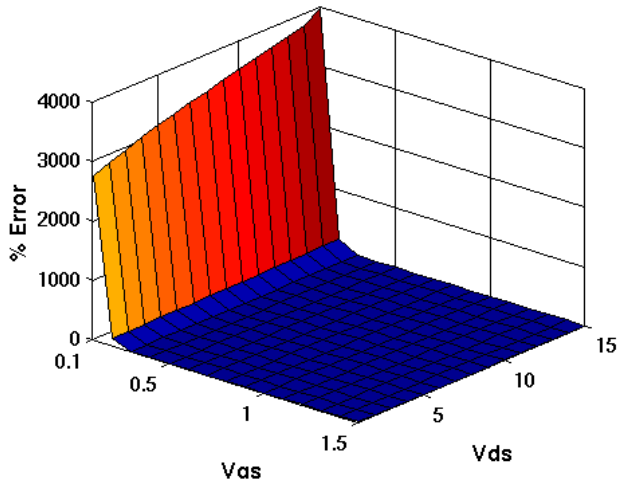

Figure 5.2: Measured current error of a single device in the 3D test structure (without background subtraction).

This error can be significantly reduced through the use of background subtraction, a technique described in section 4.1. The combined leakage current of the total test structure was measured, and this value (approx. 10nA) was subtracted from the simulated DC currents shown in Fig. 5.2. The result is a significant reduction in subthreshold current error, as shown in Fig. 5.3. Here, the low V_{GS} error is reduced by a factor of 500, to a maximum of 5.5%. This represents a flaw in the test structure's design: even with error reduction techniques, very low sub-threshold measurements will have a significant error, and therefore inaccuracy. However, the error for simulated DC currents measured for V_{GS} > 0.1V is somewhat low, and pertinent device parameters such as threshold voltage can still be extracted from these simulated results with a certain degree of accuracy.

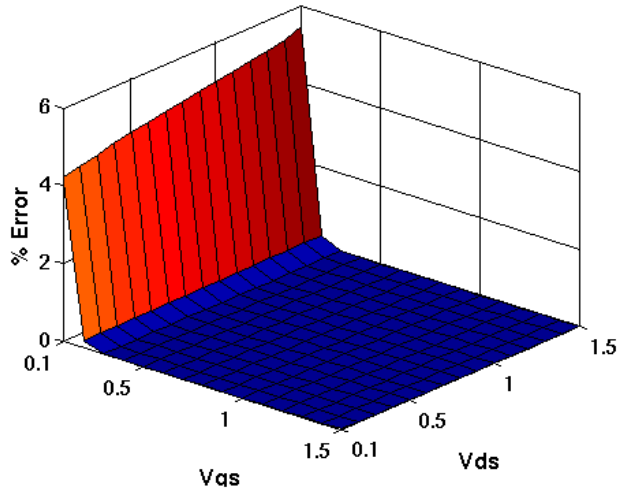

Figure 5.3: Measured current error of a single device in the 3D test structure (with background subtraction).

The RC parasitic netlist of the test structure also demonstrates the low resistive voltage drop error of the test chip, relative to the multiplexer design. Fig. 5.4 shows spatial plots of current error for each structure. It shows plots of the simulated DC current error for every device in every row and column of each structure, for an applied V_{GS} and V_{DS} of 1.5V. Even without RC parasitics taken into account, the overall error of the multiplexer is significantly higher than the 3D test structure, with a worst-case error of 0.62%. The 3D test structure's parasitic resistance was kept relatively low during its layout by wiring its devices together

in a mesh, using wide metal lines to lower the resistance of the interconnects. As a result, the worst-case error is only 0.21%. This worst-case error is measured from devices located furthest from the I/O pin of the structure.

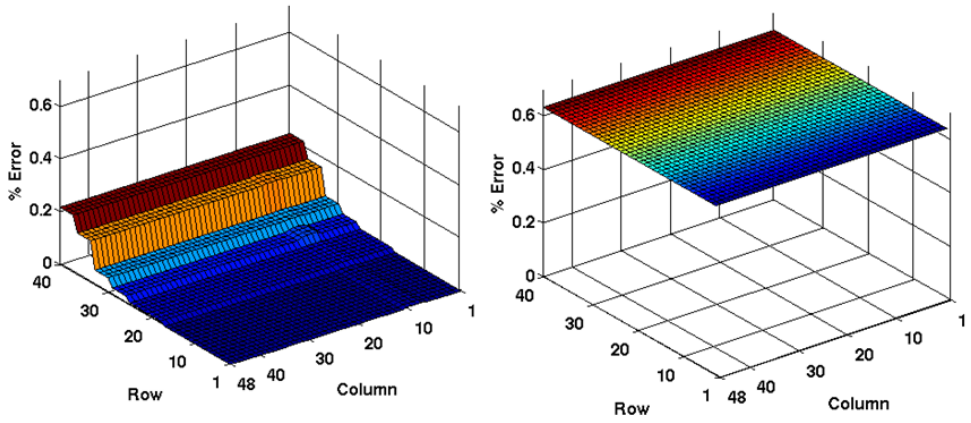

Figure 5.4: Spatial current error of the 3D test structure (left) and multiplexer structure (right).

5.2 Variability

To test the accuracy of the test chip's measurement of variability, Monte Carlo simulations were performed on it using Cadence Spectre's process and mismatch models, which introduce variations in device parameters including oxide thickness, drain and source region dimensions, and threshold voltage. The results are graphed in Fig. 5.5, simulating the DC current of devices with applied V_{GS} and V_{DS} values of 1.5V. The results are graphed along with Monte Carlo simulations of devices from the multiplexer structure under the same conditions, as well as Monte Carlo simulations of an ideal transistor. The error between the test chip structure's estimated mean and standard deviation and the ideal mean and standard deviation are relatively small (0.8% and 11%, respectively), and are comparable to the multiplexer (MUX) test structure's error. This is not surprising since the resistive error for both structures is relatively small at this applied V_{GS} and V_{DS}. Fig. 5.6 shows

similar Monte Carlo simulations for V_{GS} of 1.5V and a V_{DS} of 0.8V. The test chip structure estimated mean and standard deviation error is relatively unchanged (0.3% and 10.6%, respectively), but the mean estimated by the MUX test structure is very large, as expected from the results shown in Fig. 5.1. The MUX's mean error at these values of V_{GS} and V_{DS} is 14.2%.

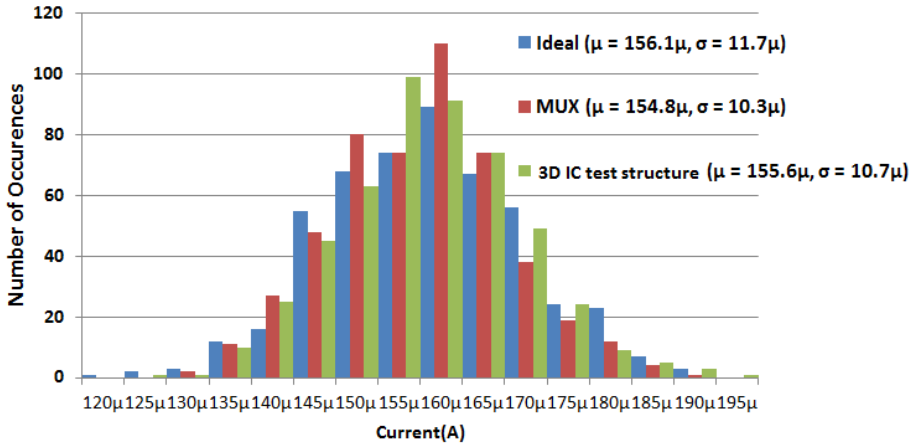

Figure 5.5: Monte Carlo simulations of the DUT current of an ideal device, a DUT in the test chip structure's array, and a DUT in the MUX array (V_{GS} = 1.5V, V_{DS} = 1.5V, number of runs N = 500).

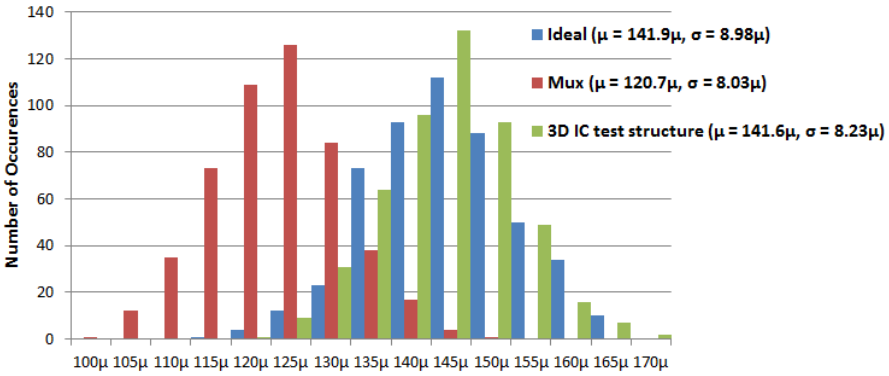

Figure 5.6: Monte Carlo simulations of the DUT current of an ideal device, a DUT in the test chip structure's array, and a DUT in the MUX array (V_{GS} = 1.5V, V_{DS} = 0.8V, number of runs N = 500).

Welch t tests were performed on these distributions to determine the probability that the mux and test chip distributions really do have different means from the ideal distribution, and to determine which distribution is more likely to match more closely to the ideal distribution. The Welch t test is only valid for Gaussian distributions, and so normal probability plots were made using the data of the six distributions in Fig. 5.5 and Fig. 5.6. The normal probability plot for the test chip distribution of Fig. 5.5 is shown in Fig. 5.7. A normal probability plot of a purely Gaussian distribution would plot all data points linearly, with any deviation from that linearity indicating a deviation from a true Gaussian distribution. Fig. 5.7 shows that all data from the test chip's distribution within at least 2.5σ of its mean falls within 3% of true linearity, i.e. 95% of the data from this distribution is approximately Gaussian. Generally, the normal probability plots of all six distributions have all of their data within at least 2.5σ of their mean fall within 3% of true linearity (see Appendix A). This is enough statistical confidence to assume that the Monte Carlo distributions are approximately Gaussian and that the results of the Welch t tests are valid.

The Welch t tests were performed using a Graphpad Software calculator. The results are shown in Table 5.1 and Table 5.2. A Welch t test produces a P value between 0 and 1. For example, the P value of the ideal and test chip distributions for a V_{DS} value of 1.5V, can be used to determine whether the means of the ideal and test chip distributions differ, or that they are the same, and the 500 iteration Monte Carlo simulation recorded two slightly different means as a result of random sampling. Under the assumption that the two distributions are equal, P gives the probability of the differences between their means being equal to or greater than their recorded difference (in the case of the ideal and test chip distributions for a V_{DS} value of 1.5V, the difference is only 0.5μA). From Table 5.1, the low P value between the ideal and MUX distributions shows that, if their means are truly identical, the probability that their means differ by at least 1.3μA is only 6.25%. This probability is quite small, and so it is likely that there *is* a difference between the ideal and MUX Monte Carlo data. By contrast, the P value between the ideal and test chip distributions is quite large. There is a 48% chance that, if the two distributions were the same, random sampling *could* produce a mean difference of at least 0.5μA. A

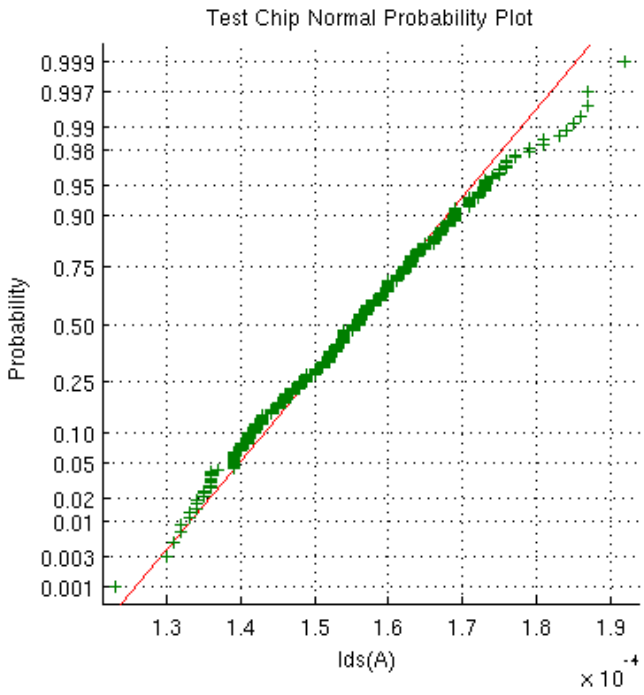

Figure 5.7: Normal Probability Plot of a Monte Carlo simulation for the current of a DUT in the test chip structure's array(V_{GS} = 1.5V, V_{DS} = 1.5V, μ = 155.6μA, σ = 10.7μA, number of runs N = 500).

large P value simply means that it is unlikely that the two distributions differ. Therefore, Table 5.1 shows that, statistically, it is more likely that the MUX distribution differs from the ideal distribution, and less likely that the test chip distributions differs from the ideal distribution. While this does not definitively prove that the test chip distribution matches closer to the ideal distribution than the MUX distribution does, it shows that this is far more likely true than the opposite case.

Table 5.1: Welch t test results for Monte Carlo Distributions(V_{GS} = 1.5V, V_{DS} = 1.5V)

Monte Carlo distributions	P
Ideal, MUX	0.0625
Ideal, Test Chip	0.4809

Table 5.2 shows that the probability of the test chip distribution matching closer to the ideal distribution than the MUX distribution does is even higher for the case of a V_{DS} value of 0.8V, as the difference between the two P values has grown even larger. The P value of the Ideal and MUX distributions is almost insignificant due to the large shift of the MUX structure's mean current caused by an increased voltage drop in its multiplexer at these operating conditions.

Table 5.2: Welch t test results for Monte Carlo Distributions(V_{GS} = 1.5V, V_{DS} = 0.8V)

Monte Carlo distributions	P
Ideal, MUX	0.0001
Ideal, Test Chip	0.582

5.3 Stress Measurements

To simulate the effects of TSV stress on mobility variation, Spectre simulations were made using modified BSIM4 transistor models, which varied the mobility of specific NMOS DUTs by 0% to +10% in a symmetric, radial pattern around a specific point, to emulate the mobility variation patterns induced by a single TSV on neighbouring NMOS devices, as described in Yang *et al* [25]. These simulations were run as Monte Carlo simulations to simulate the effects of stress and process variations on a limited number of DUTs. The spatial profile of the simulated DC current of these devices is shown in Fig. 5.8. This figure illustrates the effectiveness of the radial filtering technique described in section 4.2. By taking advantage of the symmetry of the original mobility pattern, the currents of devices equidistant to the simulated TSV can be averaged to approximate the current of these devices without the influence of process variations. As a result, the peak current extracted using radial filtering has an error of only 0.19% from the mobility-only simulation. A modification of the Gaussian filtering scheme proposed in Chang *et al* [2] is also somewhat effective at reconstructing the spatial profile of the DC currents. This filtering technique averages neighbouring devices' current together, applying a weighted average to nearby

devices to avoid skewing the recovered data. The extracted peak current of the Gaussian filtering technique has only a 3.1% error from the mobility-only variation.

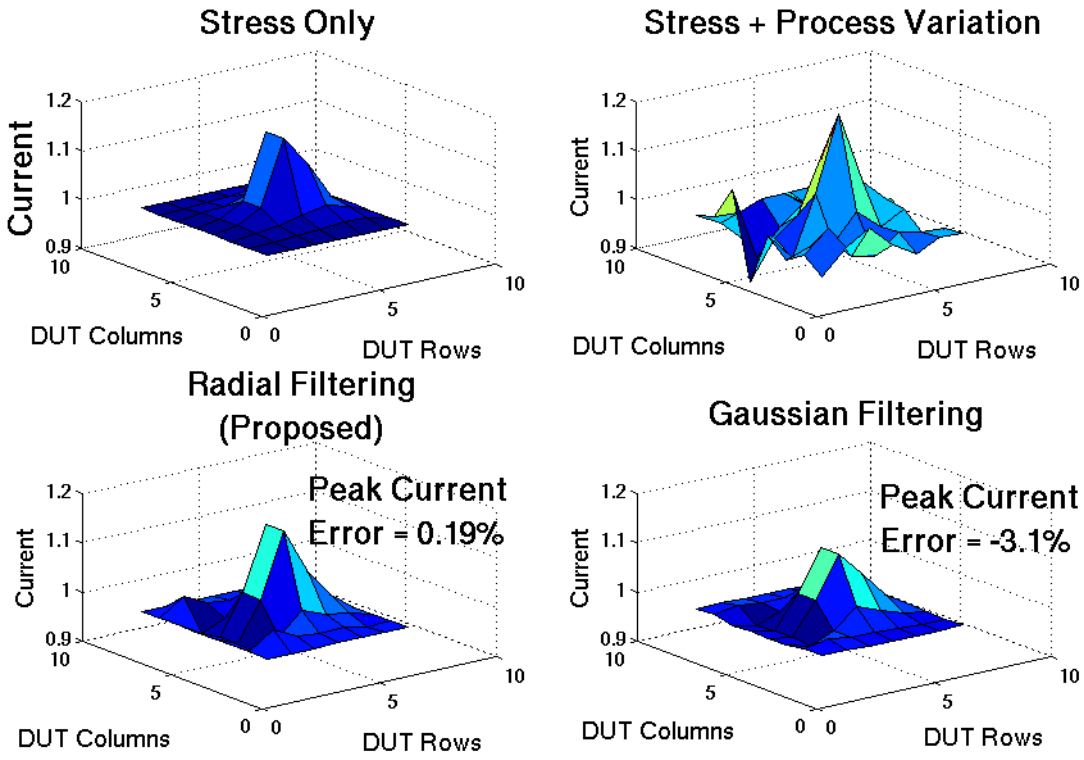

Figure 5.8: Recovered stress-induced current profiles using radial filtering and Gaussian filtering [2].

Chapter 6

Measurement Setup

6.1 Test Board

A printed circuit board (PCB) was designed to house the test chip and provide inputs and outputs to and from the chip to equipment used to record measurements from it. An image of the PCB is shown in Fig. 6.1.

The chip will be manufactured in a DIP-40 package, and placed in the DIP-40 slot on the PCB. Power (3.3V and 1.5V) and ground are provided off-chip. Off-chip digital signals(CLK, Enable, etc.) and analog signals (Vgs, eight Vds signals for the eight test structures within the chip) used to control the chip are provided by a FPGA. A circuit consisting of a manual switch and two resistors connects the FPGA's voltage signals to the output pin of each test structure to allow for the measurement of the drain-to-source current (I_{DS}) from the DUTs for different values of the DUTs' drain-to-source voltage (V_{DS}). The measured signals from the chip are routed to VNC connectors so that they may be monitored by oscilloscopes.

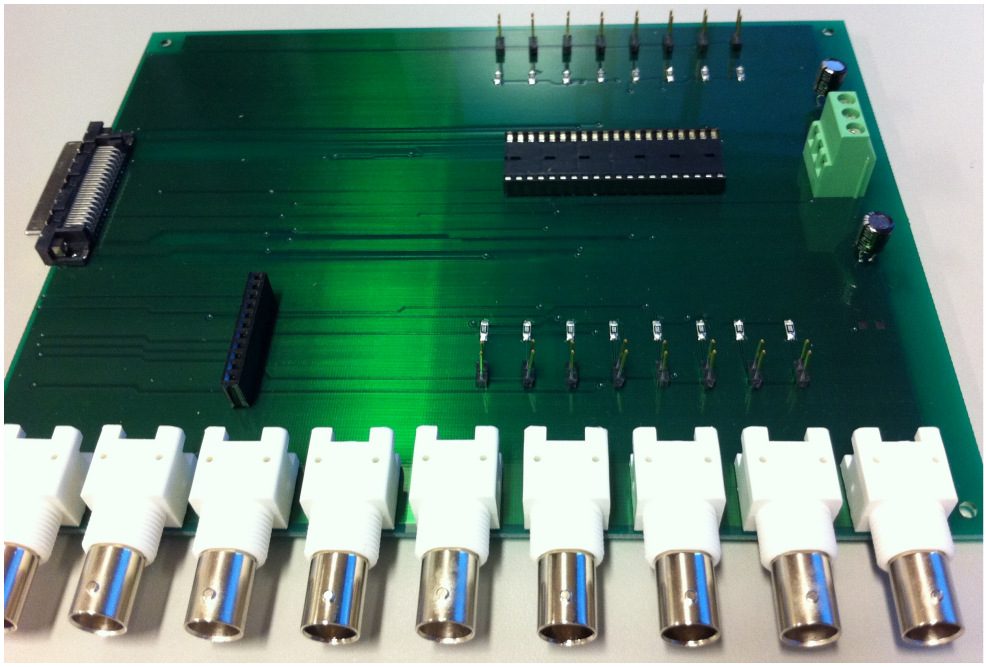

Figure 6.1: Image of the PCB designed for the test chip.

6.2 Test Chip Operation

To take measurements from each DUT in the test chip, a FPGA controlled off-chip voltage is applied to the eight I/O pads connected to the DUT arrays of the eight test structures. Next, the row select and column select scanchains of each test structure are manipulated simultaneously by the shared signals CLK, Set, Reset, and Enable provided by the FPGA to select the gate select cell of the first row and column of each test structure's selection array. Fig. 6.2 shows one DUT of one array, already selected by the digital logic of the test chip. It also shows its series connection to a resistor located on the PCB, then connected to the analog voltage provided by the FPGA. To measure I_{DS}, the analog voltage ($V_{applied}$) is set to 1.5V. The DUT will turn on and provide a current (I_{DS}) that will cause a voltage drop (ΔV) across the resistor. The voltage at both terminals of the resistor is measured by the FPGA to quantify the voltage drop, and I_{DS} can be measured as $I_{DS} = \frac{\Delta V}{R}$. This value of I_{DS} is for a specific value of V_{DS}, since the small voltage drop determines the voltage

seen at the drain of the DUT.

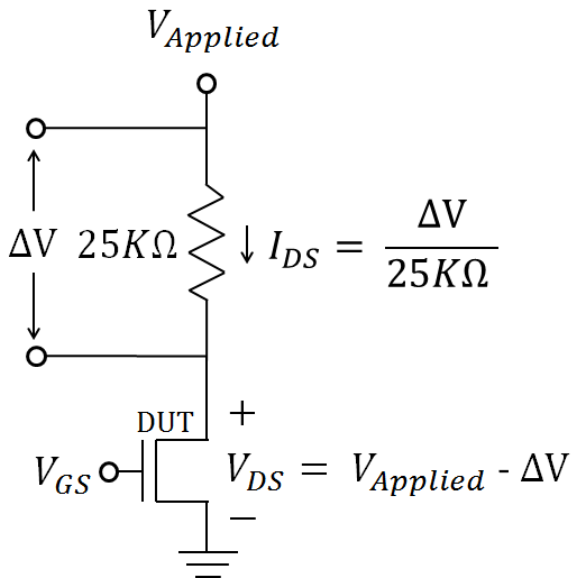

Figure 6.2: Test chip measurement setup.

The FPGA converts its voltage measurements using a 16-bit resolution ADC that has a range of -10V to 10V. The largest current that is expected to be measured from a single device in one of the test chip's DUT structures is 180μA, so the resistor value was chosen to ensure that the largest ΔV possible is created across the resistor without exceeding 5V, half of the ADC's range (ΔV should be maximized since larger voltages can be measured more accurately by the ADC then smaller ones). This resistor value was calculated to be $\frac{5V}{180\mu A}$ $\sim$ 25kΩ. By maximizing ΔV for half the ADC's range, the PCB allows a safety margin for the FPGA to measure unexpectedly large currents, which can be as twice as large as the value of the expected maximum current without exceeding the ADC's range Once the measurements of the eight DUTs are complete, the FPGA manipulates the scanchains to select a different DUT cell position among the DUT arrays and the current-voltage characteristics can be measured again. This process repeats itself until all 7200 DUTs have been measured. Both V_{GS} and $V_{applied}$ are adjustable for a range of -10V to 10V, so this process can be repeated to take current measurements from the DUTs for many values

of V_{GS} and V_{DS} to extract their current-voltage characteristics.

6.3 Measured Parameters

Once measured, each DUT provides several measurements of I_{DS} for different values of V_{GS} and V_{DS}. Two key parameters can be extracted from this data to characterize process variations: V_{th} and the current factor of the device ($\beta = \mu C_{ox}\frac{W}{L}$, where μ is the carrier mobility of the device, C_{ox} is the oxide capacitance per unit area, and W and L are the channel width and length of the device, respectively). V_{th} is affected by variations in t_{ox} and implant impurity levels [18]. β is affected by fluctuations in W, L, V_{th}, and C_{ox} (which is dependent on t_{ox}). β is also affected by changes in carrier mobility ($\frac{\Delta\mu}{\mu(\theta)}$), which can change as a result of both process variation and TSV-induced stress. Both are highly dependent on the variations of a transistor's device parameters, and can be extracted directly from a transistor's current-voltage characteristic.

There are many methods for extracting these parameters. V_{th}, for example, can be extracted by extrapolating the saturation current of each device by plotting $\sqrt{I_{DS}}$ vs. V_{GS} and extrapolating the curve to zero drain current, as shown in Fig. 6.3 [22]. There are simpler techniques available to determine V_{th}, but most of them require subthreshold current measurements. Section 5.1 has already described the increased error that the test chip's subthreshold measurements should incur as a result of background leakage in the chip's test structures, so a parameter extraction technique that uses saturation current measurements is preferred.

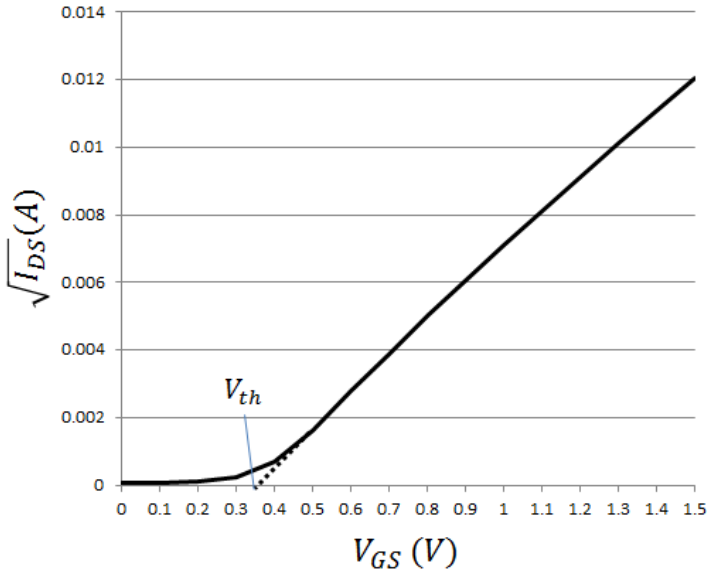

Figure 6.3: The extrapolation of V_{th} from an $\sqrt{I_{DS}}$ vs. V_{GS} plot of a single 1.2μm/1.2μm DUT from the 3D test chip. V_{DS} = 1.5V, V_{th} = 0.354V. [22]

Chapter 7

Conclusions

A test chip was designed for the purposes of characterizing process variation in 3D ICs and device performance variation caused by TSVs. The test chip design was laid out, and it will be fabricated in a 130nm Tezzaron/GlobalFoundries CMOS process. Chapter 3 detailed the design of the test chip, comparing it to previous process variation test structure designs. It was shown that the test chip hosts eight test structures featuring large, single I/O DUT arrays consisting of NMOS and PMOS transistors. The test chip was designed with minimal I/O while having a high spatial resolution and measurement fidelity. The physical parameters of the test chip were compared to those of similar designs. Chapter 4 introduced measurement techniques for the improvement of the accuracy of the measurements taken from the test chip.

Chapter 5 detailed the simulation results illustrating the accuracy of the test chip's measurements. Monte Carlo simulations of the extracted parasitic RC netlist of the test chip show that it can accurately measure Monte Carlo variations with less than 1% error, and Welch t tests showed that the Monte Carlo distribution of the test chip's DC current measurements likely align closely with the ideal DC current distribution. Simulated TSV stress-induced mobility variation measurements could also be extracted with less than 0.2% accuracy. Additionally, error caused by parasitic resistance in the DUT array was simulated to be less than 0.25%. Chapter 6 proposed a measurement setup for the test chip and a

technique for extracting threshold voltage from the measured current.

7.1 Improvements

The simulations show that the measurement of subthreshold current for $V_{GS} < 0.1$V has very high error, even after background subtraction has been employed to reduce measurement error. This could be fixed in future designs. Assuming the designs were more lenient and offered more design area and therefore more I/O pads, more accurate measurements techniques, such as four-probe kelvin measurements, could be employed to reduce measurement error. At a cost to the test structure's spatial density, additional circuitry can be introduced to its design to reduce background leakage current [20].

If more I/O were made available, changes to the layout could be made to improve the statistical confidence of the measured data. With more I/O available, more test structures of smaller sizes could be placed on the two tiers of each array to provide a better spatial measurement of process variations across the area of both dies. Other small layout changes could improve the test structures as well. For example, rows and columns of non-connected DUTs could be added to the periphery of each gate to ensure that all measured devices along the edges of the test structures have identical lithographic neighbourhoods.

7.2 Contributions

A 3D test chip architecture was introduced in this thesis. Unlike similar architectures, the structures of the test chip take advantage of the 3D nature of the IC to separate the selection logic and the measurement devices onto the two tiers of the chip, significantly increasing the density of the devices compared to similar, 2D designs. A PCB design and test setup were also proposed created and proposed for the purposes of taking automated measurements from the chip using a FPGA. Measurement techniques were proposed to reduce the error of current measurements taken from the test chip.

7.3 Future Work

Upon its return from fabrication in October 2013, the test chip can be used to verify the measurement accuracy results of chapters 5 and 6. Measurements taken from the chip will indicate if a significant process variation mismatch exists between two tiers contained within a 3D IC, as well as how the measurement accuracy of the test chip compares to the accuracy of other process variation characterization designs. Measurements will also verify if the TSVs employed in Tezzaron/GlobalFoundries 3D ICs contribute significant variance to the performance of nearby devices.

Further work can be made to directly compare the 3D test chip's design to other 2D designs. Investigations can be made exploring the limits of the scalability and density of each design.

Other test chip architectures can be made such that they take advantage of 3D ICs in the way the proposed test structure's design does. A test structure that uses SRAM cells instead of single NMOS and PMOS transistors for its DUT array can further classify process variation mismatch in 3D IC in memory applications. The test structure design has usefulness in other applications as well. For example, light sensitive diodes placed in the DUT array could be used to design a very dense CMOS sensor array.

APPENDICES

Appendix A

Normal Probability plots of Monte Carlo Simulations

A.1 Single device Normal Probability plots of Monte Carlo Simulations

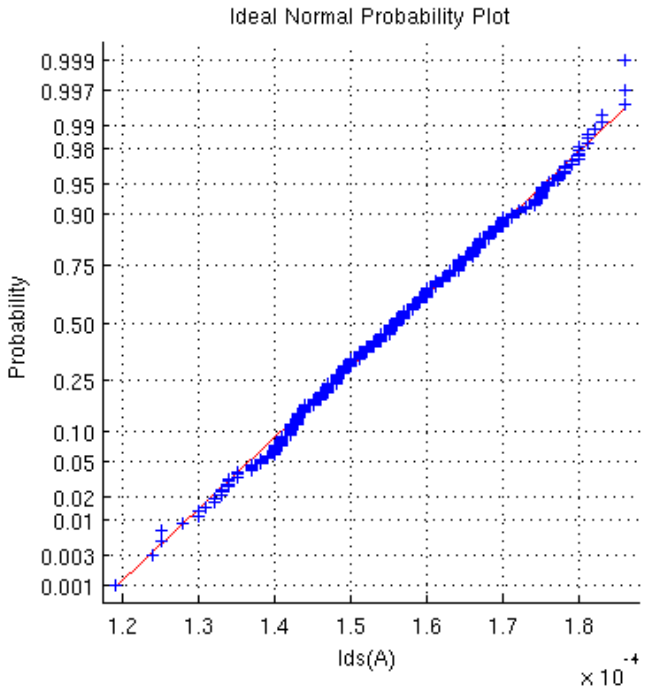

Figure A.1: Normal Probability Plot of a Monte Carlo simulation for the current of an ideal device(V_{GS} = 1.5V, V_{DS} = 1.5V, μ = 156.1μA, σ = 11.7μA, number of runs N = 500).

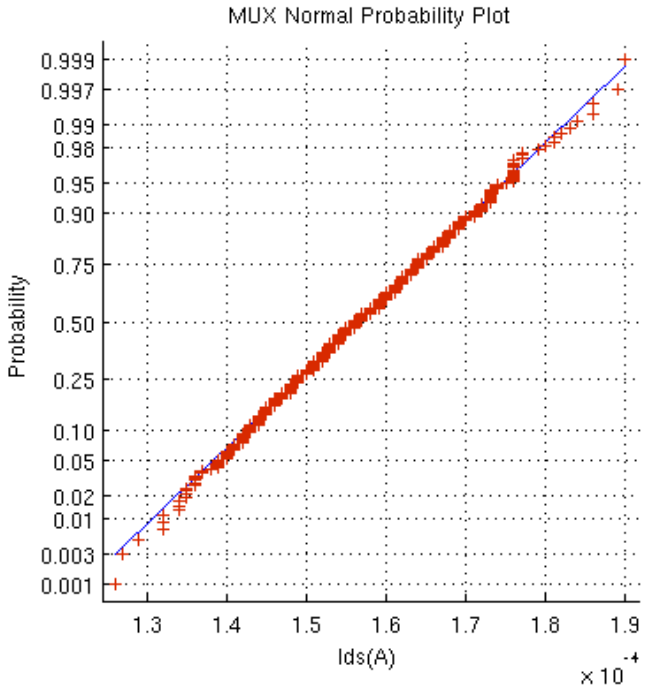

Figure A.2: Normal Probability Plot of a Monte Carlo simulation for the current of a DUT in the simulated MUX structure(V_{GS} = 1.5V, V_{DS} = 1.5V, μ = 154.8μA, σ = 10.3μA, number of runs N = 500).

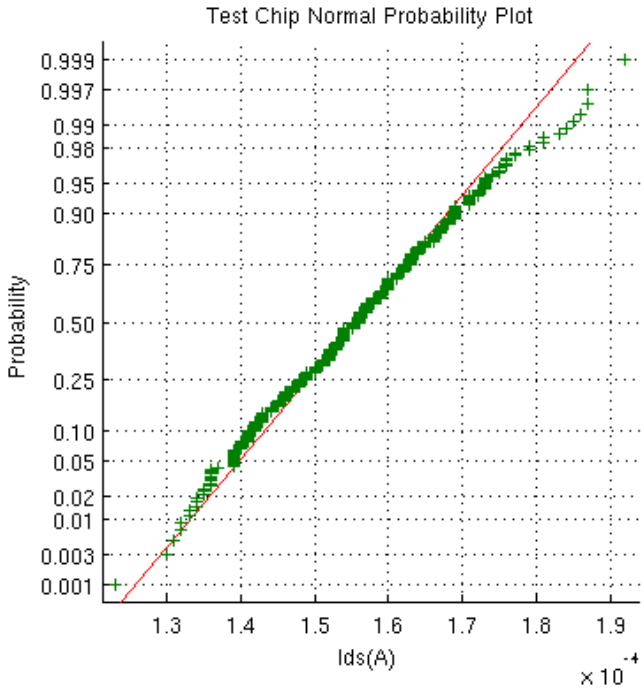

Figure A.3: Normal Probability Plot of a Monte Carlo simulation for the current of a DUT in the test chip structure's array(V_{GS} = 1.5V, V_{DS} = 1.5V, μ = 155.6μA, σ = 10.7μA, number of runs N = 500).

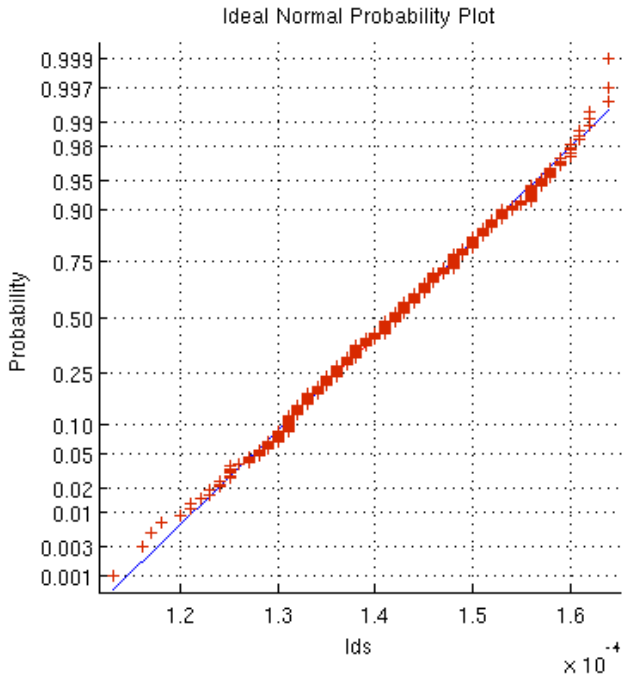

Figure A.4: Normal Probability Plot of a Monte Carlo simulation for the current of an ideal device(V_{GS} = 1.5V, V_{DS} = 0.8V, μ = 141.9μA, σ = 8.98μA, number of runs N = 500).

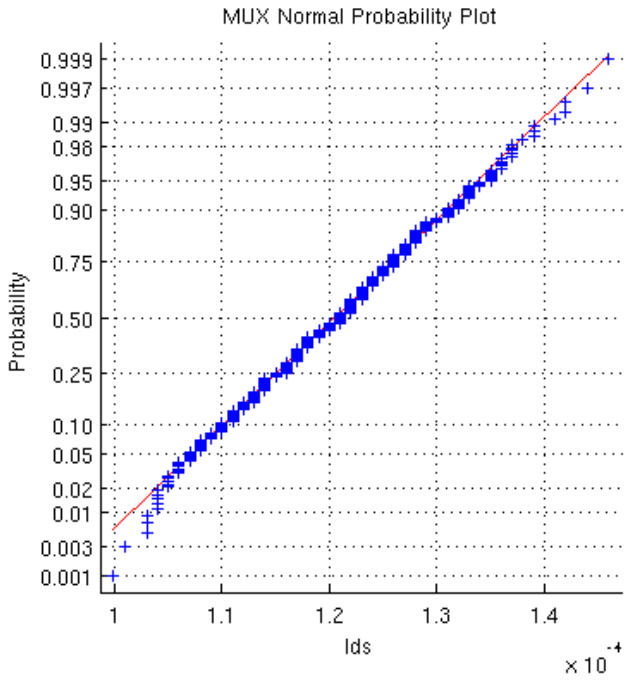

Figure A.5: Normal Probability Plot of a Monte Carlo simulation for the current of a DUT in the simulated MUX structure(V_{GS} = 0.8V, V_{DS} = 1.5V, μ = 120.7μA, σ = 8.03μA, number of runs N = 500).

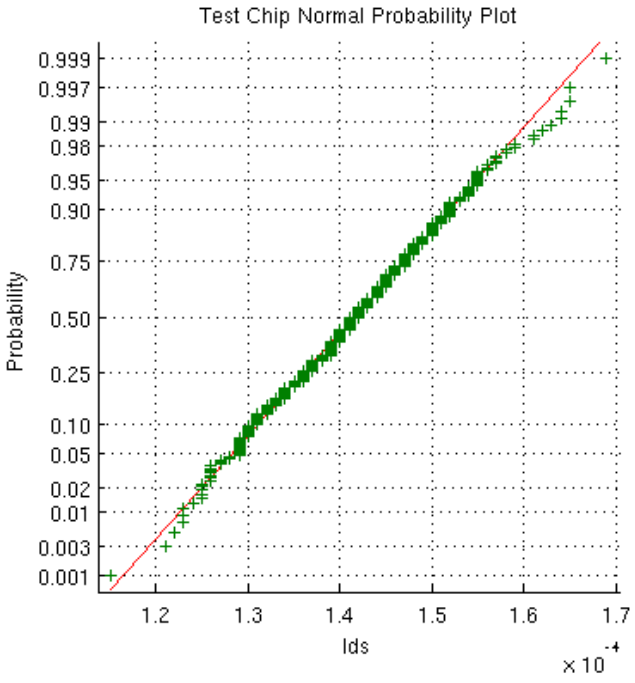

Figure A.6: Normal Probability Plot of a Monte Carlo simulation for the current of a DUT in the test chip structure's array(V_{GS} = 0.8V, V_{DS} = 1.5V, μ = 141.6μA, σ = 8.23μA, number of runs N = 500).

References

[1] P. Apte, B. Doering, and P. Gargini. The international technology roadmap for semiconductors, 2007.

[2] A.H. Chang, K. Zuo, J. Wang, D. Yu, and D. Boning. Test structure, circuits and extraction methods to determine the radius of infuence of STI and polysilicon pattern density. In *13th International Symposium on Quality Electronic Design*, pages 185–192, 2012.

[3] W.R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A.M. Sule, M. Steer, and P.D. Franzon. Demystifying 3D ICs: the pros and cons of going vertical. *IEEE Design & Test of Computers*, 22(6):498–510, 2005.

[4] M. Dietrich and J. Haase, editors. *Process Variations and Probabilistic Integrated Circuit Design*. Springer Science+Business Media, 233 Sring St., New York, NY, 10013, USA, 2012.

[5] N. Drego, A. Chandrakasan, and D. Boning. A test-structure to efficiently study threshold-voltage variation in large MOSFET arrays. In *Proc. International Symposium on Quality Electronic Design*, pages 281–286, 2007.

[6] S. Garg and D. Marculescu. 3D-GCP: An analytical model for the impact of process variations on the critical path delay distribution of3D ICs. In *International Symposium on Quality Electronic Design, 2009*, pages 147–155, 2009.

[7] S. Han and D.D. Wentzloff. Wireless power transfer using resonant inductive coupling for 3D integrated ICs. In *IEEE International 3D Integration Conference*, pages 1–5, 2010.

[8] C. Hess, S. Saxena, H. Karbasi, S. Subramanian, M. Quarantelli, A. Rossoni, S. Tonello, S. Zhao, and D. Slisher. Device array scribe characterization vehicle test chip for ultra fast product wafer variability monitoring. In *Proc. IEEE International Conference on Microelectronic Test Structures*, pages 145–149, 2007.

[9] J.W. Joyner, P. Zarkesh-Ha, and J.D. Meindl. A stochastic global net-length distribution for a three-dimensional system-on-a-chip (3d-soc). In *14th annual IEEE International ASIC/SOC Conference*, pages 147–151, 2001.

[10] A. P. Karmarkar. Performanace and reliability analysis of 3d-integration structures employing Through Silicon Via (TSV). In *Reliability Physics Symposium, IEEE International*, pages 682–687, 2009.

[11] N.H. Khan, S.M. Alam, and S. Hassoun. Through-Silicon Via (TSV)-induced noise characterization and noise mitigation using coaxial TSVs. In *IEEE International Conference on 3D System Integration*, pages 1–7, 2009.

[12] D. Kim, J Kim, J Pak, H. Lee, J. Lee, K. Park, and j. Kim. Vertical tree 3-dimensional tsv clock distribution network in 3D IC. In *IEEE 62nd Electronics Components and Technology Conference*, pages 1945–1950, 2012.

[13] A. Mercha, A. Redolfi, M. Stucchi, N. Minas, J. Van Olmen, S. Thangaraju, D. Velinis, Domae S., Y. Yang, G. Katti, R. Labie, C. Okoro, M. Zhao, P. Asimakopoulos, I. De Wolf, T. Chiarella, T. Schram, E. Rohr, A. Van Ammel, A. Jourdain, W. Ruythooren, S. Armini, A. Radisic, H. Philipsen, N. Heylen, M. Kostermans, P. Jaenen, E. Sleeckx, D. Sabuncuoglu Tezcan, I. Debusschere, P. Soussan, D. Perry, G. Van Der Plas, J.H. Cho, P. Marchal, Y. Travaly, E. Beye, S. Biesemans, and B. Swinnen. Impact of thinning and through silicon via proximity on high-k / metal gate first CMOS performance. In *Symosium on VLSI Technology*, pages 109–110, 2010.

[14] S. Mick, J. Wilson, and P. Franzon. 4 Gbps high-density AC coupled interconnection. In *Proceedings of the IEEE Custom Integrated Circuits Conference*, pages 133–140, 2002.

[15] Robert S. Patti. Three-dimensional integrated circuits and the future of system-on-chip designs. *Proceedings of the IEEE*, 94(6):1214–1224, 2006.

[16] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers. Matching properties of MOS transistors. *IEEE Journal of Solid-State Circuits*, 24(5):1433–1440, 1989.

[17] D. Perry, J. Cho, S. Domae, P. Asimakopoulos, A. Yakovlev, P. Marchal, G. Can der Plas, and N. Minas. An efficient array structure to characterize the impact of through silicon vias on FET devices. In *IEEE International Conference on Microelectronic Test Structures*, pages 118–122, 2011.

[18] J.M. Rabaey, A. Chandrakasan, and B. Nikolic. *Digital Integrated Circuits: A Design Perspective*. Pearson Education, Inc., Upper Saddle River, NJ, 07458, USA, 2nd edition, 2003.

[19] R. Rao, A. Srivastava, D. Baauw, and D. Sylvester. Statistical estimation of leakage current considering inter- and intra-die process variation. In *Proceedings of the 2003 International Symposium on Low Power Electronics and Design*, pages 84–89, 2003.

[20] T. Sato, N. Ueyama, N. Nakayama, and K. Masu. Accurate array-based measurement for subthreshold-current of MOS transistors. *IEEE Journal of Solid-State Circuits*, 44(11):2977–2986, 2009.

[21] S. Saxena, S. Minehane, J. Cheng, M. Sengupta, C. Hess, M. Quarantelli, G.M. Kramer, and M. Redford. Test structures and analysis techniques for estimation of the impact of layout on MOSFET performance and variability. In *Proc. IEEE International Conference on Microelectronic Test Structures*, pages 263–266, 2004.

[22] D.K. Schroeder. *Semiconductor Material and Device Characterization*. Wiley Interscience, Piscataway, NJ, 3rd edition, 2006.

[23] T. Thorolfsson, S. Melamed, G. Charles, and P.D. Franzon. Comparative analysis of two 3D integration implementations of a SAR processor. In *IEEE International Conference on 3D System Integration*, pages 1–4, 2009.

[24] S. Uemura, Y. Hiraoka, T. Kai, and S. Dosho. Isolation techniques against substrate noise coupling utilizing through silicon via (TSV) process for rf/mixed-signal socs. *IEEE Journal of Solid-State Circuits*, 47(4):810–816, 2006.

[25] J.S. Yang, K. Athikulwongse, Y.J. Lee, S. K. Lim, and D. Z. Pan. TSV stress aware timing analysis with applications to (3D-IC) layout optimization. In *Proc. Design Automation Conference*, pages 803–806, 2010.

[26] L. Yu. A study of through-silicon-via (TSV) induced transistor variation. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2011.