Measurement and Analysis of Variability in CMOS circuits

Liang Teck Pang

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2008-108
http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-108.html

August 29, 2008

Acknowledgement

I wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center, the National Science Foundation Infrastructure Grant
No. 0403427, wafer fabrication donation of STMicroelectronics, and the support of the Center for Circuit & System Solutions (C2S2) Focus Center, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program.

Measurement and Analysis of Variability in CMOS circuits

by

Liang Teck Pang

Diplôme D'Ingénieur (Ecole Centrale de Paris, France) 1997
Master of Philosophy (University of Cambridge, United Kingdom) 1997

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Borivoje Nikolić, Chair
Professor Andrew Neureuther
Professor Robert Leachman

Fall 2008

The dissertation of Liang Teck Pang is approved:

Chair Date

Date

Date

University of California, Berkeley

Fall 2008

Measurement and Analysis of Variability in CMOS circuits

Liang Teck Pang

Abstract

Measurement and Analysis of Variability in CMOS circuits

by

Liang Teck Pang

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Borivoje Nikolić, Chair

The scaling of CMOS technology into the deep sub-micron regime has resulted in increased impact of process variability on circuits, to the point where it is considered a major bottleneck to further scaling. In order to continue scaling, there is a need to reduce margins in the design by classifying process variations as systematic or random. In this work, a methodology to characterize variability through measurement and analysis has been developed. Systematic and random, die-to-die (D2D) and within-die (WID) components of variability are quantified and corresponding sources of variability are identified.

This methodology was developed for an early 90nm CMOS process and further refined for an early 45nm CMOS process. Test-chips have been designed to study the effects of layout, and characterize variability of delay and leakage current using an array of test-structures. Delay is obtained through the measurement of ring oscillator frequencies, and transistor leakage current is measured by an on-chip analog-to-digital converter (ADC).

In 90nm, it has been found that transistor performance depends strongly on polysilicon (poly-Si) gate density and that spatial correlation depends on gate orientation and the direction of gate spacing. WID variation is small with three standard deviations over mean $(3\sigma/\mu) \approx 3.5\%$, whereas D2D and systematic layout-induced variations are significant, with $3\sigma/\mu$ D2D variation of $\approx 15\%$ and a maximum layout-induced frequency shift of 10%.

In 45nm, a process which features immersion lithography, strained-Si and more restrictive design rules for gate spacing, it has been found that systematic layout-induced variability has decreased. However, new sources of variability due to the dependence of stress on layout were found. WID has increased to $3\sigma/\mu \approx 6.6\%$ and can be attributed to a smaller transistor area whereas D2D variation has remained at $3\sigma/\mu \approx 15\%$.

This methodology is effective in characterizing variability. It improves the accuracy of statistical models and allows process corners to be set up for WID or D2D variations. In addition, sources of systematic variations are identified and the impact of layout design rules are measured. As scaling continues and variation increases, characterization of variability will become an integral part of the IC design process.

Professor Borivoje Nikolić
Dissertation Committee Chair

To my wife and our parents

Contents

List of Figures **v**

List of Tables **ix**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Research Goal . . . 3
1.3 Related Work on Process Characterization . . . 5
1.3.1 Metrology Techniques . . . 5
1.3.2 Measurement of transistor performance . . . 6
1.3.3 Other circuit techniques for process characterization . . . 8
1.4 Dissertation Organization . . . 8
1.5 Acknowledgements . . . 9

2 Characterization of Variations **11**
2.1 Characteristics of Variations . . . 12
2.2 Impact of Variability on Circuits . . . 13
2.3 Sources of Process Variations . . . 15
2.3.1 Interface roughness . . . 16
2.3.2 Line-edge roughness . . . 16
2.3.3 Random Dopant Fluctuation . . . 17
2.3.4 Subthreshold Voltage Variation . . . 17
2.3.5 Strain . . . 18
2.4 Summary . . . 20

3 Process-induced Systematic Variations **21**
3.1 Variations in the Exposure System . . . 22
3.2 Variations in Patterning and Other Manufacturing Steps . . . 26
3.3 Layout Dependent Variations . . . 27
3.4 Summary . . . 29

4 Measuring CMOS Performance Variability **30**
4.1 Leakage Current . . . 31
4.1.1 Leakage Transistor Array . . . 31
4.1.2 Leakage Current Measurement . . . 32
4.2 Frequency . . . 47
4.2.1 Ring Oscillator Array . . . 47
4.2.2 Ring Oscillators with Different Number of Stages . . . 50
4.3 Summary . . . 51

5 Measuring SRAM Performance Variability **53**
5.1 SRAM Static Noise Margins Measurement . . . 54
5.2 SRAM Dynamic Write Performance Measurement . . . 56
5.3 SRAM Dynamic Read Performance Measurement . . . 57
5.4 SRAM Access Times Measurement . . . 61
5.5 Summary . . . 62

6 Measurement Results and Analysis **63**
6.1 Layout Structures and Test-Chips Overview . . . 64
6.2 Effects of Layout on Frequency and Leakage . . . 75
6.2.1 Proximity of Poly-Si Gates . . . 76
6.2.2 Gate Rotation . . . 81
6.2.3 Symmetry . . . 82
6.2.4 Metal Coverage Over Gate . . . 83
6.2.5 Fixed Poly-Si Gate Pitch . . . 86
6.2.6 Shallow Trench Isolation (STI) . . . 86
6.2.7 Source/Drain Diffusion area . . . 87
6.3 Characteristics of D2D and WID Variations . . . 89
6.3.1 Statistics of D2D and WID Variations . . . 89
6.3.2 WID Spatial Correlation of ROs . . . 91
6.3.3 Spatial Correlation of Gates in a RO . . . 95
6.4 Inferring Process Parameters (90nm) . . . 96
6.5 Summary . . . 101

7 Conclusion **103**
7.1 Key Contributions . . . 103
7.2 Methods to Mitigate Variability . . . 105
7.3 Future Work . . . 106
7.4 Conclusion . . . 107

Bibliography **108**

A Mathematics **117**
A.1 Integrator equation derivations. . . . 117
A.2 Least squares estimate, error estimate. . . . 119

B Layouts **121**
B.1 Arrangements of the Layouts in a Tile. 121
B.2 Library cells in 90nm. 121
B.2.1 90nm Library Setup. 121
B.2.2 90nm Library. 121
B.3 Library cells in 45nm. 124
B.3.1 45nm Library Setup. 124
B.3.2 45nm Library. 124

C Test Setup **126**
C.1 GPIB-ENET/100 Guide 126
C.2 DG2020A Pattern Generator 130
C.2.1 Configuring the DG2020A 130
C.2.2 Programming the DG2020A 131
C.3 Measurement Setup and Procedures 133

List of Figures

1.1 Evolution of the wavelength of light used in photolithography over the years. Presently, in the 90nm and 45nm nodes, 193nm wavelength is used. Rounding of printed features is shown on the right. [1] 3
1.2 a) Present design methodology. b) Categorizing variations. 4

2.1 Correlation of path delays 14
2.2 Correlation of gate delays in a critical path 14

3.1 Step and scan photolithography using a slit of light. 22
3.2 Formation of poly-Si gates. 24
3.3 Isolated and dense lines. 24
3.4 Lens imperfections. 24
3.5 Effect of flare. 25

4.1 Standby leakage current measurement. 32
4.2 45nm NMOS leakage current measurement. I_{nb} refers to drain to bulk junction leakage current and I_{ng} refers to drain to gate leakage current. Taking the difference between I_{s1} and I_{s0} gives I_s, the subthreshold leakage current of the selected NMOS transistor. Both NMOS and PMOS leakage currents are measured in the 45nm test-chip and only NMOS leakage current is measured in the 90nm test-chip. 33
4.3 Single slope ADC for current measurement (90nm test-chip). 34
4.4 Timing diagrams of the integrator control signals and the output signals. . 35
4.5 Current measurement procedure. 35
4.6 Schematics of the folded cascode amplifier for measuring NMOS leakage current. Gain is optimized for input bias of VDD=1V. Transistor widths are indicated in μm next to the individual transistors. All gate lengths are $0.35\mu m$ and $In = 90\mu A$. All transistors whose bulk connections are not indicated have it connected to VDDA(PMOS) or GNDA(NMOS). 41

4.7 Schematics of the folded cascode amplifier for measuring PMOS leakage current. Gain is optimized for input bias of GND=0V. Transistor widths are indicated in μm next to the individual transistors. All gate lengths are $0.35\mu m$ and $Ip = 120\mu A$. All transistors whose bulk connections are not indicated have it connected to VDDA(PMOS) or GNDA(NMOS). 41
4.8 BODE plots of the amplifier used for NMOS I_{LEAK} measurement in the 90nm test-chip. Top plot shows a DC gain of $\approx$ 60dB and bottom plot shows a phase margin of $\approx 80°$ 42
4.9 Schematics of the comparator for generating the start and stop signals when the output of the integrator crosses the threshold voltages. 43
4.10 Simulation results for the current measurement circuit. Plots of leakage current measurement error vs leakage current. a) shows post-layout simulation results and b) shows pre-layout simulation results. 44
4.11 Repeated measurement of I_{LEAK} to demonstrate accuracy in the 90nm process. Plots 160 measurements of NMOS I_{LEAK} of the same 3 layouts in the same tile taken over 27 hours. A σ/μ of $\leq 1.3\%$ is measured. 44
4.12 BODE plots of the amplifier for NMOS I_{LEAK} measurement in the 45nm test-chip. Top plot shows a DC gain of $\approx$ 60dB and bottom plot shows a phase margin of $\approx 75°$ 45
4.13 BODE plots of the amplifier for PMOS I_{LEAK} measurement in the 45nm test-chip. Top plot shows a DC gain of $\approx$ 60dB and bottom plot shows a phase margin of $\approx 73°$ 46
4.14 PMOS leakage current measurement for the 45nm test-chip. 46
4.15 Repeated measurement of I_{LEAK} to demonstrate accuracy in the 45nm process. a) and b) plot 200 measurements of PMOS and NMOS I_{LEAK} of the same layout in the same tile. σ/μ of 1.4% and 11.5% are measured for PMOS and NMOS I_{LEAK} respectively. 47
4.16 RO with n stages. Address bits will enable the RO by applying VDD to the *enable* signal. The output of the RO (*out*) goes to the input of the multiplexer. 48
4.17 RO frequency measurement. 49
4.18 Repeated measurement of RO frequency to demonstrate accuracy in the 90nm process. Plots 236 measurements of RO frequency of 3 layouts in the same tile taken over 18 hours. A σ/μ of 0.15% is measured. 49
4.19 Repeated measurement of RO frequency to demonstrate accuracy in the 45nm process. Plots 1000 measurements of RO frequency of the same layouts in the same tile. A σ/μ of $\leq 0.02\%$ is measured. 50
4.20 RO with programmable number of stages using NAND gates 51

5.1 A SRAM column showing the bitline switches. 55
5.2 Bitline IV plot showing write margin indicators 55
5.3 Bitline IV plot showing read margin indicators 56
5.4 RO to measure write access time and the differential buffer used in the RO. 58
5.5 Pulse RO for characterizing cell read timing. 60
5.6 Current starved buffers for widening the pulse width and creating delay for the regenerated pulse. 60

5.7 SRAM access time measurement 61

6.1 Die photo of the 90nm test-chip 64
6.2 12 layout structures in 90nm 65
6.3 Model of spatial correlation coefficient. Plot of correlation coefficient ρ vs spacing i. k is a model parameter. 66
6.4 Arrangement of ROs of different lengths. 67
6.5 Layout of a bigtile and the 10×4 array of bigtile in the 90nm test-chip. . . 68
6.6 17 layout configurations in 45nm 72
6.7 Illustrates (a)horizontally vs (b)vertically placed inverters in a RO. 73
6.8 Die photo of the 45nm test-chip 73
6.9 Frequency distribution for single-finger configurations in the 90nm test-chip. Vertical lines correspond to typical and fast corner simulation results. Bar plots correspond to the WID distribution of the fastest and slowest chip. . . 77
6.10 $Log(I_{LEAK})$ distribution for single-finger configuration in the 90nm test-chip. 78
6.11 RO frequency distribution for stacked gates plotted next to its corresponding layout in the 90nm test-chip. Vertical lines represent simulation corners from extracted layout. Distribution plots of a) RO frequency normalized to a constant, b) RO frequency normalized to its corresponding TT simulation corner. In b), less dense gates show a 5% increase in frequency compared to dense gates. 79
6.12 Effect of nearest neighbor poly-Si pitch on RO frequency and transistor leakage currents in the 45nm test-chip. Plots of RO frequency distribution normalized to the SS corners, NMOS $log(I_{LEAK})$ distribution and PMOS $log(I_{LEAK})$ distribution for 22 dies. 80
6.13 Bottom left plot shows the frequency distribution for 90° rotated ROs with horizontal gates in the 90nm test-chip. Non-rotated ROs in the top left plot show more variation. 99% confidence interval of σ/μ and σ^2 shown on the right verify that the top plot has more variations. 81
6.14 I_{LEAK} distributions for non-rotated and rotated ROs in the 90nm test-chip. 82
6.15 Effect of placement of gates horizontally and vertically in the 45nm test-chip. 83
6.16 Effect of symmetrical layouts which are mirror images of each other in the 45nm test-chip. 84
6.17 Effect of M1 coverage over gates in the 90nm test-chips. 85
6.18 Effect of M1 coverage over gates in the 45nm test-chips. 85
6.19 Effect of fixed poly-Si gate pitch in the 45nm test-chips. 86
6.20 Effect of STI on RO frequency and transistor leakage currents in the 45nm test-chips. NSTI uses gate isolation [7] instead of STI. 87
6.21 Effect of a longer S/D diffusion on RO frequency and transistor leakage currents in the 45nm test-chips. 88
6.22 Wafer maps of mean RO frequency and mean Log(ILEAK) in the 90nm process. X marks a defective chip. Location of dies on only half the wafer are known. 90
6.23 Comparison of RO frequency D2D variations for 2 wafers and the overall D2D variations in the 45nm process. 90

6.24 45nm WID statistics of RO frequency for 22 dies. The frequency is normalized to the SS corner frequency. 91
6.25 Spatial correlation coefficient (ρ) vs a) horizontal column spacing and b) vertical row spacing, for vertical and horizontal gates in the 90nm process. Dotted lines represent 99% confidence bounds for ρ. 92
6.26 Surface plot of normalized data of 36 chips in the 90nm process. RO of layout with vertically oriented gates show a 1% frequency shift between rows 4 and 5. 93
6.27 Plot of 45nm RO frequency normalized to zero mean and unit variance showing mask characteristics. Small systematic variation of $\approx 0.2 \times \sigma$. 94
6.28 Plots of correlation coefficient (ρ) vs spacing between inverters in a RO obtained from fitting measurement data to a exponential model. 95
6.29 90nm scatter plot of $Log(I_{LEAK})$ vs frequency showing a) mean $Log(I_{LEAK})$ and mean RO frequency for 36 dies, b) all $Log(I_{LEAK})$ and RO frequency for 3 layouts on 1 die. 96
6.30 Illustrates the estimation of the nominal value of L (L_m) in the case where only one parameter (L) is inferred. I_0 and F_0 are the measurement data. L_m is used to obtain the gradients and to find I_m and F_m which are used to normalize the data in the least squares estimation. 98
6.31 De-embedded distribution of L. 100
6.32 Top plots: surface plots of within-die RO frequency (F) and $Log(I_{LEAK})$ (I) of a typical die; Bottom plot: corresponding surface plots of estimated normalized L. 101

A.1 Small signal model of the integrator. 118

B.1 Arrangement of the 12 layouts in a tile for the 90nm test-chip. 122
B.2 Arrangement of the 17 layouts in a tile for the 45nm test-chip. 123

C.1 DG2020A data generator from Tektronix. 127
C.2 Keithley 2400 source meter. 128
C.3 Infiniium 54855A DSO 20GSPS oscilloscope from Agilent. 128
C.4 TLA5202 logic analyzer from Tektronix. 129
C.5 GPIB-ENET/100 box from National Instruments. 129
C.6 Test setup for collecting data. 133
C.7 PCB board for measuring the 90nm test-chips. 134
C.8 PCB board for measuring the 45nm test-chips. 135
C.9 Test setup for collecting leakage measurement data for the 45nm test-chips. 136

List of Tables

2.1 Random and systematic sources of process variations 20

6.1 Summary of the 45nm process. [2, 3, 4] 69
6.2 Characteristics of the layout configurations. 71
6.3 Dimensions of the transistors in the I_{LEAK} arrays. 74
6.4 Dimensions of the transistors in the inverters of the ROs. 74
6.5 Spatial correlation coefficient of neighboring ROs for different gate orientation and spacing direction. 93
6.6 Measurement conditions. 97
6.7 Sensitivities used in the 90nm process parameters extraction for one chip and one layout structure. 99
6.8 Comparison of 45nm and 90nm technology results. 102

Acknowledgments

I would like to acknowledge my friends and colleagues who have helped me throughout the last six years of my life in U.C. Berkeley.

First of all, I would like to acknowledge my dear Singaporean friends. Starting from the beginning, Engling and Ailee, whose hospitality I abused when I stayed with them for over a month during my first semester in Berkeley. I'm glad they are now happily installed in Silicon Valley with two happy kids. Yuenhui, Siewleng, Shaokun and Xiaofang, for their company and help in getting me through the prelims and quals, and Voon Bin, Ann, Bingru, Kah Cheong, Shawn, Mien Ling, Yingshi, Chuohao and Wanling for their patience in tolerating my less than exciting company.

My colleagues in the DCDG group are the most impressive bunch of people I have ever worked with. Both intelligent and fun loving, these people make graduate school interesting and less stressful. I will definitely miss our lunch outings and the long lunch time discussions that provide much entertainment as well as information. Engling, Radu, Bill, Josh and Sean are part of the previous generation and have all left school. Farhana, Zhengya, Renaldi, Zheng Guo, Seng-Oon, Ji-Hoon, Dusan, Kenny, Lauren, Jason and Vinayak are now the main actors in the DCDG group. All of them have created and continue to create a lively and fun environment in the group and made life tolerable, even during tapeouts. Throughout the years, I have come to know the staff at BWRC and realized what great people they are. Gary, Tom, Brian, Kevin, Brad, Brenda and Susan are all experts in their own areas and work hard to keep BWRC a center of excellence for research and an ideal oasis for graduate students to work and hang out.

Finally, the friends I made at IBM Austin Research Labs during my three summer internships played a very important part in my Ph.D endeavour. They were instrumental both in developing my research as well as helping me find a research position in IBM. J.B. Kuang is the nicest man I have ever met. His intelligence, wealth of experience and knowledge and his willingness to go out of his way to help made my stay in Austin very pleasant and accounts for my repeated internship there. ARL staff Fadi, Tuyet, Jay and others, other interns I met, Jerry, John, Vicky and Praveen made Austin, Texas feel fun and enjoyable. In addition, Kevin Nowka is unanimously voted as the best manager by the people in ARL and I agree totally.

I am sure I forgot a few friends or colleagues in this acknowledgment. Do send me an email to remind me so that I can save some time looking for your name. I recognize faces better than remember names and rest assured I have you in my head.

Thank you all for being part of this.

Chapter 1

Introduction

Scaling of CMOS technology into the sub-100nm range has introduced new design challenges in the form of increased variability. Studies have predicted the end of CMOS scaling if this trend continues and if the present design methodology remains unchanged. This thesis describes a technique to characterize and quantify systematic, random, die-to-die (D2D) and within-die (WID) transistor variability in order to manage it both from the manufacturing side and the design side. It demonstrates the viability of this methodology through the design, measurement and analysis of two test-chips in early 90nm and 45nm processes [5, 6, 7].

1.1 Motivation

CMOS device scaling has increased the impact of process variability to the point where it is now regarded as a major roadblock to further scaling [8]. The control of process fluctuations has not kept pace with rapidly shrinking device dimensions [9]. Furthermore,

the drive to improve performance has enticed device and circuit designers to operate at conditions which are more sensitive to variability.

Transistor oxide thickness (T_{ox}), threshold voltage (V_{th}) and gate length (L) variations have become more significant as the manufacturing process enters sub-100nm technology nodes. Currently, the technology is enhancing the capabilities of photolithography in volume production with sophisticated resolution enhancement techniques (RET) and increasing constraints in the form of restrictive layout design rules to continue printing ever smaller sub-wavelength features. The evolution of the wavelength of light used in photolithography over the years and the difficulty in printing sub-wavelength features is shown in Figure 1.1. Optical proximity corrections (OPC) and phase shift masks are being used in the sub-100nm technology nodes and immersion lithography will enter production at the 45nm node. Double patterning will soon be necessary at 32nm node and beyond. Even with shorter gate-lengths, the difficulty in scaling oxide thickness and supply voltage results in transistor performance that does not scale sufficiently, leading to the use of strained silicon and high-K dielectrics in order to continue the exponential growth in transistor performance. These techniques have, however, contributed to an already complex manufacturing process and compounded the sources of process variability.

Presently, the foundry measures a number of test-structures on the wafers and fit the I-V data into a model such as the BSIM SPICE model. Variability is captured in the statistics of the model parameters. This information is then used to generate the process corners whereby certain parameters are varied by a number of standard deviations from their nominal values. Monte Carlo simulations can also be performed using this statistical

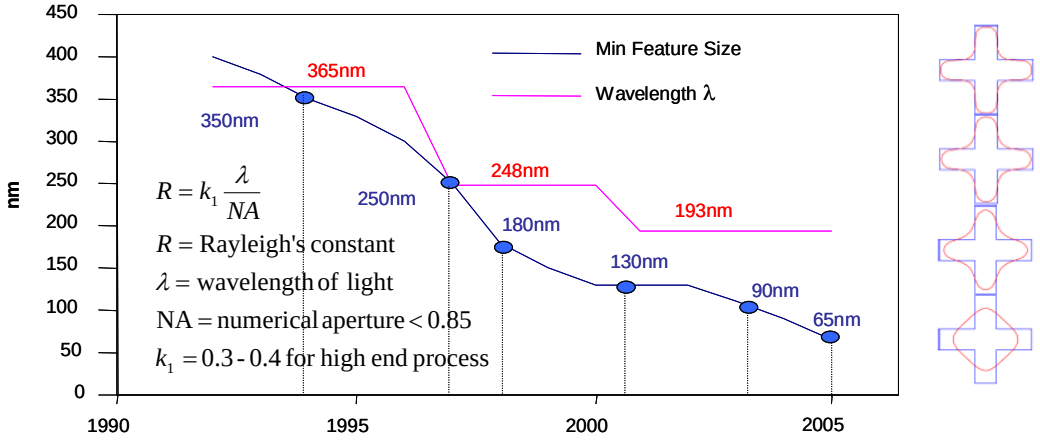


Figure 1.1: Evolution of the wavelength of light used in photolithography over the years. Presently, in the 90nm and 45nm nodes, 193nm wavelength is used. Rounding of printed features is shown on the right. [1]

information. In a typical VLSI design process, satisfying design corners is often necessary and sufficient to validate a design (Figure 1.2a). This approach regards all variations as D2D, with all devices on a chip having correlated process parameters. Deep sub-micron scaling has compounded the impact of variability and increased the amount of design margin to cope with worst-case scenarios. Since the tolerances on the process parameters do not track the scaling of their nominal values, designers are experiencing a proliferation of various design corners. Characterizing variability in a more detailed way (Figure 1.2b) would allow designers to reduce systematic variations and use the right amount of margins to obtain an optimal design that maximizes performance, power and yield.

1.2 Research Goal

The goal of this research is to help circuit designers reduce the impact of variations on circuit performance. This can be achieved in the following ways:

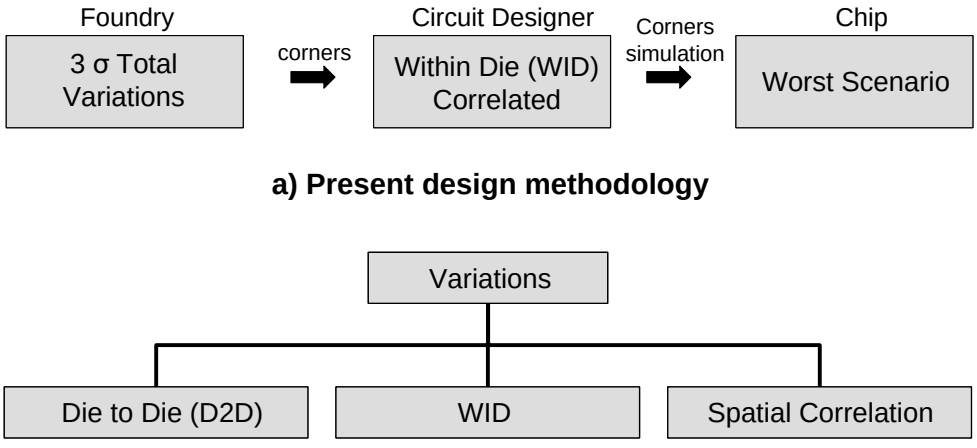


Figure 1.2: a) Present design methodology. b) Categorizing variations.

- Reduce the total amount of variations.

  This could be done by removing as much systematic variations as possible, selecting the right design parameters that minimize variations, and using circuit techniques to reduce variations. This research focuses on the impact of systematic layout-dependent variations and the design of circuits to measure variations.

- Design for the right amount of variations.

  This involves quantifying D2D, WID, systematic, random variations and quantifying the spatial correlation characteristics of variations and applying the correct amount of variations and correlation to the process parameters used in simulations.

- Exploiting the characteristics of spatial correlation.

  By using the fact that randomness can be reduced by averaging, and that correlated paths will be more likely to meet timing/power constraints, the placement of gates and paths can be maneuvered to minimize the impact of variations.

In order to do the above, a methodology has been developed whereby test chips to characterize the manufacturing process have been designed and fabricated. These chips contain arrays of closely spaced test-structures such as ring-oscillators, transistors in the off state and SRAM arrays. Individually addressable array of test-structures and circuit techniques to provide high resolution measurements were developed for characterizing transistor variations whereas non-invasive, high resolution techniques were developed for measuring variability in the individual SRAM cells of memory arrays. Measurement results provide quantitative information on layout dependent systematic effects, D2D variations, WID variations and spatial correlation. Finally, analysis of these results will lead to a set of guidelines for circuit designers to reduce the impact of variations.

1.3 Related Work on Process Characterization

1.3.1 Metrology Techniques

Electrical metrology techniques are used to collect large amounts of data on material properties and process information. These techniques measure current, voltage or charge and extract from them the physical dimensions of the shapes or the properties of the material. The quality of lines produced through photolithography has been evaluated in this way. ELectrical line-width metrology (ELM) measures the resistance of polysilicon (poly-Si) lines and uses that as a measure of line widths [10, 11, 12, 13]. Orshansky et al [12] used this method to investigate the effects of proximity and rotation on the printed gatelength for different configurations of poly-Si lines distributed over the reticle field and found layout-dependent behavior and systematic variation within the reticle field. Resis-

tance measurements can also be used to find the placement of the feature, metal step coverage over topography, wire edge-taper width and layer to layer alignment [14]. Capacitance measurement can be used to find wire to wire spacing and dielectric thickness.

Non-electrical metrology techniques include optical interferometry, profilometry, atomic force microscopy (AFM), and scanning electron microscopy (SEM) metrology. Optical interferometry is often used to measure film thickness, and profilometry is used to measure the height profile of a surface. High resolution SEM and AFM images can also be used to measure feature sizes such as line-widths. The advantage is that there is no need for special pads or circuitry for measurement, but the disadvantages are that throughput is low and the process is electrically destructive to the device.

These measurements are accurate but may not correlate strongly with the parameter that determines the electrical characteristics of the circuit. For example, measurement of resistance of poly-Si lines reflect the resolution and image quality of the regularly spaced poly-Si lines printed with lithography but does not directly measure the effective gatelength that affects transistor performance.

1.3.2 Measurement of transistor performance

Increased variability has prompted the development of more accurate process characterization techniques. In order to obtain accurate statistical models, a large number of transistors, in a wide variety of layout configurations with different combinations of transistor dimensions have to be measured. Several work have been published which measure a large array of transistor accurately in a short time. Prior to this work, Boning and his students [15, 16, 17] measured the frequency of ring oscillators (RO) specially laid out to

investigate the effects of layout on gates and interconnects. The RO array is made of small tiles that can be placed together to form a large array. The RO frequency is routed through multiplexers to one output where it is divided down and measured off-chip. Scan chains provide address bits to enable a single RO in the array and generate control signals for the multiplexers. In this way, the measurement process is simple, fast and accurate.

Other work involves inferring process parameters through measurement of transistor performance characteristics. Ouyang et al [18] measured transistor saturation current (I_{DSAT}) for an addressable array of transistors and inferred the transistor gate length L. Ohkawa et al [19] used a device matrix array which contains an array of transistors, capacitors, resistors and ring oscillators to evaluate WID variation in device parameters. Agarwal et al [20, 21] measured gate leakage and V_{th} in an addressable array of devices in IBM's 65nm SOI technology. Levacq et al [22] constructed a long array to measure the spatial frequency of intra-die variations using current measurements. A wide range of spatial frequency can be measured using the long array with closely spaced devices.

The traditional test structures for characterizing MOSFETs consist of measuring I-V data for different gate dimensions. Hence the most accurate way to statistically characterize the MOSFET is to collect I-V data for a large array of devices. This method requires huge amount of measurement and achieving efficiency is the main challenge. Wang et al [23] has designed a circuit that allows for rapid measurement of I-V data for an array of densely placed devices. This design consists of a dual-slope ADC and uses current mirrors to keep the integrated current at the same range. It achieves very good accuracy and a short measurement time.

Using the test structure in [20], Zhao et al [24] demonstrated that only 3 points in the I-V curve are needed to extract the process parameters. Wang et al [25] attempts to bridge the gap between statistical metrology and circuit design by modeling random process variability using current distribution measured at saturation and linear region.

1.3.3 Other circuit techniques for process characterization

Other interesting circuit techniques to characterize single inverter delays and SRAM performance have been published. To measure inverter delays, programmable ring-oscillators which allow for the selection of two ROs with overlapping inverter stages have been designed. By taking the difference between the periods of oscillation of the two ROs, the delay of the non-overlapped inverters can be extracted [26, 27]. Kim et al [28] uses a fixed RO to sample another RO that has been subjected to negative bias temperature instability (NBTI). By measuring the resulting beat frequency, the shift in frequency due to NBTI can be measured very accurately.

For SRAM measurement, besides measuring static currents from the bit lines as described in [29], work on connecting the SRAM columns together to form a RO whose frequency gives a measure of the performance of the SRAM array has also been shown in [30].

1.4 Dissertation Organization

Chapter 2 describes the characteristics of variability and how variations impact circuit design. It explains how spatial correlation affects performance and gives an overview

of the sources of process variations. Chapter 3 describes the sources of variations in the manufacturing process. It starts with an overview of the lithography process and its related sources of variations and goes on to cover variations from other manufacturing steps. Chapter 4 describes the circuits developed for measuring variability. These circuits were implemented in a 90nm and a 45nm test-chip that are described in chapter 6. Chapter 5 describes circuits for measuring SRAM performance. These circuits have not been implemented in test-chips and are part of the subject of future work. Chapter 6 presents the measurement and analysis results. In the recent 45nm process, restricted design rules and more complex OPC have been introduced in order to mitigate systematic poly-Si density and other layout-induced effects. The effectiveness of these measures will be evaluated. Finally, Chapter 7 concludes with a a set of guidelines for mitigating variability and a brief summary of our research accomplishments.

1.5 Acknowledgements

I wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center, the National Science Foundation Infrastructure Grant No. 0403427, wafer fabrication donation of STMicroelectronics, and the support of the Center for Circuit & System Solutions (C2S2) Focus Center, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program. I would also like to thank Ernesto Perea and the engineers at STMicroelectronics, Crolles for their support, Profs Andrew Neureuther, Tsu-Jae King-Liu and Costas Spanos and their students for helpful discussions, and Christopher Siow and Emmanuel Adeagbo

for their help with the measurements. Profs Jan Rabaey and Robert Leachman for agreeing to be in my qualifying examination and dissertation committees. Most importantly, I would like to thank my advisor Prof Borivoje Nikolić for his guidance and support.

Chapter 2

Characterization of Variations

Variability has to be measured and characterized in order to reduce its magnitude and impact on circuit performance. Categorizing variability into D2D and WID will allow for circuit designers to design for WID variations and use other techniques to reduce D2D variations. Categorizing variability into systematic and random will allow us to identify significant sources of variation. Systematic variation can then be reduced through tuning of the process, introduction of new process techniques or new design rules. Understanding the magnitude of the different sources of variations will allow process engineers to focus their resources on targeting the more important ones. Systematic layout-induced variations can also be reduced with the use of restricted design rules.

This chapter describes the characteristics of variability and gives an overview of the sources of transistor process parameter variations. Section 2.1 describes the characteristics of variability and how it is classified. Section 2.2 shows how spatial correlation affects the yield of a chip, and section 2.3 gives an overview of the sources of random and systematic

process parameter variability. Finally section 2.4 gives a summary of the characteristics of variations.

2.1 Characteristics of Variations

Variations in CMOS performance can be dynamic or static [31]. Dynamic variations such as electromigration, hot-electron effect, negative bias temperature instability (NBTI), self heating/temperature, SOI history effect, voltage drops across wire resistance, capacitive coupling or residual source/drain channel charge are time varying and change with the operations. Static variations are non time-dependent and are due mainly to the manufacturing process. Other causes of static variations such as voltage drop due to static current across a resistance can be accounted for in the design. In this work, the characteristics of static process variations will be studied.

Process variations can be systematic or random, and are generally characterized as within-die (WID), die-to-die (D2D) and wafer-to-wafer (W2W) [32]. WID and D2D classifications reflect some of the spatial characteristics of the variations. Those which vary rapidly over distances smaller than the dimension of a die result in WID variations whereas variations that change gradually over the wafer will cause D2D variations. D2D variations can be due to intra-field variations or field to field variations over the same wafer and also include other inter-die variations such as W2W variations. W2W variations reflect both the spatial as well as temporal characteristics of the process and cause different wafers to have different properties. In a typical design methodology, designs are made to satisfy the worst case corners which consist of the total WID and D2D variations.

Systematic variations are deterministic shifts in process parameters, whereas random variations change the performance of any individual instance in the design in an unpredictable manner. In practice, although many of the systematic variations have a deterministic source, they are not known at the design time, or are too complex to model and are thus treated as random. The resulting random variation component will have a varying degree of spatial correlation.

2.2 Impact of Variability on Circuits

Variability affects the yield of the chips. Yield is defined as the probability that a chip will meet the constraints of timing and power. A circuit with more design margin will have a higher yield. The challenge is in finding the smallest margin necessary to give the required yield so that performance is not overly constraint. In order to model the statistics of the circuit performance accurately both the amount of parameter variation and the spatial correlation of the these parameters between different gates have to be known. Characterizing the amount of variation involves making measurements of many devices and obtaining the standard deviation. This has traditionally been done in order to obtain corner information. However, spatial correlation has not been captured and incorporated into existing design flows.

By characterizing spatial correlation, excessive margins can be avoided and the effects of variations can be mitigated. Figure 2.1 illustrates how the correlation of path delays affects the yield. In a circuit with P parallel critical paths, if all the path delays are correlated, the probability of all the paths satisfying a timing constraint is equal to

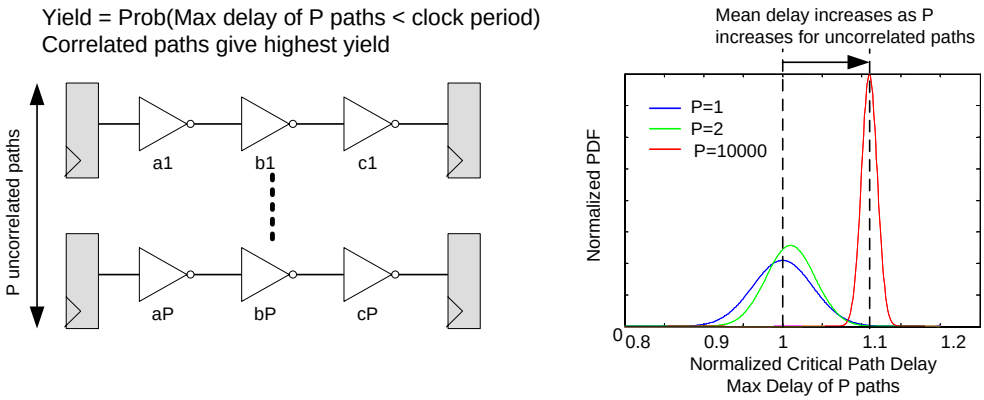


Figure 2.1: Correlation of path delays

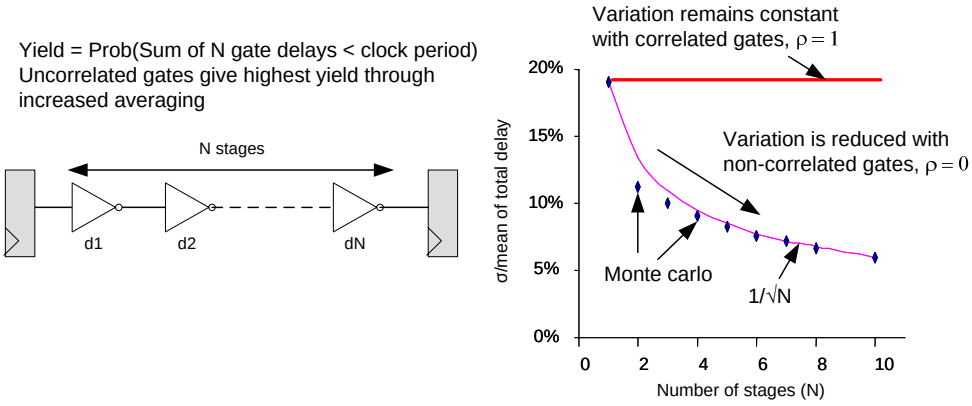


Figure 2.2: Correlation of gate delays in a critical path

the probability of one path meeting the constraint. On the other hand, if the paths are uncorrelated, the yield will be the product of the probabilities of the individual paths meeting the timing constraint, making it much lower and dependent on the number of paths. Figure 2.2 illustrates how the correlation of gate delays in a single path affects its yield. If the gate delays in a critical path are all uncorrelated ($\rho = 0$), the total path delay will have less variation due to averaging. Random variations can also be suppressed using longer logic paths and larger gate area transistors [33].

2.3 Sources of Process Variations

The primary sources of variability are the transistors themselves, interconnects, and the operating environment (supply and temperature) [31]. Many sources of systematic variability can be attributed to the different steps of the manufacturing process. The photolithography and etching processes contribute significantly to variations in nominal lengths and widths due to the complexity required to fabricate lines that are much narrower than the wavelength of light used to print them [1]. Variation in film thicknesses (e.g., oxide thickness, gate stacks, wire and dielectric layer height) is due to the deposition and growth process, as well as the chemical-mechanical planarization (CMP) step. Additional electrical properties of CMOS devices are affected by variations in the dosage of implants, as well as the temperature of annealing steps. In recent technologies, overlap error, mask error, shift in wafer scan speed, rapid thermal anneal and the dependence of stress on layout have become notable sources of systematic variations. Systematic variations will be discussed in detail in Chapter 3.

The most important reason for increased random device parameter fluctuations is that CMOS technology is scaling into atomic-scale dimensions [34]. The size of an atom is in the order of $\approx 1\ \mathring{A}$ or 0.1nm. In this regime, small variation of discrete numbers is significant when the total number is small. Smooth, continuous and distinct interfaces become granular and pebbled with atoms. Quantum mechanical properties of these discrete entities come into play and the classical models become increasingly inaccurate.

Random device parameter fluctuations stem mainly from Si/SiO_2 and poly-Si/SiO_2 interface roughness, line-edge roughness (LER) and doping fluctuations. Other effects in-

clude random strain distribution and non uniform temperature during annealing.

2.3.1 Interface roughness

The dielectric that separates the transistor gate from the substrate strongly affects the electrical properties of the device. The thickness of this layer (T_{ox}) and its dielectric constant are used to model the device and a thinner layer with higher dielectric constant gives better transistor performance. SiO_2 has been the gate dielectric of choice for CMOS technology due to its excellent interface properties and manufacturability. Scaling of T_{ox} has reached its limit of $\approx 2nm$ or 10 atomic layers of SiO_2 [35]. This limit is dictated by the exponential increase in gate leakage current due to quantum mechanical tunneling across the thin layer of dielectric. T_{ox} variation is around 1-2 atomic layers causing significant gate leakage variation. The impact of T_{ox} variation has been analyzed using 3D numerical simulation and incorporating quantum mechanical effects [36]. Results show that the impact of T_{ox} variation on V_{th} is small and strongly dependent on the correlation length. A pessimistic estimation of its effects for longer correlation length generates a variation of $\sigma_{V_{th}} \approx 35\text{mV}$.

2.3.2 Line-edge roughness

Line-edge roughness is due to statistical variation in the incident photon count during exposure, contrast of the aerial image, absorption rate, chemical reactivity and molecular composition of the resist [37, 31]. Atomic-scale behavior of the manufacturing process creates missing chunks of atoms from the surface of the gate along the width giving rise to the edge roughness. The impact of LER is stronger than that of interface roughness.

Subthreshold leakage current (I_{LEAK}) is strongly affected due to its exponential dependence on gate length [38]. LER becomes significant below the 50nm node and will severely affect performance at the 32nm node if gate-length control does not improve [39].

2.3.3 Random Dopant Fluctuation

Dopant atoms are injected into the channel region through ion implantation. This is followed by an annealing step to activate the dopant atoms and repair the damaged surface. The volume of channel region and the total number of dopant atoms are both so small that the number and position of dopant atoms in the channel of the transistor is effectively random. The impact of random dopant fluctuation (RDF) is exhibited through a large V_{th} variation. [40, 41, 42, 43, 44, 45, 46, 47]. It accounts for most of the variations observed in SRAM arrays and in analog circuits where systematic variation is small and random uncorrelated variation can cause mismatch that results in reduced noise margins. Studies have shown that retrograde body doping has the least amount of V_{th} fluctuation while uniformly doped case is worst.

2.3.4 Subthreshold Voltage Variation

The amount of V_{th} variation is reflected in $\sigma_{V_{th}}$ which varies with $\frac{1}{\sqrt{W \cdot L}}$ [48] where W is the width of the gate. $\sigma_{V_{th}}$ due to RDF and thin film interface roughness also varies with $\frac{1}{\sqrt{W \cdot L}}$ [42, 36]. Other effects such as non-uniform temperature distribution during annealing and random distribution of biaxial strain also cause V_{th} variation. These are described in Sections 2.3.5 and 3.2.

2.3.5 Strain

The impact of strain on the electrical resistance of silicon was discovered soon after Si was widely accepted as the material of choice for solid state devices [49]. Strain alters the band structure of Si, causing changes to properties such as bandgap, effective mass, mobility (μ), diffusivity of dopants, and oxidation rates [50, 51, 52].

The impact of strain on transistors depends on the crystal orientation (<100>or <110>direction), the type of strain (compressive or tensile), the direction of strain (one direction [53] (uniaxial) along the channel or orthogonal to the channel or in both directions (biaxial)), and the type of device (NMOS or PMOS). In the <110>channel direction tensile strain increases NMOS mobility by lowering the effective mass and the scattering rate of electrons. Compressive strain increases PMOS mobility by lowering the effective mass of holes [54, 53, 55]. In the <100>direction, PMOS mobility is insensitive to strain [56]. Threshold voltage (V_{th}) is affected by biaxial strain and less so by uniaxial strain [57]. Experimental studies have shown improvements in drive current of up to 60% due to strain [58, 59, 60, 61, 62].

Starting from the 90nm node, devices are intentionally strained to improve performance [63, 64, 65]. Shallow trench isolation(STI) using SiO_2 creates compressive strain on the substrate that varies with the distance from the edge of the STI/diffusion interface to the channel region. This effect is systematic and has been modeled in recent BSIM models. Other standard process-induced strain [50] include the use of silicide layers, and the contact etch stop layer (CESL). Silicide layers deposited onto the active and poly-Si region of the transistor introduce a compressive strain on the Si channel that is sandwiched between the

silicides due to differences in thermal expansion coefficient. Nitride films are deposited on Si to act as the CESL. These layers can also be used to generate large stresses which can be tensile or compressive. The amount of stress increases with the thickness of the nitride layer that can fill between the gates and the amount of contact of the CESL with the source/drain region. The latter depends on poly-Si pitches and spacer sizes. As these dimensions get smaller with scaling, the effectiveness of this method is reduced. Additional process steps have also been introduced to generate strain. A stress memorization technique consists of depositing a stressed film onto the device, annealing it at high temperature and stripping the stressed film. Re-crystallization of the poly-Si at high temperature under stress leads to the generation of residual stress [66]. Finally, strain is generated when SiGe is grown on Si because both materials have slightly different lattice constants. These epitaxial films generate global strain (when used as the substrate for the channel region), or local strain (when used only in the source/drain diffusion region of the transistor). Similarly to STI, the amount of strain depends on the length of source/drain diffusion.

The manufacturing process subjects the Si wafers to various temperatures and chemical, mechanical interactions. Different materials are grown or deposited on the wafer to create devices and interconnects. Strain is created on the substrate due to differences in thermal expansion coefficients or lattice constants of the different materials. This causes the electrical properties of devices to vary randomly [67]. Simulations show that the use of strain to improve performance also exacerbates the impacts of random process variability [68, 69].

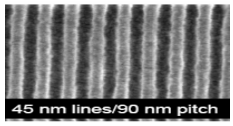
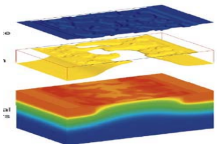
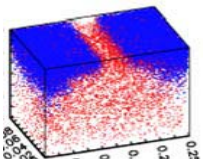
Parameter	Random	Systematic
Gate Length L	Line edge roughness (LER) [38] 	Lithography and etching: proximity effects, orientation. [10]
Gate Oxide Thickness T_{ox}	Si/SiO_2 & SiO_2/Poly-Si interface roughness [36] 	Non uniformity in the process of oxide growth.
Channel Dopant Concentration N_{ch}	Affects $\sigma_{V_{th}}$ [43] 	Non uniformity in the process of dopant implantation, dosage, diffusion.
Threshold Voltage V_{th} (non N_{ch} related)	Random anneal temperature and strain effects	Non-uniform annealing temperature [70] (metal coverage over gate). Biaxial strain.
Mobility μ	Random strain distributions	Systematic variation of strain in the Si due to STI, S/D area, contacts, gate density, etc.

Table 2.1: Random and systematic sources of process variations

2.4 Summary

The sources of transistor process parameter variations can be classified as random or systematic, D2D or WID. This classification together with information on the spatial correlation of these parameters can help designers manage variability in circuit performance. Random variations are mostly due to physical limits whereas systematic variations are due to the manufacturing process. A summary of the sources of random and systematic process variations is shown in Table 2.1.

Chapter 3

Process-induced Systematic Variations

Systematic sources of variations have been dominant in sub-100nm technologies. In many cases they can be reduced or accounted for in the design process. This chapter describes important sources of variation in the manufacturing process. Section 3.1 describes the sources of variations in a standard exposure system. This includes the mechanical system, the source of light, the optics and the mask. Section 3.2 describes other process steps involved in patterning of the transistor gate and formation of the gate channel. Section 3.3 gives a list of effects that are dependent on the layout of the circuits and finally, section 3.4 gives a summary of this chapter.

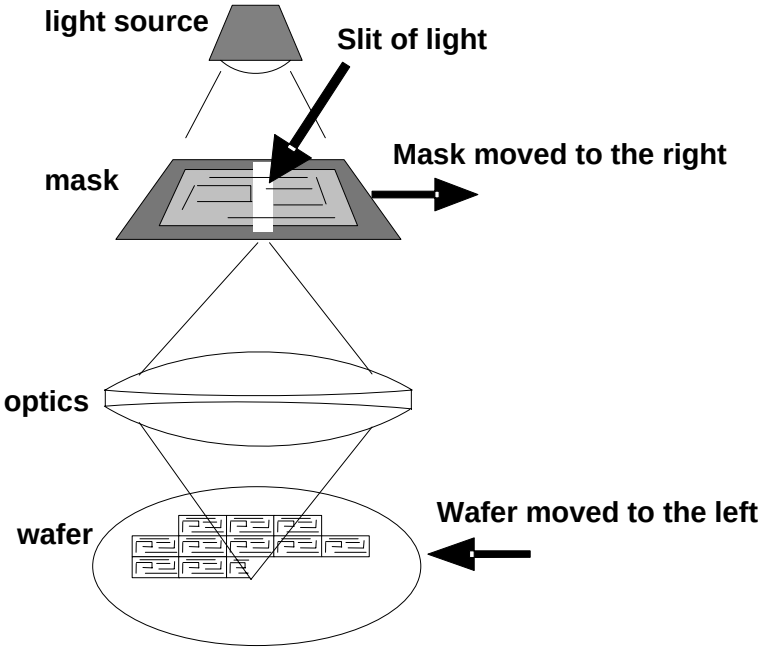


Figure 3.1: Step and scan photolithography using a slit of light.

3.1 Variations in the Exposure System

Present lithography systems employ a step-and-scan method, where the stepping is used to move the wafer between exposure fields. Within an exposure field, a narrow slit of light illuminates the mask and projects the mask pattern optically onto a wafer [71]. The mask and the wafer are moved simultaneously in opposite directions such that the slit of light scans the entire mask and projects the image onto the wafer. This is illustrated in Figure 3.1.

Modified illumination [72], refers to the use of partially coherent light sources and off-axis illumination schemes such as dipole, quadrupole and annular illumination. This technique improves resolution but also causes line-widths to vary according to their

orientation [73] and exacerbates the effects of lens aberrations [74]. Non uniformity and asymmetry of the light source can cause systematic variations in the reticle field. Movement of the wafer stage is controlled by look-ahead sensors. Changes in speeds during the scan may cause systematic variations in effective channel lengths over the reticle. Mask overlay error due to misalignment of the different masks can cause shifts in the position of poly-Si gate with respect to diffusion and source-drain contacts.

In sub-wavelength lithography, the effective line width depends on the surrounding features [72, 10, 12]. The process step of fabricating polysilicon (poly-Si) gates is shown in Figure 3.2. When exposed beyond a certain light intensity, positive resist will dissolve in the developer fluid. The exposed poly-Si gate stack layer will then be etched away leaving behind the transistor gates. Narrow poly-Si lines with varying pitch will have different channel lengths when exposed with the 193nm wavelength light, as illustrated in Figure 3.3. Dense lines also have higher depth of focus, and are more immune to defocusing of the optical system [75]. Optical proximity correction techniques in the mask processing add sub-lithographic assist features to control the printed critical dimensions (CD). However, their effect is limited due to the shallow depth of focus.

Lens imperfections are often characterized through aberrations, Figure 3.4. Aberrations create optical path differences for each pair of rays through the imaging system. Effects such as astigmatism and spherical aberrations cause differences in exposed patterns at the level of a reticle. Coma effect [74] is an aberration due to lens imperfection, which causes a gate surrounded by non-symmetrical structure to print differently from its mirror image [76].

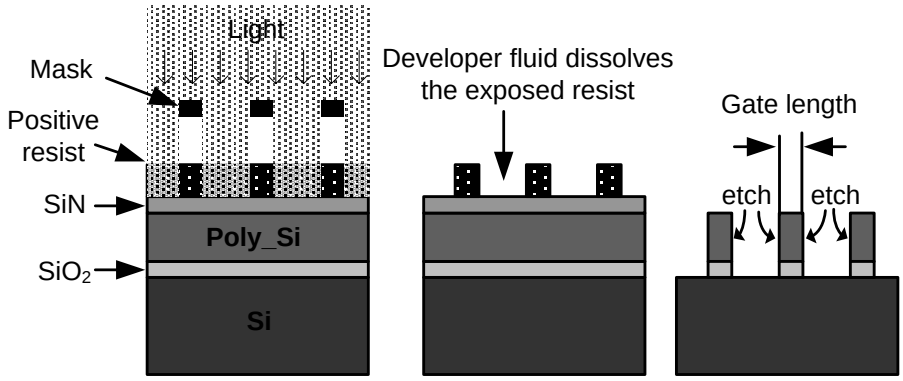


Figure 3.2: Formation of poly-Si gates.

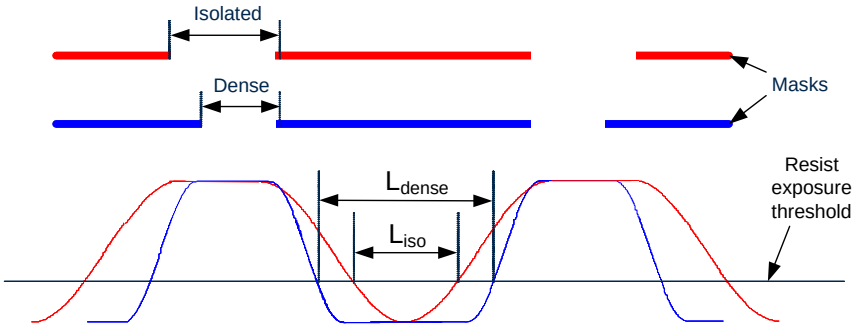


Figure 3.3: Isolated and dense lines.

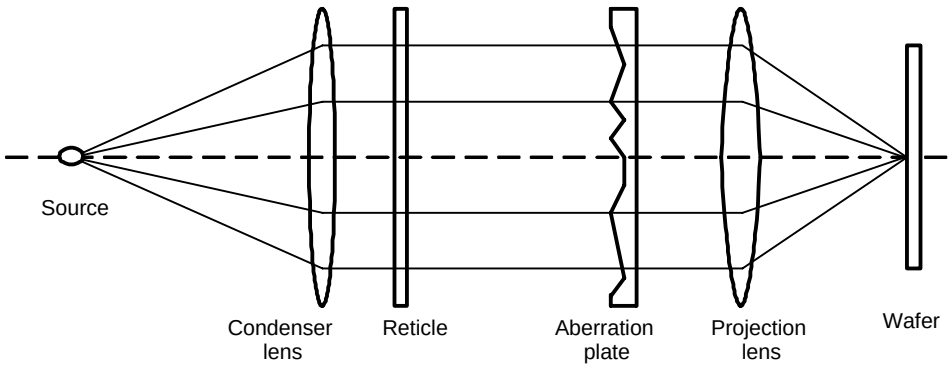


Figure 3.4: Lens imperfections.

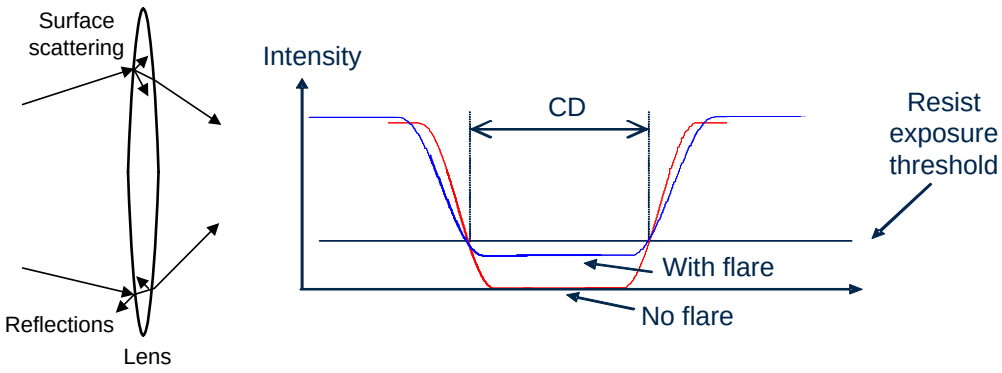


Figure 3.5: Effect of flare.

Flare results from the scattering and reflection of light through the projection system and causes variations in the effective CDs, Figure 3.5. In general, the amount of flare is dependent on the local pattern density in the mask [74]. Proximity effects, aberrations, and flare are usually not captured in the design process, and they induce layout-dependent systematic variations in the design.

Another source of variability comes from e-beam mask stitching [77] discontinuity. E-beam lithography is employed in writing the optical mask. In order to cover the large reticle field of the mask with a small e-beam field, it is necessary to construct the mask by drawing smaller e-beam fields and stitching the whole image together. Discontinuity at the boundaries between the smaller e-beam fields will result in stitching errors in the mask. Even though optics will reduce imperfections on the mask by a few times, aggressive scaling has made stitching discontinuities more significant.

3.2 Variations in Patterning and Other Manufacturing Steps

After the resist is spun onto the wafer and exposed, it undergoes post-exposure bake (PEB). This step is essential to activate the photo-active compound and set the resist exposure threshold [78]. If the temperature over the entire wafer is not even, this will result in different resist exposure thresholds over the wafer and cause D2D gate-length variation. This effect is mostly systematic, and wafers have exhibited a radial temperature profile during baking.

Resist development can also be a source of systematic wafer scale variation. The application of developer fluid, if not uniform over the wafer surface, can result in some parts of the wafer being more developed than others. This will also cause systematic gate-length variation.

Dry etching is employed in patterning of the poly-Si gates. This process suffers from microscopic loading effects [71] in which densely spaced poly-Si gates experience a lower etch rate than isolated poly-Si, resulting in dense poly-Si gates having longer gate-lengths than isolated gates.

Rapid thermal annealing (RTA) is used in activating implanted dopants and annealing defects without causing significant dopant diffusion [71]. Temperature distribution over the wafer surface could be non-uniform resulting in wafer-level systematic variations. For example, the edge of the wafer could experience more cooling and hence a lower temperature, resulting in a radial temperature gradient across the wafer. This effect could also cause wafer warping that results in variation in substrate stress. Finally, local variation in reflectivity of the wafer surface could cause local non-uniform heating, resulting in variation

in V_{th} and in external resistance (R_{EXT})[79]. The later includes mainly source-drain and contact resistances. It has also been shown in [70] that differences in anneal temperature can cause shifts in V_{th} due to non-passivation of interface states.

Shadowing during pocket implantation can result in different properties in the source and drain of transistors. These effects, like the coma effect, could cause symmetrical structures to have non-symmetrical properties.

Finally, well proximity effects (WPE) in which transistor performance depends on its distance from the edge of the well in which it sits, has been observed to cause significant variations [80].

3.3 Layout Dependent Variations

Layout-dependent effects that will be studied are shown below:

- Proximity effects

  Proximity effects refer to the dependence of L on transistor gate density or gate pitch. In the previous sections, proximity effects due to optical and etch loading effects have been described. These effects are also present in the mask fabrication since the same type of process steps are employed.

- Direction

  The step and scan lithography creates different properties in the two orthogonal directions of slit and scan. In the slit direction, variations are due to lens aberrations and result in more correlated features. In the scan direction, variations are due to dosage

and scan speed which are usually better controlled. Hence there is less systematic variation and features are less correlated. However, sudden discontinuities in scan speed can occur, resulting in the formation of lines along the slit direction, indicating the locations where there is a sudden change in scan speed. Mask stitching errors also contribute to systematic variations that are dependent on the x-y directions.

- Metal coverage over gate

  RTA can cause local non-uniformity in annealing temperature, resulting in changes in V_{th} with metal coverage over gate as described in section 3.2. Existing layout extraction tools only capture the capacitance of metal layers over the transistor gate. Shifts in V_{th} will be observed through measurement of the transistor I_{LEAK}.

- Strain effects

  In 45nm, the use of strained Si introduces mobility variations due to non-uniform stress in the Si substrate. The stress induced can be layout dependent. Sources of stress come from the shallow trench isolation (STI), a capping layer, silicide layers, stress memorization techniques and SiGe epitaxial films and they are affected by the layout such as gate pitch, contact separation, source/drain area, etc [81, 82, 83]. In the case of SiGe films in the source/drain diffusion region, a shorter source/drain length creates less stress. In the case of CESL nitride capping layer, the amount of strain depends on the thickness of CESL layer and the area of source/drain region covered with CESL which varies with the poly to poly pitch and the size of the gate spacers. A smaller contact area between the CESL and the source/drain region results in less strain on the transistors. The amount of stress starts to roll off quickly when the

sidewalls of the CESL starts touching each other. This occurs for thicker CESL layer.

In the design of analog circuits whereby matching of transistors and passives are very important, techniques for improving matching have been developed. These include the use of fixed gate pitches, dummy gates, common centroid layout, uniform distribution of metal over gate and larger design rule margins. Digital circuits have been slow to adopt them as the problem of matching was small and the area penalty was important. However, as variability begins to dominate the design, layouts are beginning to show more regularity. Work on design for manufacturability (DFM) involves setting design rules to ensure that variations in transistor process parameters are small. This has resulted in the implementation of fixed gate pitches, non-rotation of gates and regular layouts. In this work, the impact of layout on transistor performance in the presence of design rules can be evaluated by measuring an array of test circuits. These circuits, their implementation on test-chips and the measurement results and analysis of those chips will be described in Chapters 4, 5 and 6.

3.4 Summary

In summary, many sources of systematic variations are derived from the different steps in the manufacturing process as described in this chapter. Differences in transistor layouts and its proximity also contribute to systematic variation in performance.

Chapter 4

Measuring CMOS Performance Variability

Circuits that measure CMOS performance rapidly and accurately were designed to collect statistical data for a large array of devices. This chapter describes circuits that have been developed to make this type of measurements. It focuses on investigating the properties of transistors and the performance of CMOS logic circuits. Section 4.1 describes the circuits used to measure the subthreshold leakage current (I_{LEAK}) of an array of transistors. Section 4.2 describes the circuits for measuring the frequencies of an array of ring-oscillators (ROs) and for characterizing the spatial correlation of neighboring gate delays. Finally section 4.3 summarizes this chapter.

4.1 Leakage Current

Since I_{LEAK} varies exponentially with V_{th}, measuring I_{LEAK} gives a good idea of V_{th} variation. I_{LEAK} measurement of an array of transistors has been implemented in two test-chips in a 90nm and a 45nm technology.

4.1.1 Leakage Transistor Array

Figure 4.1 shows how the leakage current of a NMOS device in an array is measured. In the figure, the gates of the NMOS devices are connected to ground. The selection of the device to be measured is done with row and column bits obtained from two shift registers in order to reduce pad count. These address bits will supply either VDD or GND to the source of the transistor using a large inverter. The selected NMOS will have GND applied to its source and the other NMOS transistors will have VDD applied to their source. This will enable a subthreshold current to flow from the drain terminal to the source of the selected NMOS. Parasitic currents from the drain of the NMOS that are not selected will be measured and subtracted from the final measurement. Similarly for PMOS leakage, the address bits will apply VDD to the source of the selected PMOS transistor and GND to the source of the other non-selected PMOS transistors. The current will then be measured using an on-chip single-slope analog-to-digital converter (ADC) that will be described in the next section.

Figure 4.2 shows how the parasitic drain-to-bulk junction leakage (I_{nb}) and drain-to-gate leakage (I_{ng}) are removed from the measurements. Only NMOS leakage measurement is shown in the figure. The gates of the NMOS transistors are tied to GND and there

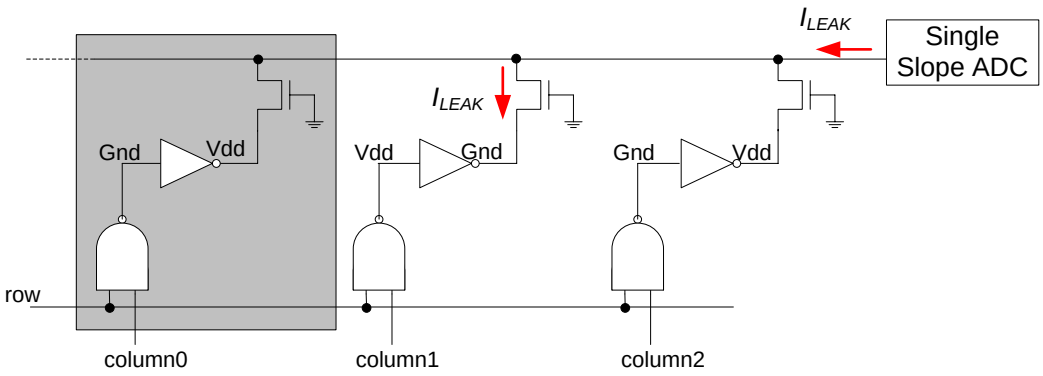


Figure 4.1: Standby leakage current measurement.

are parasitic I_{nb} and I_{ng} for all devices whether they are selected or not. Initially, none of the NMOS devices are selected and the parasitic current is measured as I_{s0}. Then one NMOS is selected and the current is measured as I_{s1}. Taking the difference between I_{s1} and I_{s0} gives the subthreshold leakage current of the selected NMOS, I_s. The size of the inverter is large enough so that even when subthreshold leakage is high, the voltage drop across the large inverter when GND is applied to the source of the selected NMOS is small. This is important as V_{GS} for the selected NMOS is the voltage across the large inverter.

4.1.2 Leakage Current Measurement

In the test-chips, an on-chip single-slope analog-to-digital converter (ADC) is used to measure transistor off-currents from 1nA to $1\mu A$ with 1nA resolution. The ADC (Figure 4.3) consists of a high-gain folded-cascode amplifier implemented with high voltage, thick oxide devices, a large on-chip metal fringe capacitor (10pF), and comparators. The amplifier uses large devices in order to increase gain and reduce the offset voltage. A large capacitance is necessary in order to remove the effects of parasitic capacitances and increase the integration time so that it can be measured more precisely. In addition, a fringe ca-

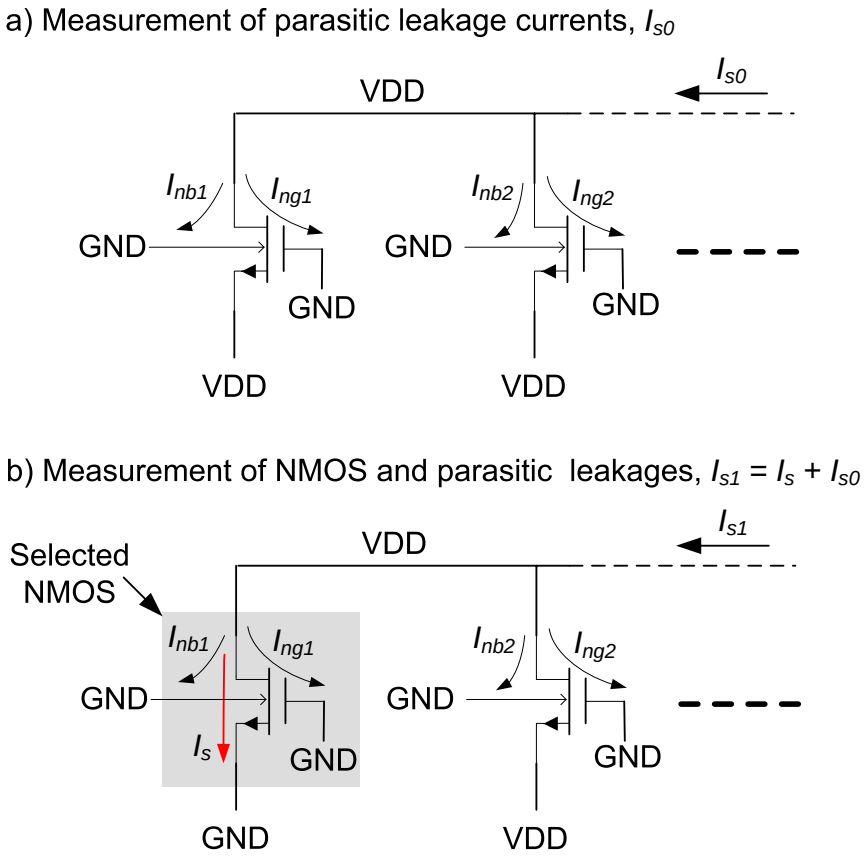


Figure 4.2: 45nm NMOS leakage current measurement. I_{nb} refers to drain to bulk junction leakage current and I_{ng} refers to drain to gate leakage current. Taking the difference between I_{s1} and I_{s0} gives I_s, the subthreshold leakage current of the selected NMOS transistor. Both NMOS and PMOS leakage currents are measured in the 45nm test-chip and only NMOS leakage current is measured in the 90nm test-chip.

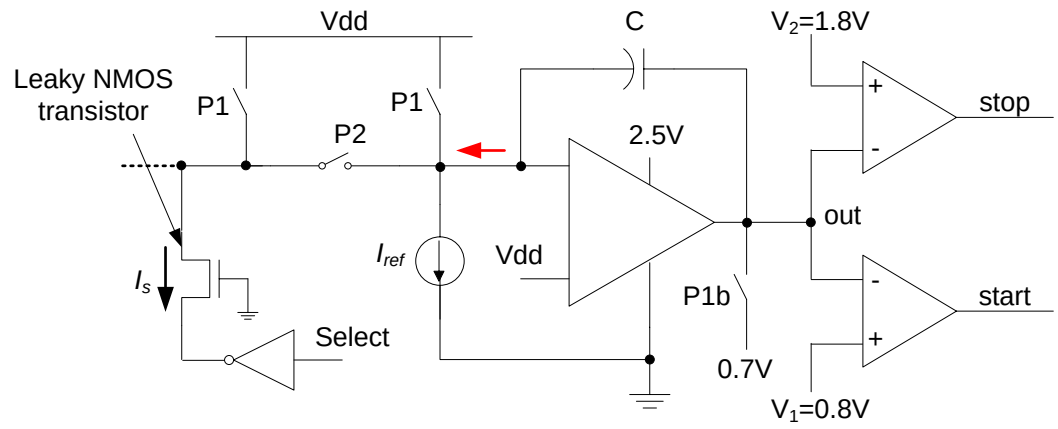


Figure 4.3: Single slope ADC for current measurement (90nm test-chip).

pacitor was used to reduce capacitance leakage current that is more significant with MOS capacitors.

The switches, controlled by signals P1, P2, and P1b provide the correct bias to the input and output of the amplifier and determine whether to measure the reference current (I_{ref}) only or the sum of I_{ref} and I_s. The latter is the sum of all parasitic gate and substrate leakage currents and the I_{LEAK} of the selected device when a device is selected. As mentioned in the previous section, by not selecting any of the devices, it is possible to measure the parasitic leakage currents and subtract them from the measurement. I_{ref} is obtained from an external source through current mirrors that divide the current down by a factor of ten.

During current integration, the output of the op-amp will ramp up, and as it crosses the voltages V1 and V2, the start and stop signals will be generated. The timing diagram for the signals in the integrator is shown in Figure 4.4. Measuring the time interval between start and stop signals will give the current (Figure 4.5).

In the integrator the relationship between the integrator output and the integrated

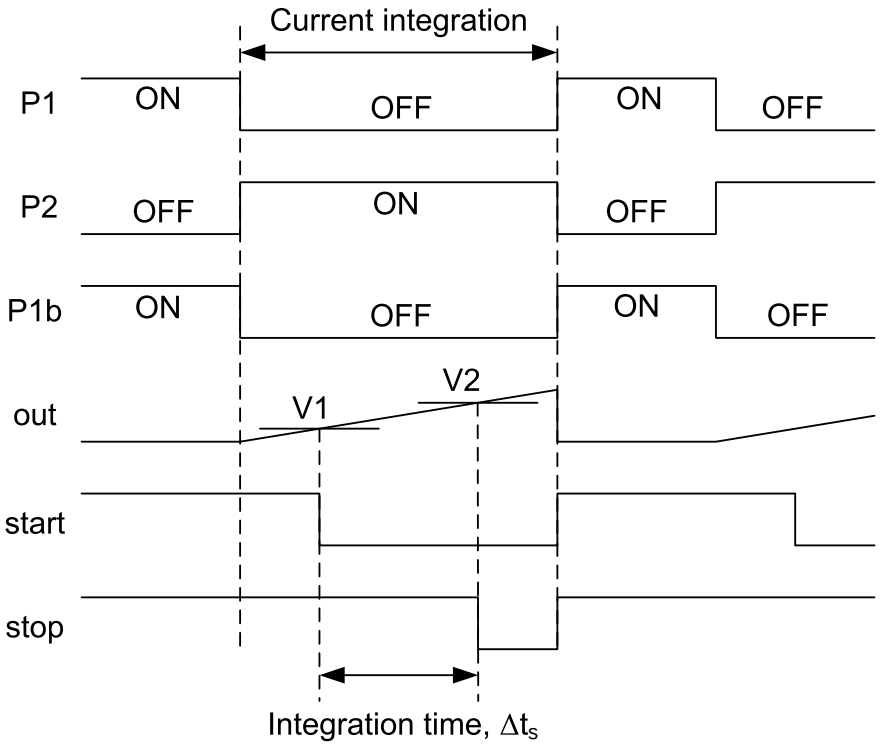
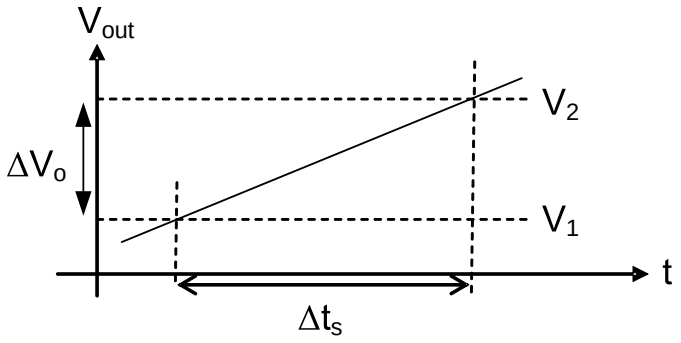


Figure 4.4: Timing diagrams of the integrator control signals and the output signals.



$$\frac{\Delta t_{ref} - \Delta t_s}{\Delta t_s} \sim \frac{I_s}{I_{ref}}$$

Δt_{ref} = integration time when measuring I_{ref}

Δt_s = integration time when measuring $I_{ref} + I_s$

Figure 4.5: Current measurement procedure.

current is given in Equation 4.1.

$$V_0 = \frac{1}{G_{m0}} \cdot I - \frac{1}{C} \cdot \int I \cdot dt \tag{4.1}$$

where

V_0 = integrator output voltage

G_{m0} = transconductance gain of the amplifier for amplifier output voltage at V_0

I = integrated current

C = integrating capacitance

t = time

Integrating from V1 to V2 for constant currents I_{ref} and $I_{ref} + I_s$, the Equations 4.2 are obtained.

$$\begin{aligned} \Delta V &= I_{ref} \cdot \Delta \frac{1}{G_m} - \frac{I_{ref}}{C} \cdot \Delta t_{ref} \\ \Delta V &= (I_{ref} + I_s) \cdot \Delta \frac{1}{G_m} - \frac{(I_{ref} + I_s)}{C} \cdot \Delta t_s \end{aligned} \tag{4.2}$$

where

$\Delta V = V_1$ - V_2

$\Delta \frac{1}{G_m} = \frac{1}{G_{m1}}$ - $\frac{1}{G_{m2}}$

I_{ref} = integrator reference current

I_s = current from the array

Δt_{ref} = time to integrate I_{ref} from V_1 to V_2

Δt_s = time to integrate $I_{ref} + I_s$ from V_1 to V_2

When the gain of the amplifier is sufficiently large, these two equations will give the linear relationship shown in Figure 4.5. This is shown in Equation 4.3. By measuring the integration times Δt_{ref} and Δt_s, the measured current can be obtained in terms of I_{ref}.

$$\frac{\Delta t_{ref} - \Delta t_s}{\Delta t_s} = \frac{\frac{I_s}{I_{ref}}}{1 - (I_b + I_{ref}) \cdot \frac{\Delta \frac{1}{G_m}}{\Delta V}} \approx \frac{I_s}{I_{ref}} \tag{4.3}$$

The accuracy of the current measurement can be improved by removing uncertainty in the value of I_{ref}. I_{ref} is generated by dividing down an external current source (I_{ext}) by 10 times using a current mirror. Mismatch in the current mirror and parasitic currents in the diffusion of the transistors in the current mirror contribute to errors in the current measurement. In a simple measurement procedure, a current of 1uA is supplied to the current mirror and divided down to 100nA to act as I_{ref}. Mismatch in the current mirror and parasitic junction and gate currents in the current mirror will change this value as shown in Equation 4.4.

$$I_{ref} = ((I_{ext} + I_{pad}) \times 0.1 \times M + I_{mirr} \tag{4.4}$$

where

I_{pad} = parasitic currents from the I/O pad

I_{mirr} = parasitic currents at the diffusion terminal of the mirror that connects to the input

of the current integrator

M = mismatch of the current mirror

The parasitic currents are fixed and can vary from die to die. Mismatch is small since the transistors used are long and wide ($M \approx 1$). Since the external current source is very accurate, a better measurement procedure to remove the effect of parasitic currents is to make a measurement with $I_{ext} = 1\mu A$ followed by a second measurement with $I_{ext} = 1.1\mu A$. Knowing the difference in mirrored current is 10nA, an accurate value of I_{ref} can be obtained and used in the measurement. An improved current measurement circuit was presented by Sato et al [84] which reduces the contribution of parasitic leakage current from non-selected transistors in the array and virtually eliminates the constraint on the number of transistor allowed.

The following steps describes the current measurement procedure. The first two steps describe how I_{ref} is determined accurately and the other steps describe how parasitic leakage currents is removed from the measurement.

1. Unselect all devices and the array and make a measurement with $I_{ext} = 1\mu A$, obtain Δt_{ref}. I_{ref} contains the mirrored current and all the parasitic currents.

2. Unselect all devices and the array and make a measurement with $I_{ext} = 1.1\mu A$, obtain Δt_{s0}. The external current I_{ext} is 100nA larger in this case. After going through the current mirror this difference is divided down by 10 to 10nA. Since the current is now

10nA larger than I_{ref}, the corresponding measured current $I_{s0} = 10nA$.

$$I_{s0} = \frac{\Delta t_{ref} - \Delta t_{s0}}{\Delta t_{s0}} \cdot I_{ref} \Leftrightarrow I_{ref} = \frac{\Delta t_{s0}}{\Delta t_{ref} - \Delta t_{s0}} \cdot I_{s0} \tag{4.5}$$

3. Unselect all devices but select the array and make a measurement with $I_{ext} = 1\mu A$, obtain $\Delta t_{parasite}$

$$I_{parasite} = \frac{\Delta t_{ref} - \Delta t_{parasite}}{\Delta t_{parasite}} \cdot I_{ref} \tag{4.6}$$

substitute I_{ref} from Equation 4.5 into Equation 4.6, we obtain

$$I_{parasite} = \frac{\Delta t_{ref} - \Delta t_{parasite}}{\Delta t_{parasite}} \cdot \frac{\Delta t_{s0}}{\Delta t_{ref} - \Delta t_{s0}} \cdot I_{s0} \tag{4.7}$$

4. Select one device and make a measurement with $I_{ext} = 1\mu A$, obtain Δt_s

$$I_s = \frac{\Delta t_{ref} - \Delta t_s}{\Delta t_s} \cdot \frac{\Delta t_{s0}}{\Delta t_{ref} - \Delta t_{s0}} \cdot I_{s0} \tag{4.8}$$

5. Subtract parasitic currents from the measurement to obtain I_{LEAK} of the measured device.

$$I_{LEAK} = I_s - I_{parasite} \tag{4.9}$$

Two folded cascode amplifiers were designed, one for measuring NMOS leakage (Figure 4.6) and the other for PMOS leakage (Figure 4.7). Each amplifier is optimized for its operations at nominal conditions and has a DC gain of more than 60dB over the voltage range of interest. This is more than sufficient to achieve the required measurement

resolution. Figure 4.8 shows the simulated BODE plot for the amplifier used for NMOS I_{LEAK} measurement in the 90nm process.

The first amplifier for PMOS leakage measurement has an input bias voltage of 0V and the second amplifier for measuring NMOS leakage has an input bias voltage of VDD. In both 90nm and 45nm technologies, nominal VDD is around 1V. The transistor sizes in the amplifier remains the same and layout is also very similar in both technologies. Differences are due to design rules for the dimensions of contacts and certain spacing rules that differ between the design kits from the two technologies. A single stage amplifier was chosen in order to avoid problems with loop stability, and a folded-cascode design was chosen for its simplicity and its ease in providing the required gain. Long gate-lengths were used to increase the gain, and large transistors were used to improve matching and reduce offset voltages. Much effort was made in the layout of the transistors in order to maximize matching. Common centroid layout and dummy poly-Si lines were used to ensure close matching. In Monte Carlo simulations using the 45nm models, a $3 \times \sigma$ input offset voltage of 2.4mV was obtained. Since the offset voltage affects the voltage at the drain of the leakage transistors, it does not affect I_{LEAK} significantly.

The schematics of the comparator is shown in Figure 4.9. These comparators are used to generate the start and stop signals when the output of the integrator crosses the threshold voltages. The start and stop signals are then buffered and output to pads. The time interval between the two signals are measured off-chip with a logic analyzer, the TLA5202 from Tektronix, which is capable of 125ps timing resolution. The long integration time of more than 50 μs easily permits a current resolution of less than 1nA. Thick oxide

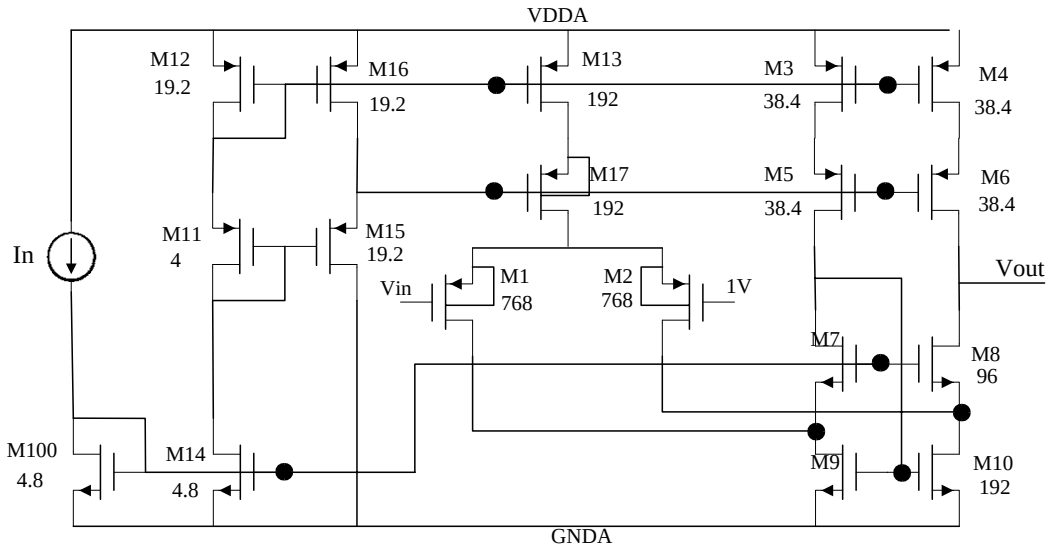


Figure 4.6: Schematics of the folded cascode amplifier for measuring NMOS leakage current. Gain is optimized for input bias of VDD=1V. Transistor widths are indicated in μm next to the individual transistors. All gate lengths are $0.35\mu m$ and $In = 90\mu A$. All transistors whose bulk connections are not indicated have it connected to VDDA(PMOS) or GNDA(NMOS).

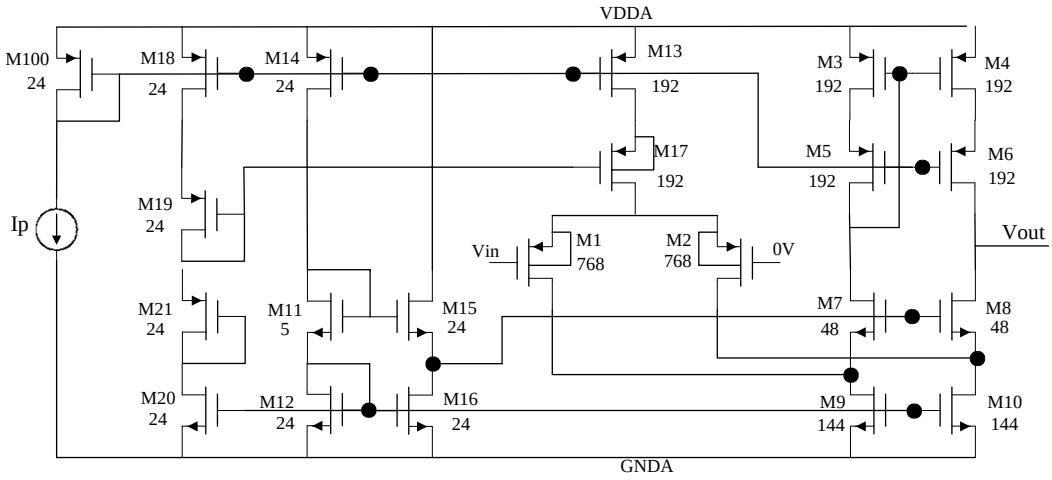


Figure 4.7: Schematics of the folded cascode amplifier for measuring PMOS leakage current. Gain is optimized for input bias of GND=0V. Transistor widths are indicated in μm next to the individual transistors. All gate lengths are $0.35\mu m$ and $Ip = 120\mu A$. All transistors whose bulk connections are not indicated have it connected to VDDA(PMOS) or GNDA(NMOS).

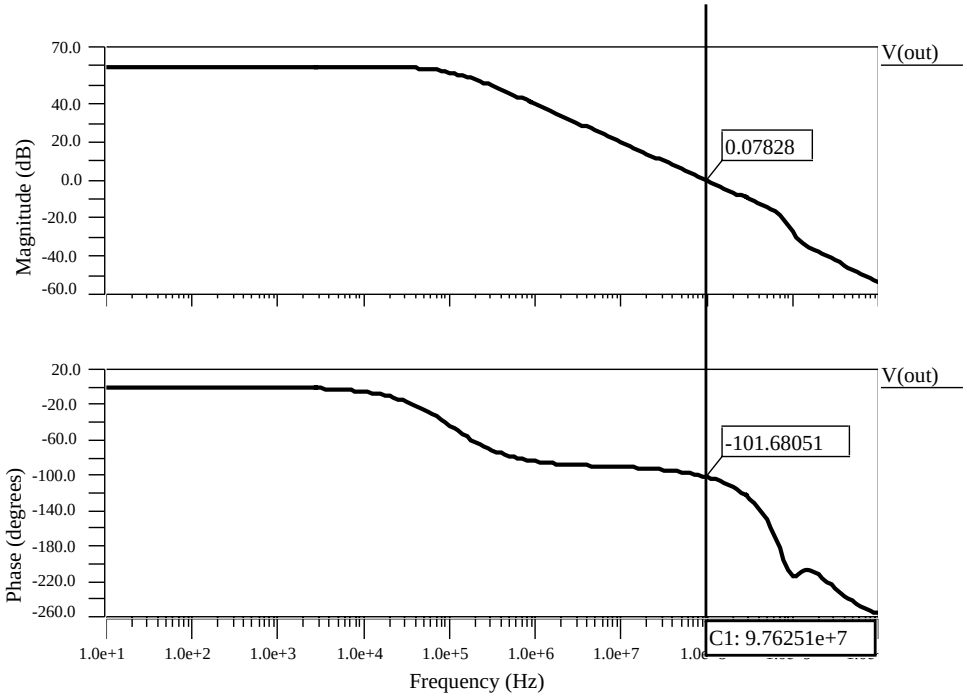


Figure 4.8: BODE plots of the amplifier used for NMOS I_{LEAK} measurement in the 90nm test-chip. Top plot shows a DC gain of $\approx$ 60dB and bottom plot shows a phase margin of $\approx 80^{\circ}$.

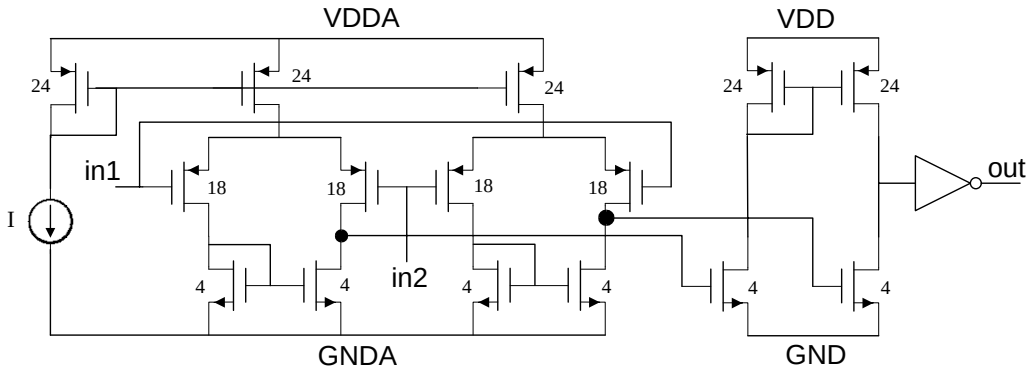


Figure 4.9: Schematics of the comparator for generating the start and stop signals when the output of the integrator crosses the threshold voltages.

transistors were also used in the comparator and transistor matching was improved with common centroid layout and dummy poly-Si lines.

Simulation results that demonstrate the precision of the current measurement circuit is shown in Figure 4.10. The measurement circuit as well as the leakage array is modeled in SPICE and simulated. I_s is the true subthreshold leakage current of the transistor under test and I_{s_cal} is the leakage current that is calculated from the integration times as described previously. A maximum error of $\approx 2\%$ is obtained for currents greater than 1nA. The accuracy of the current measurement is verified in the 90nm test-chip by measuring the same transistors repeatedly. Figure 4.11 plots 160 leakage measurements performed continuously over a 27 hour period for 3 layouts. A σ/μ of $\leq 1.3\%$ is measured.

In the 90nm test-chip, 2.5V thick oxide transistors were used and VDDA = 2.5V. However, in the 45nm test-chip, the process only had a 1.8V option and the analog circuits were migrated to VDDA = 1.8V and verified to work according to specifications. The BODE plots of the amplifiers for measuring NMOS leakage currents and PMOS leakage currents in the 45nm test-chip are shown in Figures 4.12 and 4.13 respectively. A gain of

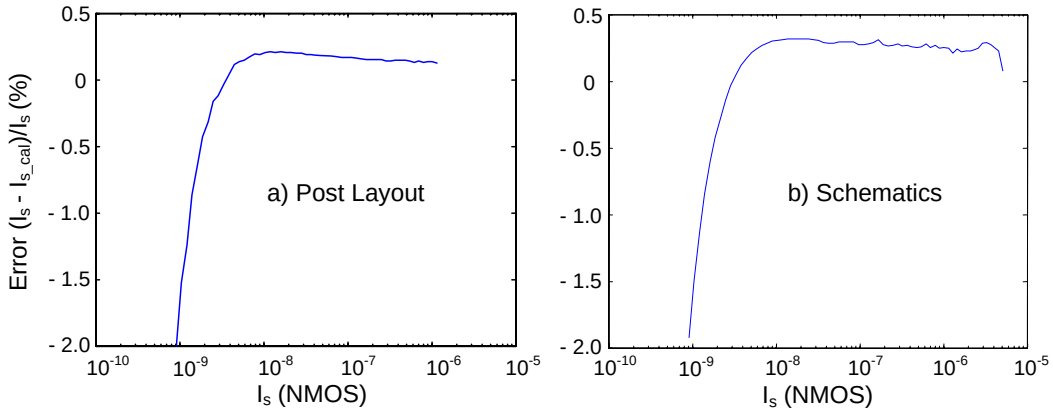


Figure 4.10: Simulation results for the current measurement circuit. Plots of leakage current measurement error vs leakage current. a) shows post-layout simulation results and b) shows pre-layout simulation results.

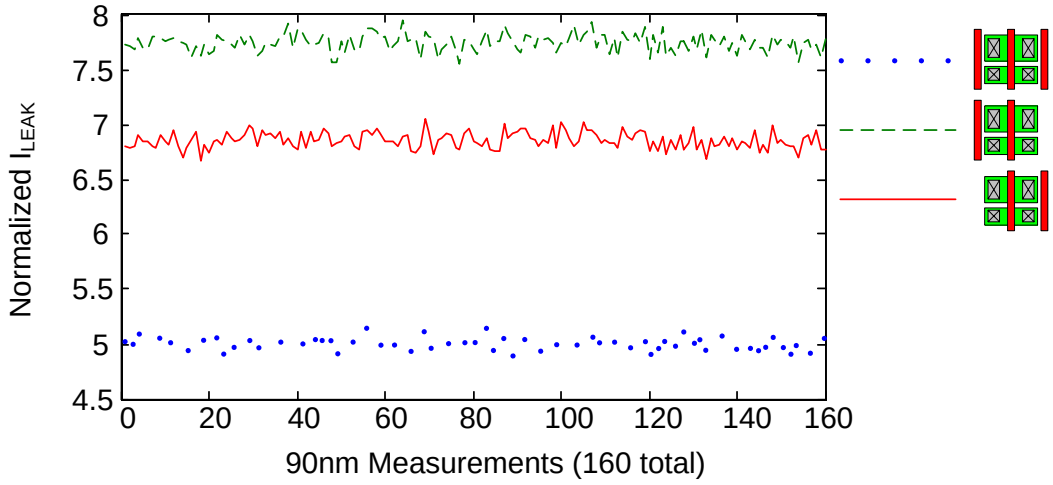


Figure 4.11: Repeated measurement of I_{LEAK} to demonstrate accuracy in the 90nm process. Plots 160 measurements of NMOS I_{LEAK} of the same 3 layouts in the same tile taken over 27 hours. A σ/μ of $\leq 1.3\%$ is measured.

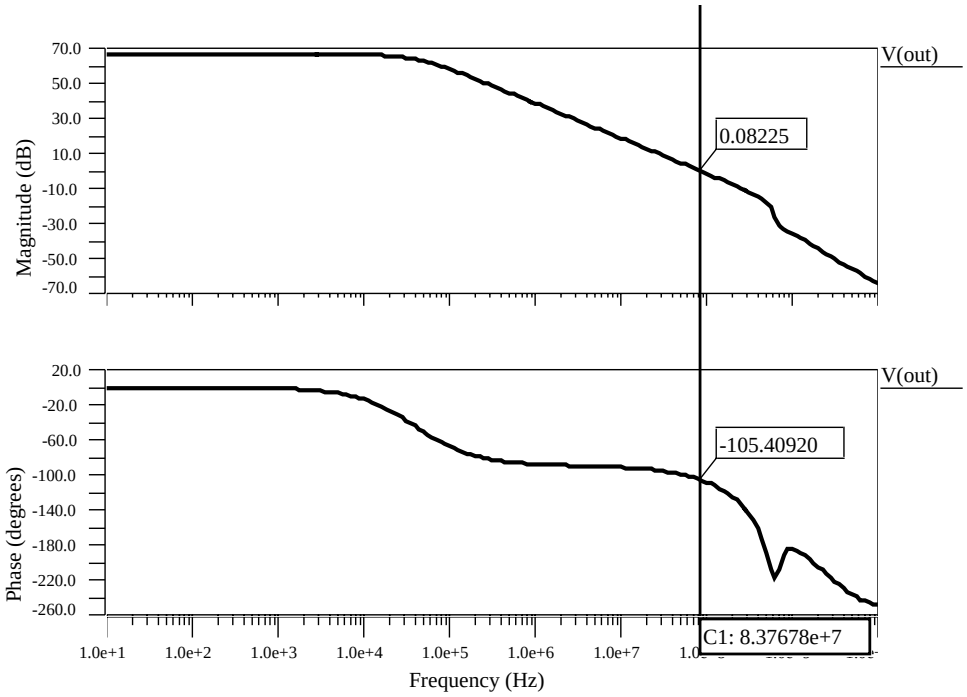


Figure 4.12: BODE plots of the amplifier for NMOS I_{LEAK} measurement in the 45nm test-chip. Top plot shows a DC gain of $\approx$ 60dB and bottom plot shows a phase margin of $\approx 75^{\circ}$.

$\approx$ 60dB is achieved in both cases with $\geq 60^{\circ}$ phase margin. Figure 4.14 shows the circuits for measuring PMOS leakage current in the 45nm test-chip. Analog supplies, comparator threshold voltages and amplifier output bias voltages are different from the 90nm test-chip. For NMOS leakage measurement, the output of the integrator, V_{out} is biased to 0.4V at the start of each measurement. Similarly to the 90nm test-chip, the accuracy of the current measurement circuit is verified by measuring the same transistor repeatedly, and the results are shown in Figure 4.15. This time, σ/μ of 1.4% and 11.5% are obtained for PMOS and NMOS I_{LEAK} respectively. The more significant noise in the NMOS I_{LEAK} measurement is due to a smaller mean NMOS I_{LEAK}.

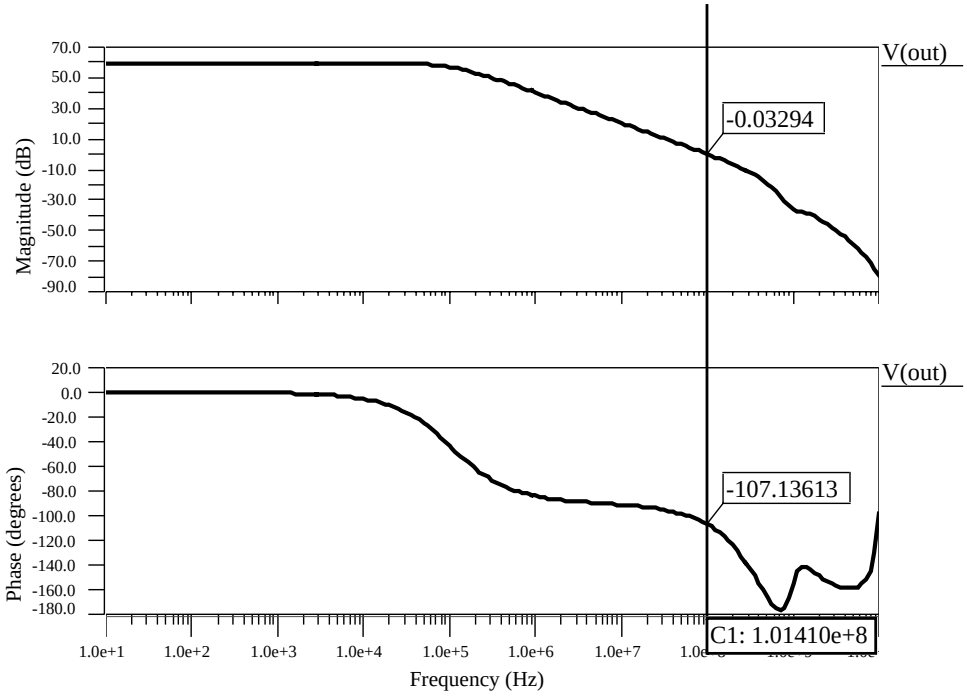


Figure 4.13: BODE plots of the amplifier for PMOS I_{LEAK} measurement in the 45nm test-chip. Top plot shows a DC gain of $\approx$ 60dB and bottom plot shows a phase margin of $\approx 73^{\circ}$.

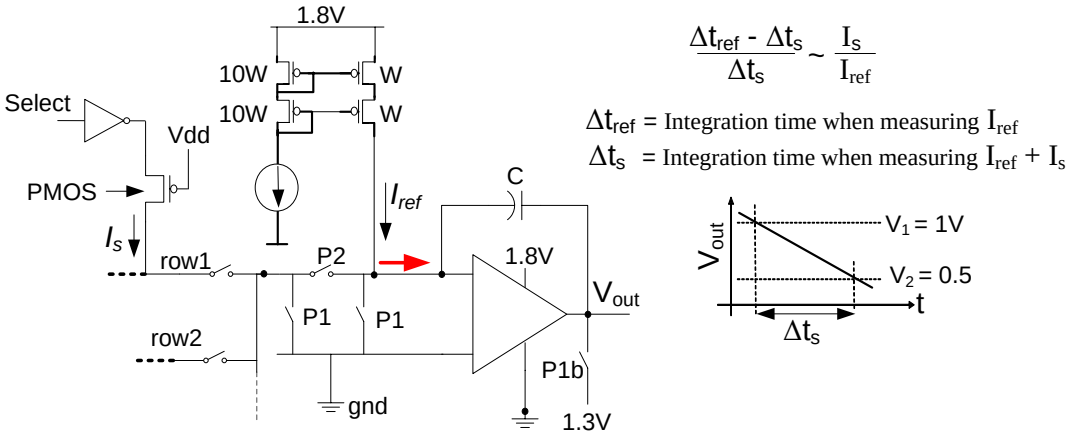


Figure 4.14: PMOS leakage current measurement for the 45nm test-chip.

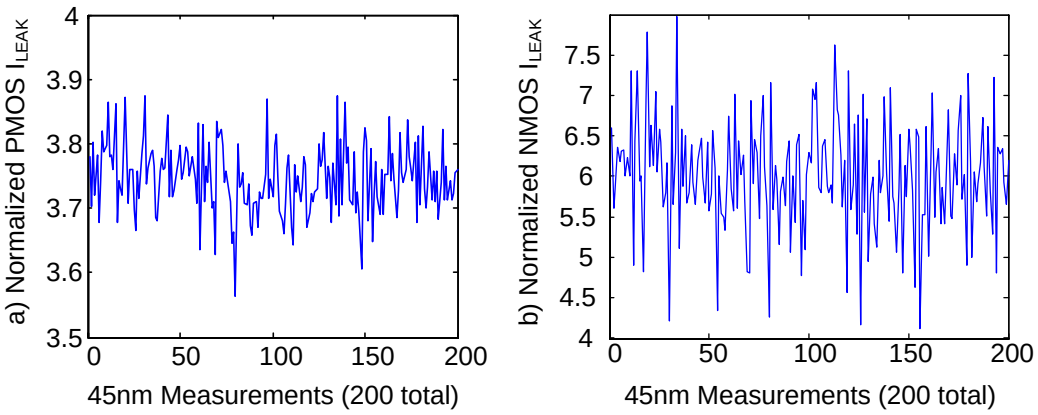


Figure 4.15: Repeated measurement of I_{LEAK} to demonstrate accuracy in the 45nm process. a) and b) plot 200 measurements of PMOS and NMOS I_{LEAK} of the same layout in the same tile. σ/μ of 1.4% and 11.5% are measured for PMOS and NMOS I_{LEAK} respectively.

4.2 Frequency

Ring-oscillators (RO) are commonly used to characterize transistor performance. They are small, easy to implement and can be measured easily, accurately and rapidly. Long ROs average random variations of the individual stages exposing systematic effects whereas short ROs have less averaging and can be used to measure random variations. Modified ROs can also be used to measure delay of individual gates with good accuracy as mentioned in chapter 1 [26, 27]. In this section, circuits that use ROs to characterize transistor performance variability and spatial correlation of gate delays will be described.

4.2.1 Ring Oscillator Array

In the test-chips, frequency measurements of a large array of ROs are made and used to characterize performance variations and systematic effects. Two scan chains for row and column addresses are used for addressing the array. Two address bits per row and a column bit are used to generate two control bits for each RO. One bit will enable

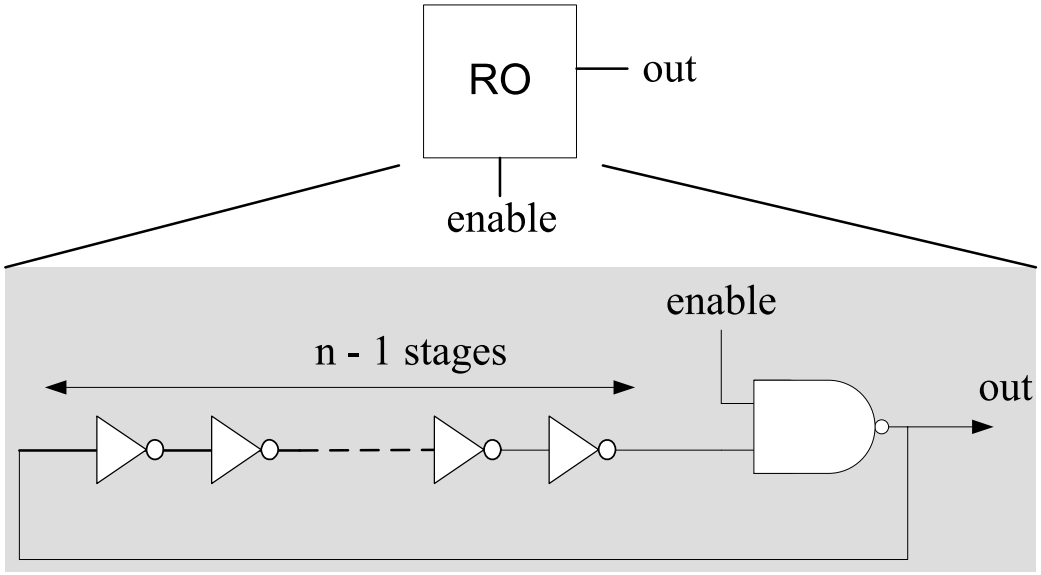


Figure 4.16: RO with n stages. Address bits will enable the RO by applying VDD to the *enable* signal. The output of the RO (*out*) goes to the input of the multiplexer.

the selected RO through a NAND gate placed in the ring ((Figure 4.16)), and the other will select the multiplexer to output its own RO frequency instead of the previous. The multiplexer output goes through a series of multiplexers to a frequency divider that outputs to a pad (Figure 4.17) [15, 16, 17, 85]. This design makes it very easy to place ROs together to form an array. Frequency is measured off-chip with a 20GSPS oscilloscope and averaged over ≈ 100 periods. Accuracy of the measurement is shown in Figure 4.18 and Figure 4.19. The same RO is repeatedly measured and noise is presented as the σ/μ of the RO frequency. A σ/μ of 0.15% and 0.02% was obtained from the 90nm and 45nm test-chips respectively.

In both 90nm and 45nm test-chips, 13-stage ROs are used for the single finger transistor layouts. For the layouts with a stack of 3 gates in the 90nm test-chip (described in chapter 5), 7-stage ROs are used. The number of stages is chosen such that the NAND gate in the RO contributes less than 10% to the total RO delay. This ensures that the

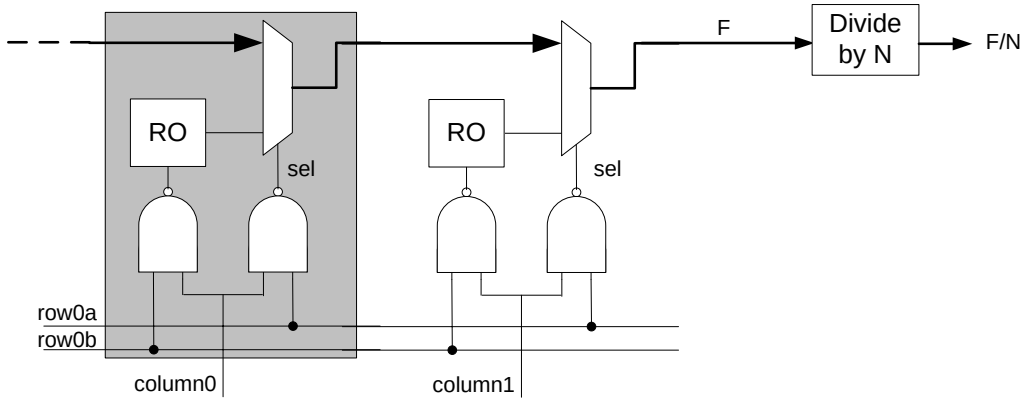


Figure 4.17: RO frequency measurement.

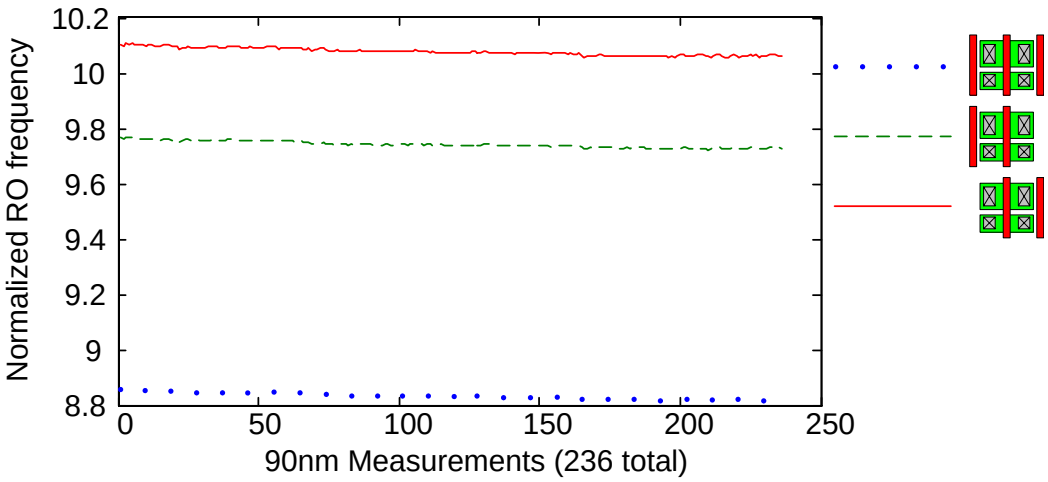


Figure 4.18: Repeated measurement of RO frequency to demonstrate accuracy in the 90nm process. Plots 236 measurements of RO frequency of 3 layouts in the same tile taken over 18 hours. A σ/μ of 0.15% is measured.

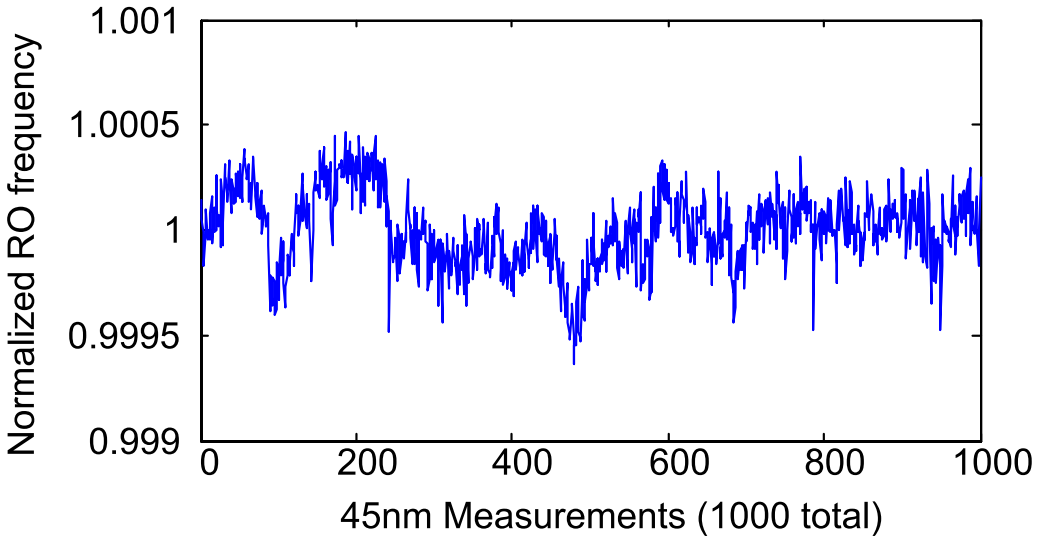


Figure 4.19: Repeated measurement of RO frequency to demonstrate accuracy in the 45nm process. Plots 1000 measurements of RO frequency of the same layouts in the same tile. A σ/μ of $\leq 0.02\%$ is measured.

RO frequency is more sensitive to the inverter delays than the NAND gate delay. It is also important to keep the RO short so that fine spatial resolution can be achieved. If the RO oscillates too fast to propagate through the multiplexers to the output pad, a local divide-by-two circuit can be used to reduce the frequency. Wide transistors are used to make variations less sensitive to fluctuations in transistor gate widths.

4.2.2 Ring Oscillators with Different Number of Stages

A set of test structures in the 90nm test-chip was designed to characterize correlation of delays between inverters that are connected to each other in a RO. This test structure consists of ROs with different number of stages (Figure 4.16). Details will be given in Section 6.1.

The delay and spatial correlation of individual gates can also be obtained using

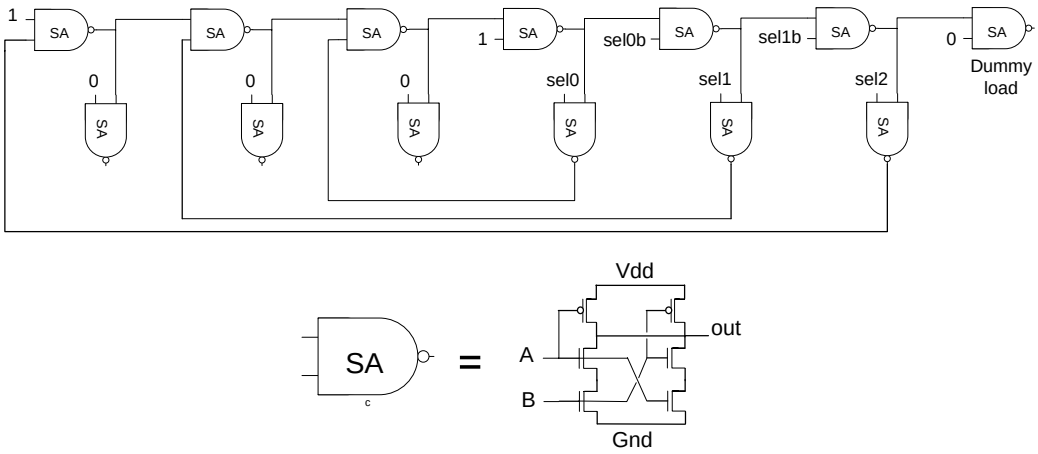


Figure 4.20: RO with programmable number of stages using NAND gates

a RO with programmable number of stages shown in Figure 4.20. This RO is made up of NAND gates and by applying the correct control signals to the NAND gates, RO of different lengths are generated. The gates of the longer RO will contain the gates of the shorter RO and the difference in delay will be due to the extra gates that are added plus the mismatch of one NAND gate. By taking the difference in delay, the delays of the extra gates can be obtained. Since these are placed next to each other, high resolution spatial correlation of gate delays can be measured.

4.3 Summary

In summary, circuits to measure CMOS performance accurately and rapidly have been proposed. Measurement of leakage current and RO frequency of an array of devices have been described in detail and implemented in a 90nm and a 45nm test-chip. RO circuits with different number of stages to characterize spatial correlation of gate delays have been presented and but have not been implemented. Future work consists of implementing these

circuits.

Chapter 5

Measuring SRAM Performance Variability

Measurement of SRAM cell noise margins and dynamic performance in large arrays with minimum perturbation of the layout is necessary in order to accurately characterize SRAM variations. This chapter proposes several circuit ideas to measure both static noise margins(Section 5.1) and the read and write access times of individual SRAM cells with minimum or no perturbation of the array layout [86]. Section 5.2 describes a ring-oscillator circuit whose frequency correlates with the write access time of the SRAM cell, and Section 5.3 describes a pulse ring-oscillator circuit whose frequency correlates with the read access time of the SRAM cell. Section 5.4 describes how the access times for read and write can be measured by adjusting the wordline pulse width and measuring it. Finally, Section 5.5 summarizes this chapter.

5.1 SRAM Static Noise Margins Measurement

SRAM cell read and write margins are important performance parameters [87]. Traditional noise margin measurements are obtained from butterfly curves or N-curves. These require measurement of the internal storage nodes of the SRAM cell which, in simulations, is easily done but in silicon, requires perturbing the layout. Instead, the IV characteristics of the bitlines can be measured and used to deduce read and write margins as well as read and write strength of the cell [86, 88][1].

Figure 5.1 shows the schematic of a SRAM column and the switches connected to the bitlines that will allow us to measure the bitline current for given voltages. A plot of the bitline IV curve is given in figure 5.2 and figure 5.3 showing indicators for the write margin and read margins respectively. The write margin IV plot is obtained by sweeping V_L from VDD to 0V and measuring the current I_L with a '1' state stored in the left node of the cell. I_L drops initially and rises suddenly when the SRAM cell flips. The amount of current required to flip the cell and the voltage at which this occurs can be used as the write margin indicators. The read margin IV plot is obtained by sweeping V_L from 0V to VDD and measuring I_L with a '0' state stored in the left node of the cell. The current when $V_L = VDD$ and the slope at this point can be used as read margin indicators. This circuit has been implemented in [88]. Using various ways of sweeping supply, wordline and bitline voltages and measuring the bitline current, different noise margin parameters are obtained and shown to correlate strongly with the tradition noise margins obtained from butterfly curves.

[1]This work was done in collaboration with Zheng Guo, U.C. Berkeley.

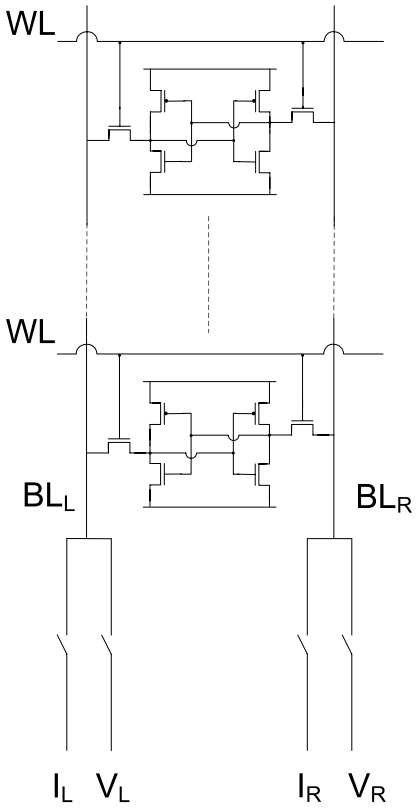


Figure 5.1: A SRAM column showing the bitline switches.

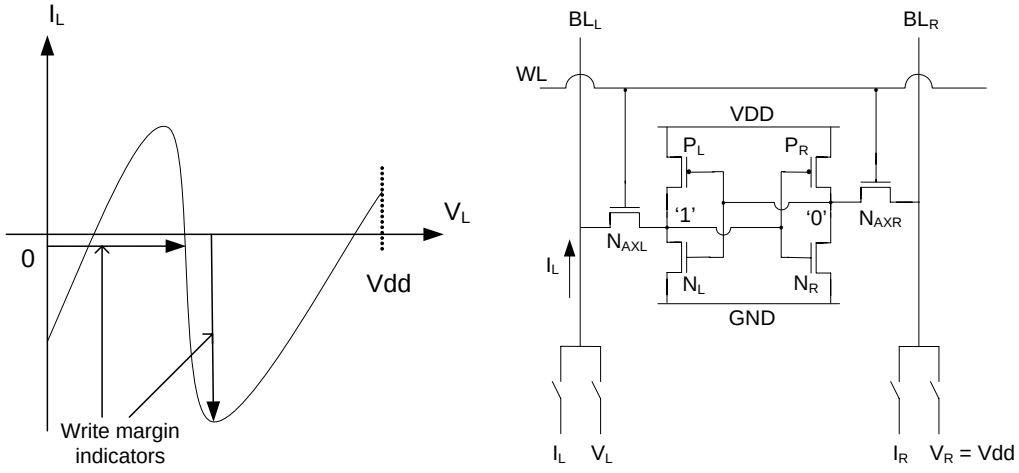


Figure 5.2: Bitline IV plot showing write margin indicators

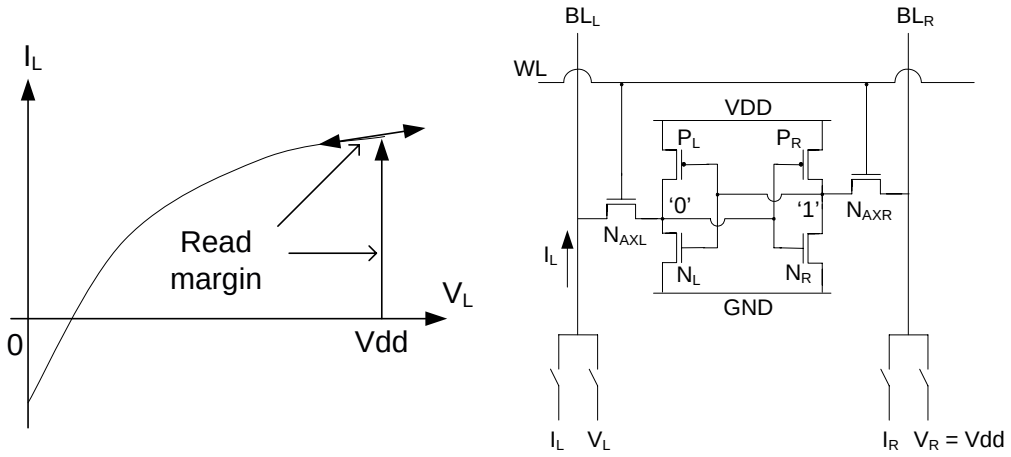


Figure 5.3: Bitline IV plot showing read margin indicators

5.2 SRAM Dynamic Write Performance Measurement

Static noise margins of the SRAM cell are useful for verifying functionality. However, the more important performance metrics are the read and write access time as these determine if the cell works during real-time operations. The read access time is the time required to perform a read, and the write access time is that required to perform a write. These dynamic parameters depend both on the SRAM cell strengths as well as the capacitances and leakage currents of the cell and the bitline which cannot be obtained through static measurement alone. As mentioned in chapter 1, [30] has shown that by connecting SRAM columns together to form a ring, the oscillation frequency thus obtained can be used to evaluate the average dynamic performance of an SRAM array. Based on similar ideas, two ring oscillators have been designed that will allow for evaluating the read and write access times of SRAM cells in an array with minimum perturbation of the layout.

Figure 5.4 shows the RO for characterizing write access time. The differential inverter forces the internal storage nodes of the SRAM cell that is being accessed to switch

at each cycle. The SRAM cell acts as a load to the RO. This changes the RO frequency which is divided down and measured. The pull-down strength of the differential inverter is controlled by Vbias. It is adjusted such that pull-down is strong enough to flip the SRAM cells across all corners but sufficiently weak such that the delay in flipping the SRAM cell is a significant portion of the total delay through the RO. In that case the RO will be sensitive to variations in the write access time of the cell. The switching threshold of the differential inverter that is driven by the bitlines should be tuned to be low so that the bitline delay reflects the writing of the SRAM cell. The number of stages should be small for increased sensitivity to the bitline delay. The RO is connected to the column bitlines of the SRAM array through addressable switches that allow to select any particular column. The write strength obtained is an average of the write "0" and write "1"' strength of the SRAM cell. The RO frequency is first measured with all wordlines turned off to measure the intrinsic frequency of the RO due to the bitline load only. The wordline of the SRAM cell under test is then turned on and a slower oscillation frequency is obtained. The difference between the two frequencies gives a measure of the average write access time of the SRAM cell.

5.3 SRAM Dynamic Read Performance Measurement

Figure 5.5 shows the RO for read access time measurement. A pulse is propagated through the ring and regenerated at each cycle. The pulse width is reduced due to the slow pull-down of the bitline (blc), enlarged by the pulse widening circuit and regenerated by the pulse generator. The oscillation frequency will vary as a function of the pull-down strength of the SRAM cell. The RO frequency is first measured with all wordlines turned

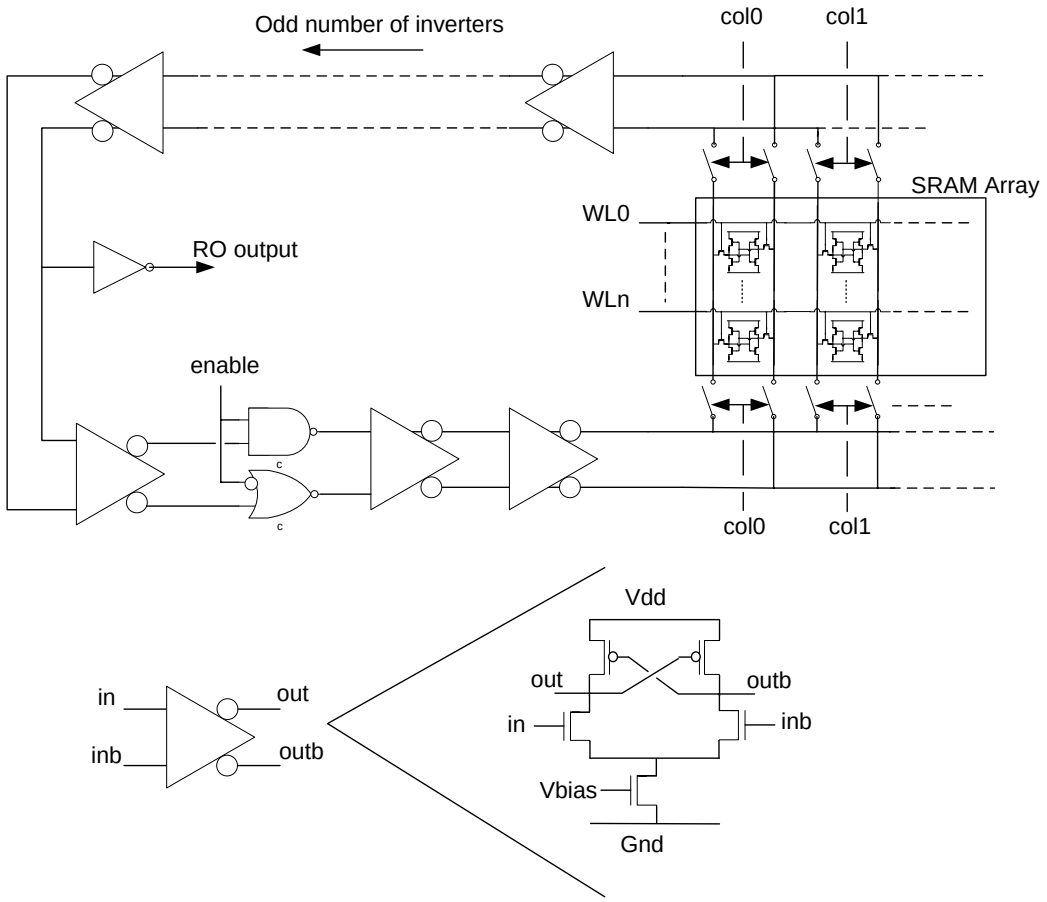


Figure 5.4: RO to measure write access time and the differential buffer used in the RO.

off. The bitline will be pulled down by the NMOS stack whose strength is controlled by voltage Vctr. The wordline of the SRAM cell under test is then turned on and a faster oscillation frequency is obtained as the NMOS pull-down stack is now aided by the SRAM cell. The difference between the 2 frequencies gives a measure of the read access time of the SRAM cell. Several knobs are necessary in this circuit to ensure the oscillations occur. In the pulse widening circuits, Vc1 and Vc2 are used to control the amount of pulse widening. The pulse has to be widened sufficiently so that the pulse generation circuits can regenerate a pulse. Next, in the pulse generation circuits, the pulse width of the generated pulse can be adjusted with Vc3 and Vc4. This pulse must be sufficiently large so that the bitline has sufficient time to pull-down the voltage to 0V. These voltage controlled buffers can be implemented as shown in Figure 5.6. If Vc2 is low to weaken the pull-down of the second stage, the falling edge of the pulse will be slowed down resulting in a widening of the pulse. Also, both Vc1 and Vc2 can be used to weaken both PMOS and NMOS of the second stage in order to increase the delay in the pulse generation circuit. This will increase the width of the regenerated pulse. Finally Vctr controls the pull-down strength of the NMOS stack. This pull-down strength must be weak relative to the SRAM pull-down strength so that the delay in pulling down the bitline is more sensitive to the SRAM cell than the NMOS stack. Overall the tuning of the oscillator has to ensure that oscillations occur and that the frequency of oscillations is sensitive to the pull-down strength of the SRAM cell. The single ended RO is connected to the column bitlines of the SRAM array through addressable switches.

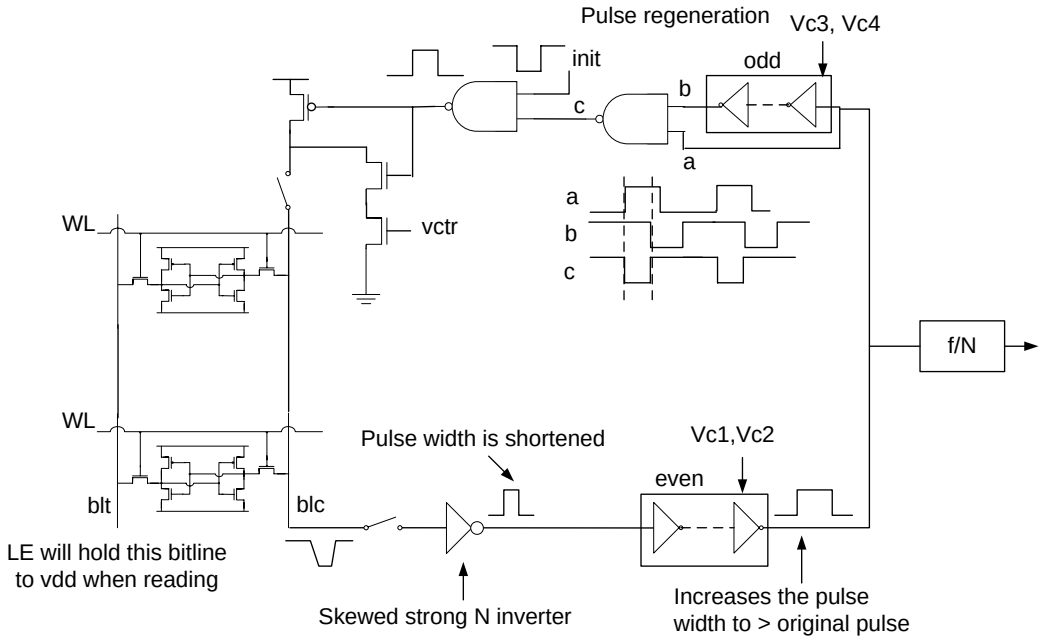


Figure 5.5: Pulse RO for characterizing cell read timing.

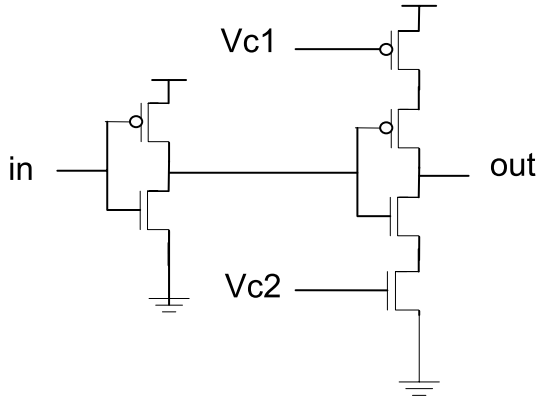


Figure 5.6: Current starved buffers for widening the pulse width and creating delay for the regenerated pulse.

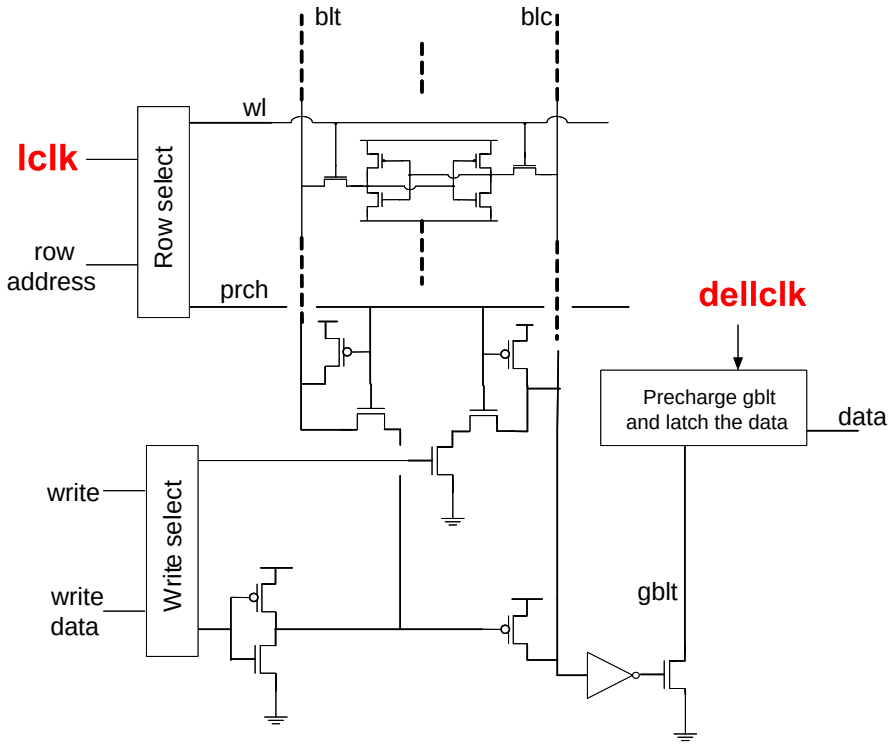


Figure 5.7: SRAM access time measurement

5.4 SRAM Access Times Measurement

Figure 5.7 illustrates how a SRAM in a 45nm SOI technology works [89]. The local bitline (*blc*) will trigger the global bitline (*gblt*) and the data read is latched. *lclk* generates the wordline pulse (*wl*) and local precharge (*prch*) signals whereas *dellclk* generates the global bitline precharge and data-latching signals. In that way, *lclk* and *dellclk* determine several important timing parameters of the SRAM. The minimum wordline pulse width for correct operation is equivalent to the access times of the SRAM cell. This pulse width is the same as the *lclk* pulse width. By generating a programmable *lclk* pulse width and measuring the minimum width for correct operation a measure of the dynamic performance of the SRAM cell can be obtained.

5.5 Summary

In summary, circuits to measure SRAM cell variability accurately and rapidly with minimum perturbation to the array layout have been proposed. However, they have either not been implemented or not been measured yet. Future work consists of implementing these circuits.

Chapter 6

Measurement Results and Analysis

Two test-chips, the first in an early 90nm process and the second in an early 45nm process, have been designed and measured. These chips contain the ring-oscillator and leakage transistor arrays and the leakage current measurement circuit described in chapter 4. Measured data show several important trends which are analyzed in detail in this chapter. The impact of layouts on transistor performance and the statistics of transistor D2D and WID variations has been measured and characterized.

In this chapter, section 6.1 will describe the layout structures that were studied and the circuits in the test-chips. Several properties of the 45nm process will also be presented. Section 6.2 will quantify the impacts of the different layout configurations and provide analysis of the causes of these systematic variations. Section 6.3 will describe the measurement results for D2D and WID variations. Spatial correlation of RO delays in the test-chips will be analyzed. Section 6.3.3 describes the measurement result of an experiment in which an array of ring oscillators with different number of stages was measured in the

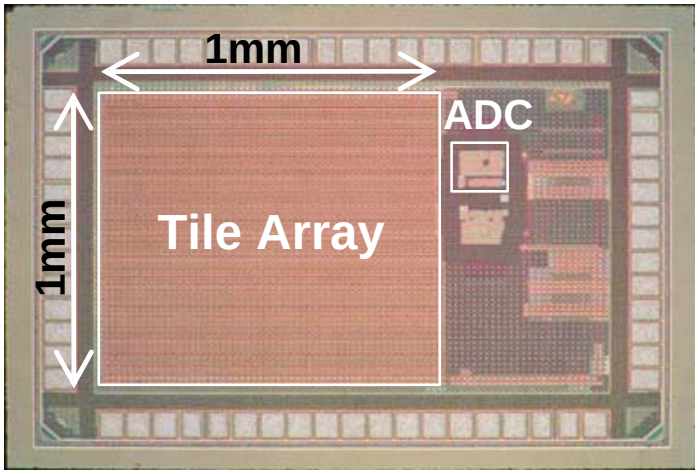


Figure 6.1: Die photo of the 90nm test-chip

90nm test-chip in order to study the spatial correlation of closely spaced gates in the RO. Section 6.4 describe the process of extracting process parameters from measurement in the 90nm process. Finally, section 6.5 summarizes the results of the measurement.

6.1 Layout Structures and Test-Chips Overview

A test-chip has been implemented in a general-purpose 90nm CMOS technology to evaluate the effects of lithography-induced variations, and to measure WID and D2D variations, as well as WID spatial correlation. This is done by measuring ring-oscillator (RO) frequencies and transistor source-drain leakage currents (I_{LEAK}) of an array of test-structures [5]. These circuits were described in Sections 4.1 and 4.2.

The chip contains 10 × 16 tiles, with a total array area of 1mm × 1mm (Figure 6.1). The tiles are separated horizontally by 62.5um and vertically by 100um. Each tile has twelve 13-stage ROs and twelve transistors in the off-state, each with a different layout (Figure

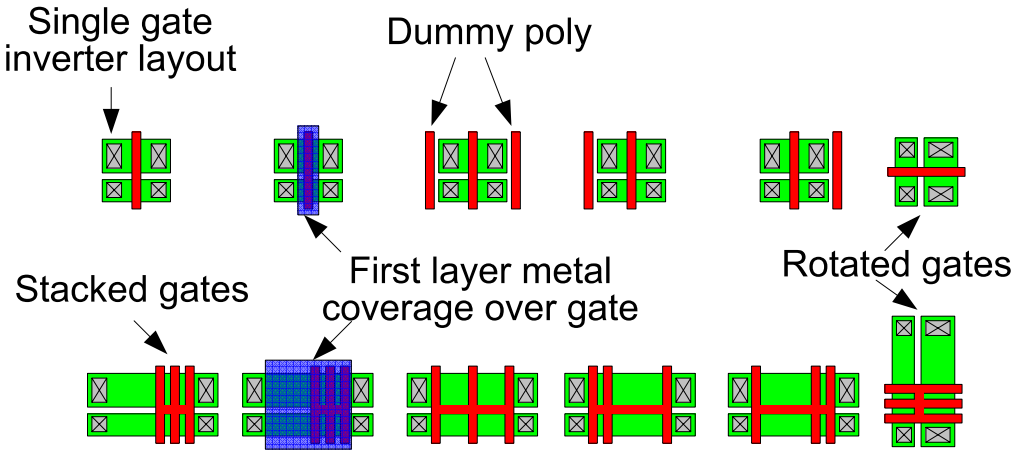


Figure 6.2: 12 layout structures in 90nm

6.2). Some transistors in both the ROs and those in the off-state consist of a single poly-Si finger, while the rest are constructed with a stack of three fingers. The poly-Si pitch of neighboring dummy poly-Si lines is varied in the test structures to capture proximity effects. Poly-Si orientations, together with the properties of the two-dimensional tile array, are used to characterize spatial correlation. Rotated poly-Si ROs are obtained by rotating standard ROs by 90^o. Non-symmetrical structures and their mirror image target measurement of the coma effect. The first layer metal coverage over certain gates are also varied in the layout to investigate the effects of anneal [70]. The chip also contains the leakage current measurement circuit described in section 4.1.2. A 2.5V analog supply was used together with thick oxide transistors to implement the analog components in this circuit.

The 90nm test-chip also contains RO of different number of stages for characterizing the correlation of delays between inverters that are connected to each other (mentioned in Section 4.2.2). Using a simple model of spatial correlation coefficient (ρ) shown in Figure 6.3, the relationship between the variance of the ROs (s_D^2) and the number of stages (n)

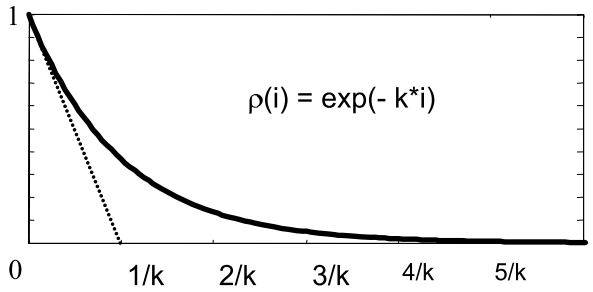


Figure 6.3: Model of spatial correlation coefficient. Plot of correlation coefficient ρ vs spacing i. k is a model parameter.

is obtained (Equation 6.1). The unknown parameters are the variance of the delay of a single inverter (s^2) and the exponential constant that characterizes the spatial correlation coefficient (k). By measuring s_D^2 for ROs of known n, the measurement data can be fitted to the equation to obtain s^2 and k.

$$s_D^2 = n \cdot s^2 + 2 \cdot s^2 \cdot \sum_{i=1}^{n-1} (n-i) \cdot e^{-k \cdot i} \tag{6.1}$$

where

i = spacing, in this case the spacing between adjacent inverters $\approx 2.1 \mu m$

k = parameter that determines the characteristics of the correlation coefficient model

s_D = standard deviation of the RO

s = standard deviation of the delay of a single inverter

n = number of stages of the RO

This test structure is implemented as an array of VRO tile, each of which contains 8 ROs with 13, 19, 25, 31, 37, 43, 49 and 55 number of stages arranged as shown in Figure 6.4. The transistors have their gates oriented vertically and the gates are spaced horizontally.

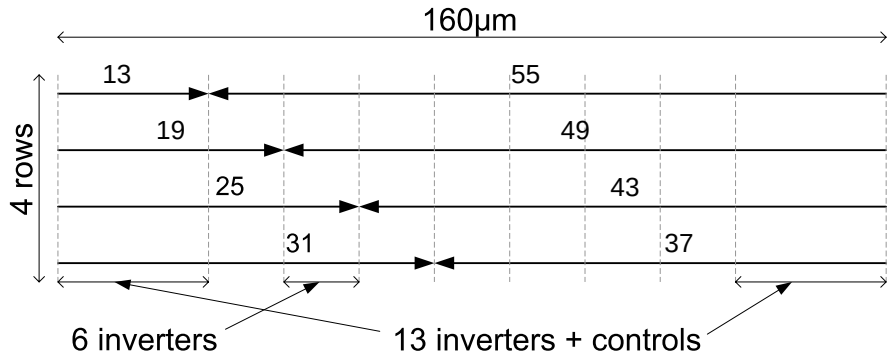


Figure 6.4: Arrangement of ROs of different lengths.

In the layout, 4 RO and off-transistor tiles combine with one VRO tile to form a bigtile that is stacked to form the 10×4 bigtile array shown in Figure 6.5. A simple exponential model of spatial correlation coefficient (ρ) shown in Figure 6.3 will be used to derive the relationship between the variance of the ROs (s_D^2) and the number of stages (n) (Equation 6.1). The unknown parameters s^2 and k will be derived by fitting the measurement results to the model equation.

The chip also contains frequency dividers and a few other control logic. The left side of the array contains a vertical column of shift registers that supply the row address bits and the bottom of the array contains a horizontal row of shift registers that supply the column address bits. Both the input and output of each shift register chain are connected to I/O pads in order to ensure that they work properly during debugging. The tiles are identical and are abutted to each other in the layout to form the array.

The methodology used to characterize variations in the 90nm process was also applied to a 45nm low power strained-Si process. A 45nm test-chip containing an array of ROs and both NMOS and PMOS leakage transistors was used to characterize the impact of layout on transistor performance and measure D2D and WID variability. In addition,

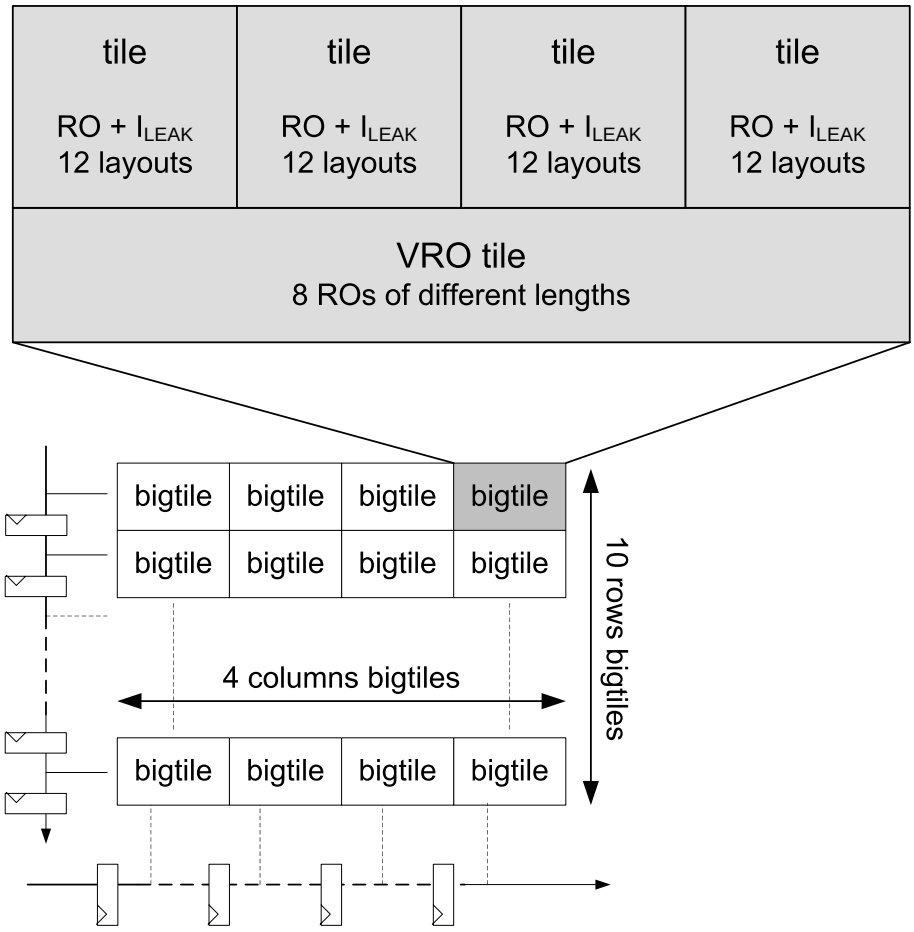


Figure 6.5: Layout of a bigtile and the 10×4 array of bigtile in the 90nm test-chip.

PROCESS FEATURE	45NM PROCESS	EFFECT
Si substrate	<100>- oriented channel	Higher PMOS mobility
Shallow trench isolation (STI)	Sub-atmospheric deposited oxide	Reduce STI stress
Contact etch stop layer (CESL)	Nitride layer create high tensile strain	Higher NMOS mobility
Immersion	NA >1	Improved resolution
Backend dielectric	Low k 2.5	Low RC delay

Table 6.1: Summary of the 45nm process. [2, 3, 4]

this chip contains SRAM structures that allow for the measurement of IV characteristics of bitlines in large SRAM arrays as described in Section 5.1.

Table 6.1 summarizes several techniques used in the low-power 45nm process [2, 3, 4]. Transistor channels are oriented in the <100>direction, which increases PMOS transistor mobility and makes it insensitive to stress. The use of sub-atmospheric chemical vapor deposition oxide (SACVD) for trench isolation further reduces stress effects. Instead of a strong compressive strain, these trenches now exert a weak tensile strain on the transistors. Strong uniaxial tensile strain is created by the nitride layer in order to increase NMOS mobility. Resolution is enhanced with immersion lithography and low-k dielectric is used for the copper interconnects. Strain induced by the shallow trench isolation (STI) and the contact etch stop layer (CESL) nitride are layout dependent and are investigated in the test chip.

In the 45nm process, many strict design rules have been introduced. Most notable is the rule on minimum length poly-Si gate orientation and spacing. Orientation of gates are now fixed and only a small continuous range followed by a discrete jump in poly-Si spacing is allowed. Taking into account the new rules and the strained process, 17 layout

styles were created to study the effects of layout. These are shown in Figure 6.6 and 6.7 and described in Table 6.2. Layouts P1, P2, P3, P4 vary the spacing of the poly-Si nearest to the transistor's gate. Measurement results from the 90nm test-chip suggest that proximity effects could dominate again. Having a range of poly-Si pitch would improve the analysis. The significance of proximity effects in the 90nm test-chip also led to the structures SP1, SP2, SP3 which vary the distance of the poly-Si that is the second nearest neighbor to the gate. S1, S2 are layouts that are symmetrical to each other. Optical aberrations can cause these structures to print differently, giving rise to different gate-lengths. D1 has a longer source/drain (S/D) diffusion area than P3, which has been observed to cause different strain in a transistor [90]. M1 has metal-2 coverage over its gate which has been shown to cause different annealing temperatures [70]. Transistors with metal covered gate tend to experience less annealing, resulting in increase V_{th} and reduced I_{LEAK}. T1 has neighboring poly-Si at the ends of its gate which could cause the gate poly-Si to print differently at the edges. R1, R2 and R3 have regular poly-Si pitches that vary from minimum to maximum. The benefits of regular poly-Si pitch has been observed in lithography but not measured in circuits. These structures permit to measure the impact of regular pitches and pitch spacing on variability. NSTI is the same as R1 except that there is no STI and isolation is achieved by turning off the adjacent transistors [91]. STI causes strain in the substrate which changes transistor mobility. It also creates unevenness on the surface of the substrate, resulting in systematic changes in transistor properties. Finally, in layout V1 the inverters of a RO are placed in the vertical direction instead of the horizontal direction as illustrated in Figure 6.7. In the 90nm test-chip, variations were dependent on the horizontal and vertical

Layouts	Targeted Effect
P1,P2,P3,P4	Primary proximity
SP1,SP2,SP3	Secondary proximity
S1,S2	Symmetry
D1	Larger S/D area
M1	Metal coverage over gate
T1	Poly-Si st extremity of gate
R1,R2,R3	Regular pitch
NSTI	No STI
V1	Vertically placed gates

Table 6.2: Characteristics of the layout configurations.

directions. Even though 90° gate rotation are not allowed by the design rules in this 45nm process, certain properties of the two orthogonal directions can be investigated with these structures.

The die photo of the 45nm test-chip is shown in Figure 6.8. The array contains 18 x 16 tiles, each tile contains 17 ROs and 17 NMOS and PMOS transistors with $V_{GS} = 0$ in each of the 17 layouts. In the RO array, a local divide-by-2 circuit within each RO allows for the use of small number of stages by reducing the frequency of the signal that is multiplexed out. The leakage NMOS and PMOS transistors have their gate connected to GND and VDD respectively so that VGS=0 only when that particular transistor is selected. Details on the circuit implementation of RO frequency and I_{LEAK} measurement can be found in Section 4.1 and 4.2.1.

In the 90nm test-chip, only NMOS leakage current is measured. Each NMOS transistor has a width of $1.92\mu m$ and is made up of 4 smaller transistors in parallel, each small transistor has a width of $0.48\mu m$. Transistor width is chosen to have sufficient I_{LEAK} such that measurement accuracy is not compromised. As the precision target is 1nA, gate widths are chosen to give I_{LEAK} of $\approx$ 10nA in the slow corner. Individual gate fingers are

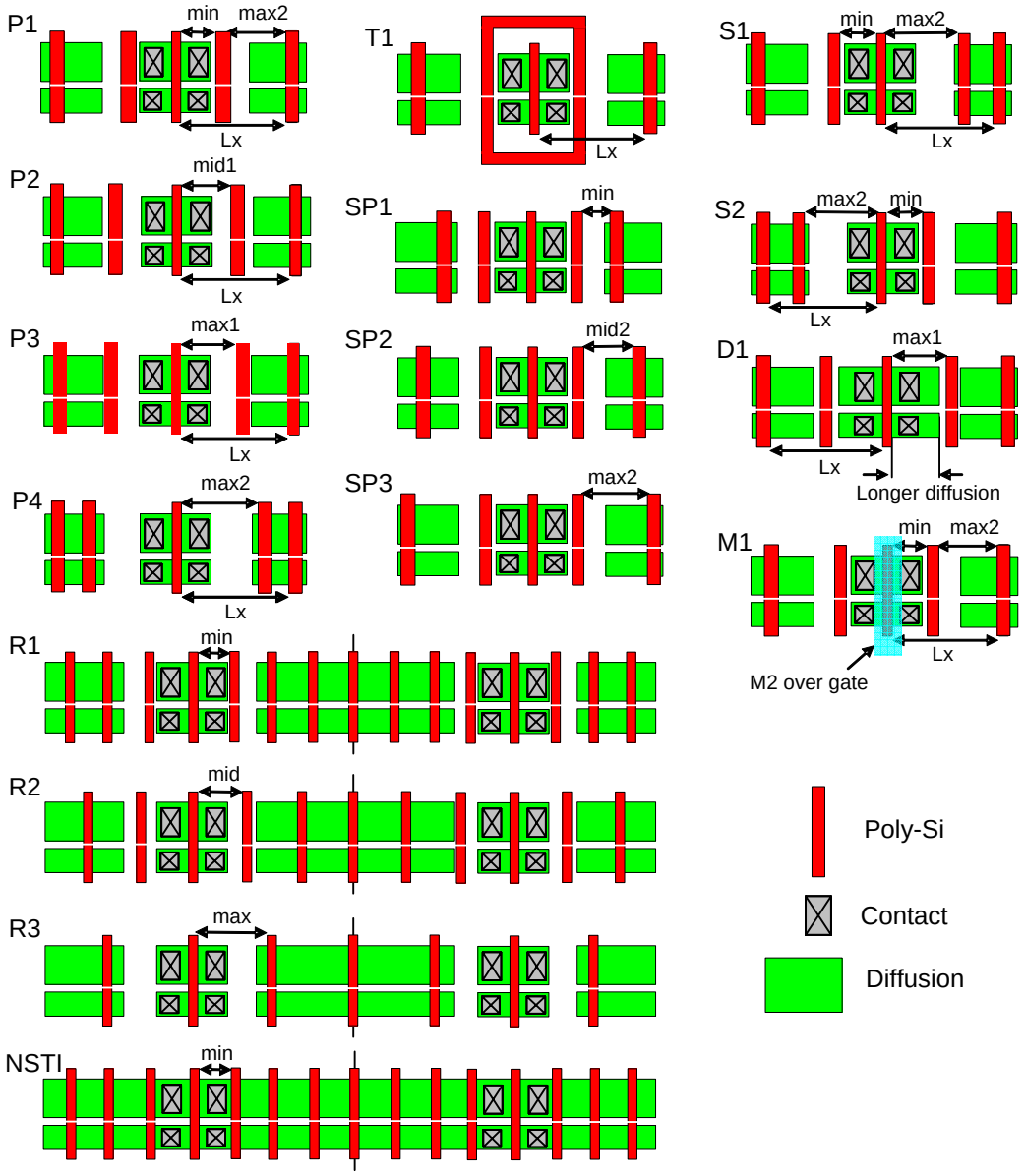


Figure 6.6: 17 layout configurations in 45nm

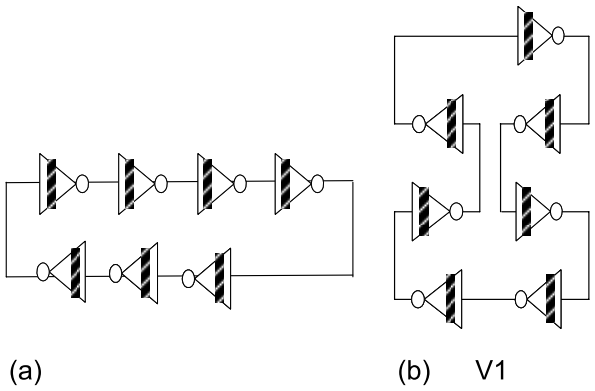


Figure 6.7: Illustrates (a)horizontally vs (b)vertically placed inverters in a RO.

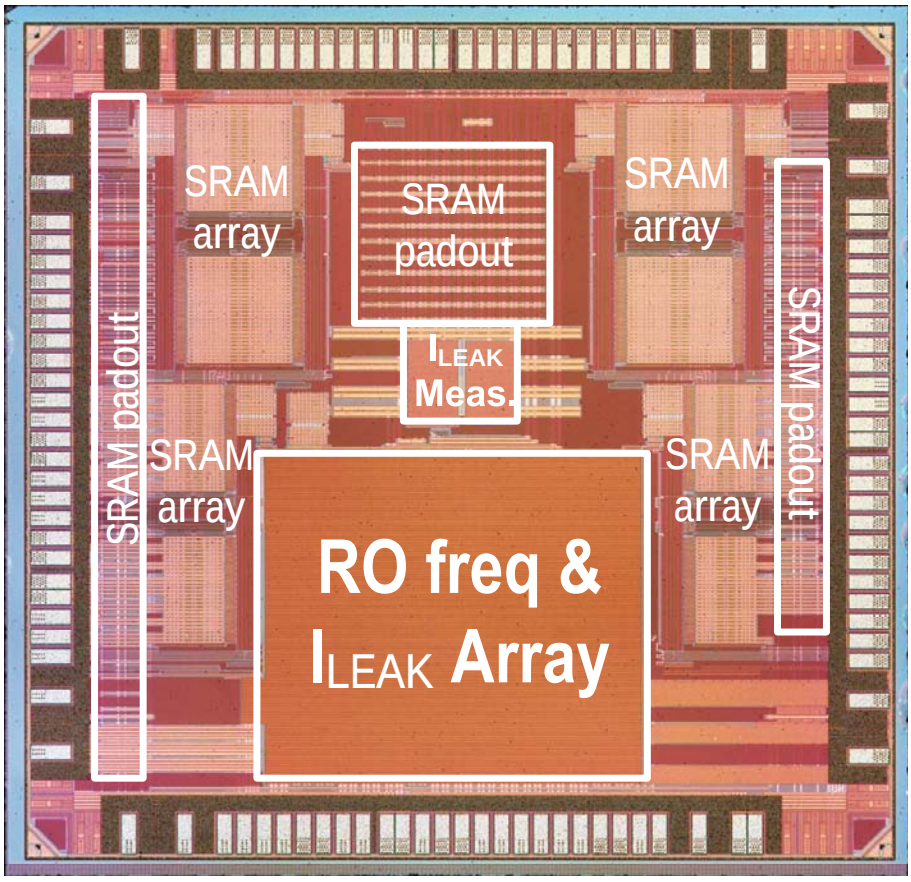


Figure 6.8: Die photo of the 45nm test-chip

Test-chip	MOS	Total Width	No. Fingers	Finger Width
90nm	NMOS	$1.92\mu(m)$	4	$0.48\mu(m)$
45nm	NMOS	$3.9\mu(m)$	13	$0.3\mu(m)$
	PMOS	$5.46\mu(m)$	13	$0.42\mu(m)$

Table 6.3: Dimensions of the transistors in the I_{LEAK} arrays.

Test-chip	Gate-stacks	No. Stages	PMOS	NMOS
90nm	single gate	13	$1\mu(m)$	$0.48\mu(m)$
	3 gate stack	7	$1\mu(m)$	$0.48\mu(m)$
45nm	single gate	13	$0.42\mu(m)$	$0.3\mu(m)$

Table 6.4: Dimensions of the transistors in the inverters of the ROs.

chosen to be much longer than the lengths in order to avoid variation in gate widths. In the 45nm test-chip, both NMOS and PMOS leakage currents are measured. The total widths of each PMOS and NMOS transistor measured are $5.46\mu m$ and $3.9\mu m$ respectively. Each transistor consists of 13 smaller transistors in parallel, each with a width of $0.42\mu m$ and $0.3\mu m$ for PMOS and NMOS respectively.

In both 90nm and 45nm test-chips, 13-stage ROs are used for the single finger transistor layouts. For the layouts with a stack of 3 gates in the 90nm test-chip, 7-stage ROs are used. Section 4.2.1 explains how the number of stages is chosen. A shorter RO achieves good spatial resolution but longer RO is necessary for good sensitivity to the inverter stages and for the signal to propagate through the multiplexers and dividers. The widths of the 90nm transistors in the inverter are 1um and 0.48um for PMOS and NMOS respectively. In the 45nm test-chip, they are 0.42um and 0.3um for PMOS and NMOS respectively.

The dimensions of the transistors in the arrays are summarized in Table 6.3 and Table 6.4.

In the analysis of the 90nm measurement results, tiles at the perimeter of the array are ignored to eliminate edge effects. The ROs with single isolated poly fingers often exceed the fast corner and are excluded in the measurements. In the 45nm test-chip, no edge effects were observed and hence all the tiles were used in the analysis. All the RO and leakage transistors were analyzed.

To distinguish systematic from the random effects, the results for different layouts are compared for each die from die to die. Measurements for identical structures within a die are compared to obtain WID correlations. Averages for various dies are used to indicate the D2D spread. Averages across several dies are used to analyze the mask writing effects.

6.2 Effects of Layout on Frequency and Leakage

In order to investigate the systematic layout effects, the RO frequency and I_{LEAK} distribution of different layout structures are compared. In the following analysis, the mean value for each layout configuration for each die is compared to observe systematic effects. RO frequencies are normalized to the nearest corner in order to remove the differences in parasitics that are captured by the layout extraction when these differences are significant. This is the case for all RO frequency measurements in 45nm and the stacked-gates RO frequency measurement in 90nm. Leakage currents are not aligned with the corners as they are independent of parasitics. The impact of layout on random variations is studied by comparing the standard deviation over mean (σ/μ) of the measurement data.

Measurement results show systematic variations due to layout that are not captured by the layout extraction tool. The impact of layout on performance is more significant

in the 90nm test-chip than in the 45nm test-chip. Proximity of poly-Si causes more than 10% shift in RO frequency (ΔF = 10%) in 90nm and around 2% shift in 45nm. The most significant effects in 45nm were due to a larger S/D area (ΔF = 5%) and the removal of STI (ΔF = 3%).

Possible causes of these systematic effects are proposed through analysis of the measurement results. Accurate analysis in the 45nm results is difficult as the variations are relatively small.

6.2.1 Proximity of Poly-Si Gates

Figures 6.9 and 6.10 plot the distribution of normalized RO frequency and normalized leakage current in the log domain respectively for the 90nm test-chips. Each colored histogram represents the distribution for a die and the overall distribution is plotted as a continuous blue curve. 90nm measured results are obtained from 36 dies located on half a wafer. Only the histograms of the fastest and slowest dies are shown. Vertical lines labeled TT and FF represent simulation results from the extracted layout for TT and FF corners.

These plots show that proximity effects due to different poly-Si gate pitch cause a shift in frequency of over 10%, and a 20x shift in I_{LEAK}. More isolated gates have shorter gate lengths and hence higher RO frequencies and higher leakage currents. This is much larger than the 1.1% frequency shift predicted by SPICE simulations of the extracted layout, which capture only changes in parasitic capacitances. D2D and WID leakage current variations are reduced with increased poly-Si density. There is also a similar but smaller effect of poly-Si density on stacked gates. Figure 6.11a plots the RO frequency distribution for stacked gates whereby the frequency is normalized to a fixed constant. Extracted

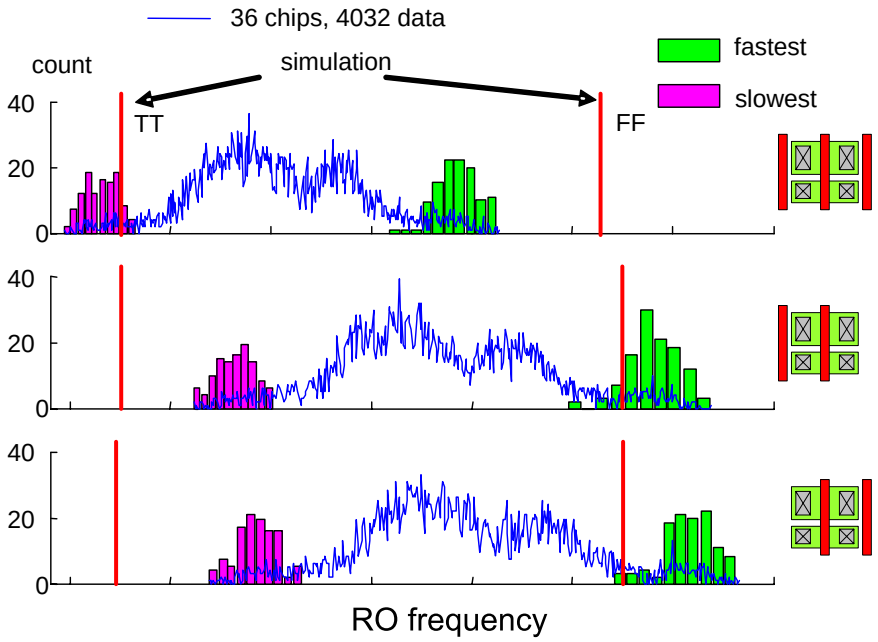


Figure 6.9: Frequency distribution for single-finger configurations in the 90nm test-chip. Vertical lines correspond to typical and fast corner simulation results. Bar plots correspond to the WID distribution of the fastest and slowest chip.

simulation results from the TT and FF corners show that RO frequencies for the different gate densities differ due to parasitics that is captured by the layout extraction. In order to show the proximity effects that are not captured by the layout extraction, the measured frequencies are normalized to the TT corner and the distributions are plotted in Figure 6.11b. These plots show that the frequency of the densest stacked gate configuration is 5% slower than the fastest layout.

Figure 6.12 shows the distributions for 4 layouts with different poly-Si gate pitches from the 45nm test-chips. Maximum systematic shift in frequency is $\approx$ 2%. Leakage currents also experience small systematic shifts. This effect could be due to a combination of small gate-length(L) variations and layout dependent variation of the strain caused by the CESL. PMOS transistors have a sharper VT roll-off and hence PMOS leakage currents

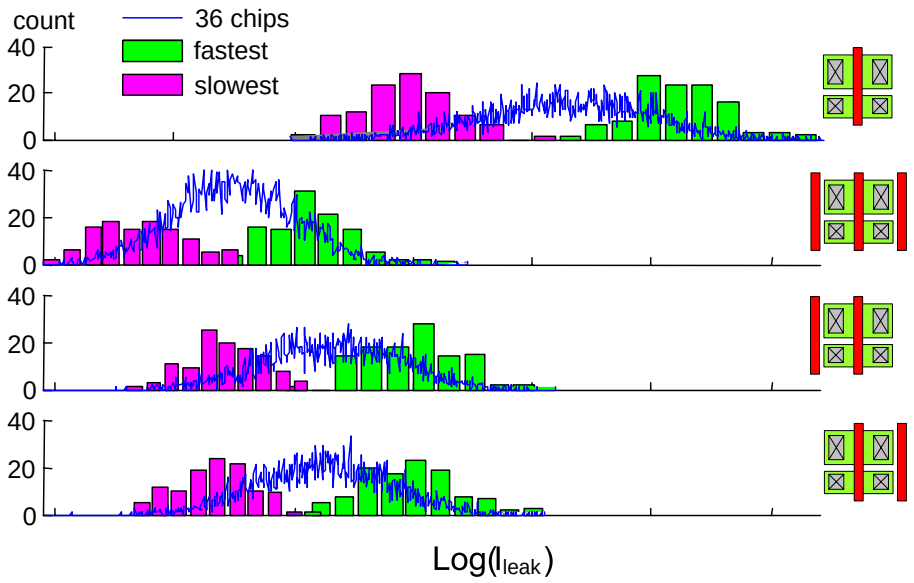


Figure 6.10: $Log(I_{LEAK})$ distribution for single-finger configuration in the 90nm test-chip.

are more sensitive to L variation. An isolated gate will generally experience more strain from the CESL than a poly-Si gate with dense pitch. However, only NMOS is affected by stress in this process.

In the plot of normalized PMOS $logI_{LEAK}$, poly density induced L variation is observed. The P2 layout on the second row likely has a shorter gate length than the others resulting in increased PMOS leakage. The effect on NMOS leakage is weaker and is completely compensated by the CESL stress. As the poly-Si pitch increases, more tensile stress is applied, increasing the mobility of NMOS transistors and raising the amount of NMOS leakage thereby offsetting the effect of increased gate length. PMOS leakage is not affected since it is insensitive to stress in a <100>-oriented channel. At the same time, RO frequency also increases and this offsets the effect of a longer L.

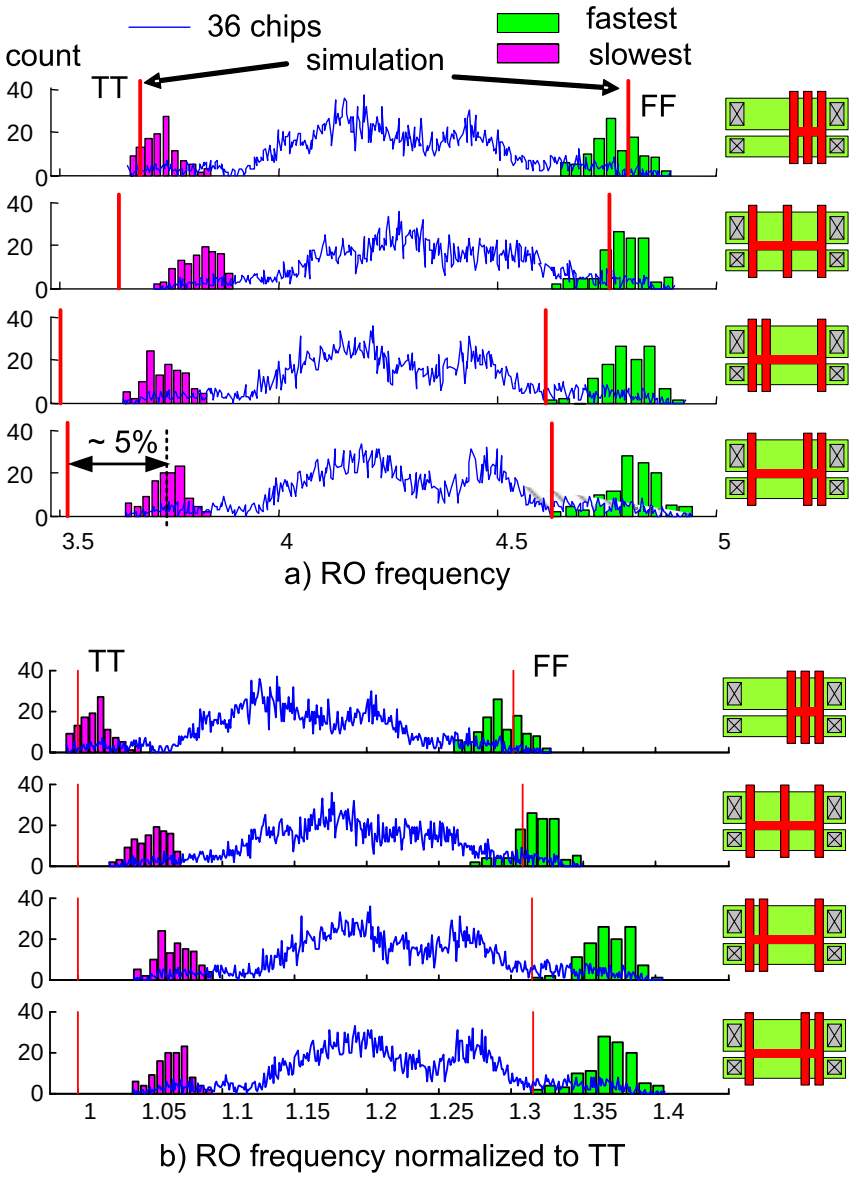


Figure 6.11: RO frequency distribution for stacked gates plotted next to its corresponding layout in the 90nm test-chip. Vertical lines represent simulation corners from extracted layout. Distribution plots of a) RO frequency normalized to a constant, b) RO frequency normalized to its corresponding TT simulation corner. In b), less dense gates show a 5% increase in frequency compared to dense gates.

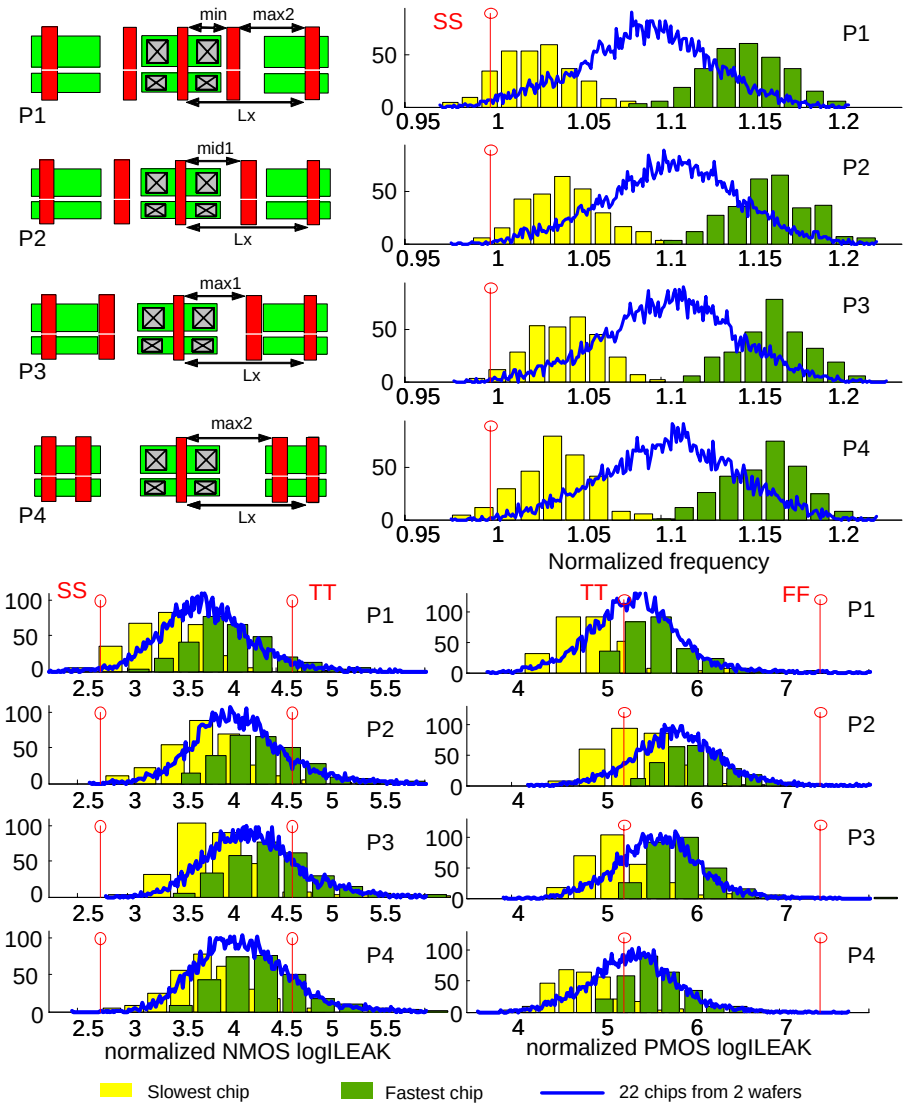


Figure 6.12: Effect of nearest neighbor poly-Si pitch on RO frequency and transistor leakage currents in the 45nm test-chip. Plots of RO frequency distribution normalized to the SS corners, NMOS $log(I_{LEAK})$ distribution and PMOS $log(I_{LEAK})$ distribution for 22 dies.

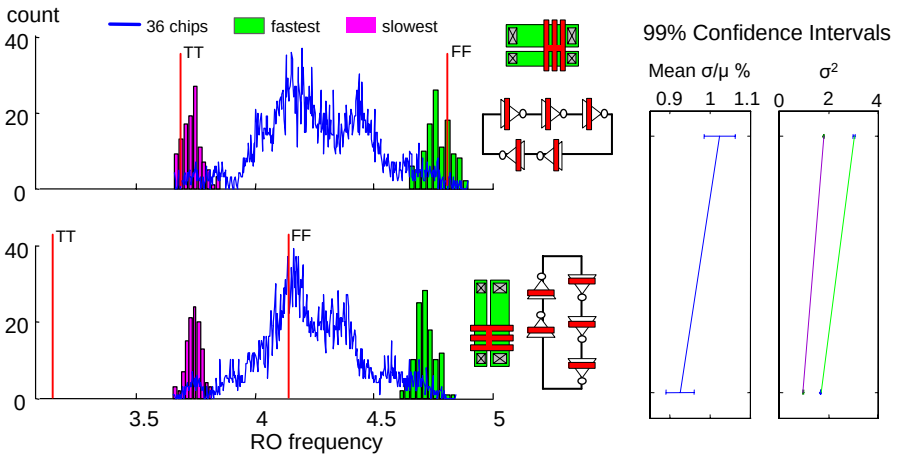


Figure 6.13: Bottom left plot shows the frequency distribution for 90° rotated ROs with horizontal gates in the 90nm test-chip. Non-rotated ROs in the top left plot show more variation. 99% confidence interval of σ/μ and σ^2 shown on the right verify that the top plot has more variations.

6.2.2 Gate Rotation

Figure 6.13 shows the impact of rotating a RO by 90° in the 90nm test-chip for stacked gates layout. All the layers below metal-2 were rotated in order to keep the parasitics as similar as possible. Variation ($3\sigma/\mu$) is smaller by $\approx 0.4\%$ for rotated gates across all chips. This is verified with a high level of confidence as shown in the confidence intervals plots. The mean value of frequency stays around the same in the measurement. However, in the simulation corners, the RO frequency is significantly slower for the rotated layout. Leakage current of single gate layout for the non-rotated and the rotated cases is shown in Figure 6.14 and verified to be similar, with the rotated layout having less leakage. This shows that there is no significant difference in the properties of the transistors in the two layouts. The difference in the simulation results is likely due to an overestimation of the parasitics in the rotated layout due to the higher level metals by the layout extraction tools.

Figure 6.15 shows the impact of placing the gates horizontally vs vertically in the

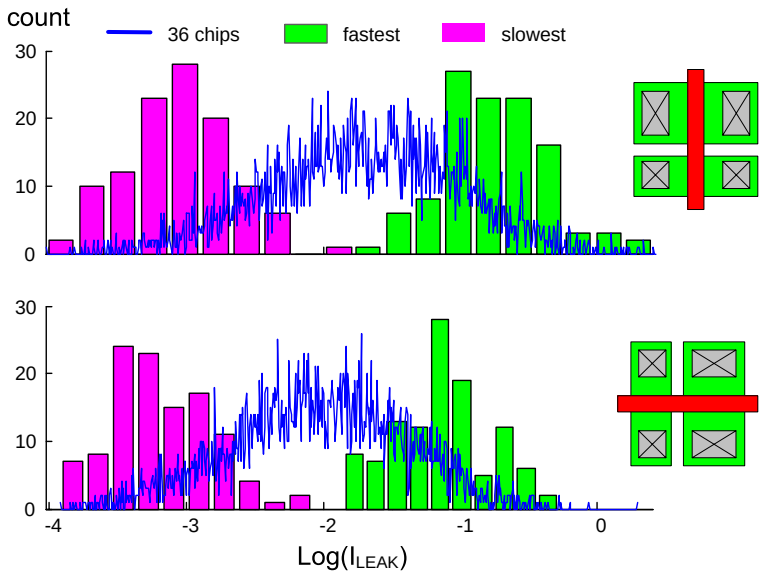


Figure 6.14: I_{LEAK} distributions for non-rotated and rotated ROs in the 90nm test-chip.

45nm test-chip. ROs were not rotated in the 45nm process because rotation of minimum length poly-Si gates is not allowed. The distribution of normalized RO frequency before normalizing to the SS corner shows that the extracted frequency and measured data are significantly different for the two layouts. After normalizing with the SS corner, a significant shift is observed and the RO frequency of the vertically placed gates is significantly lower than horizontally placed ones. However, the leakage currents remain almost the same. This is likely due to the parasitics in the metal interconnects which have not been accurately extracted. In this case, comparison of the mean values of frequency is not valid.

6.2.3 Symmetry

In the 90nm test-chips, the difference between structures that are mirror image of each other is small and is likely due to the coma effect. Roughly 1-2% shift in the mean RO frequency was observed with single gate stages as shown in the second and third plots

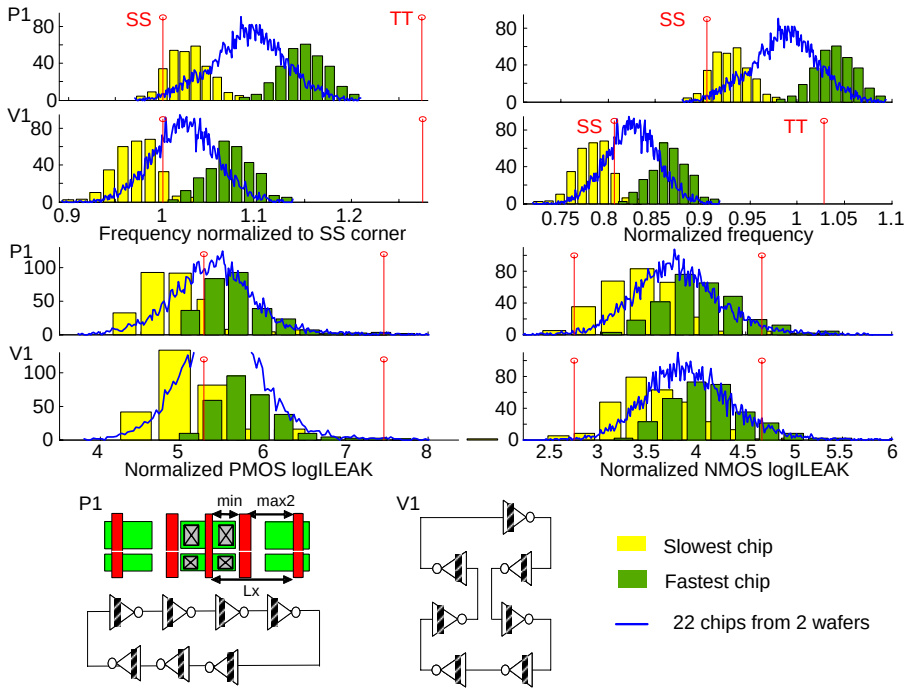


Figure 6.15: Effect of placement of gates horizontally and vertically in the 45nm test-chip.

of Figure 6.9. This was verified with a high level of confidence.

In the 45nm test-chips, this difference has also been observed as shown in Figure 6.16. However, although RO frequency increased by around 1%, both PMOS and NMOS leakage currents decreased. This could be due to a combination of the coma effect and assymmetry in the pocket doping process. However, since the difference is small, it is difficult to infer the exact cause.

6.2.4 Metal Coverage Over Gate

In the 90nm test-chips, M1 coverage over gates has negligible effects on RO frequency and causes a small reduction in leakage current, indicating a small increase in V_{th} (Figure 6.17). It has been reported [70] that metal-covered gates suffer from larger number

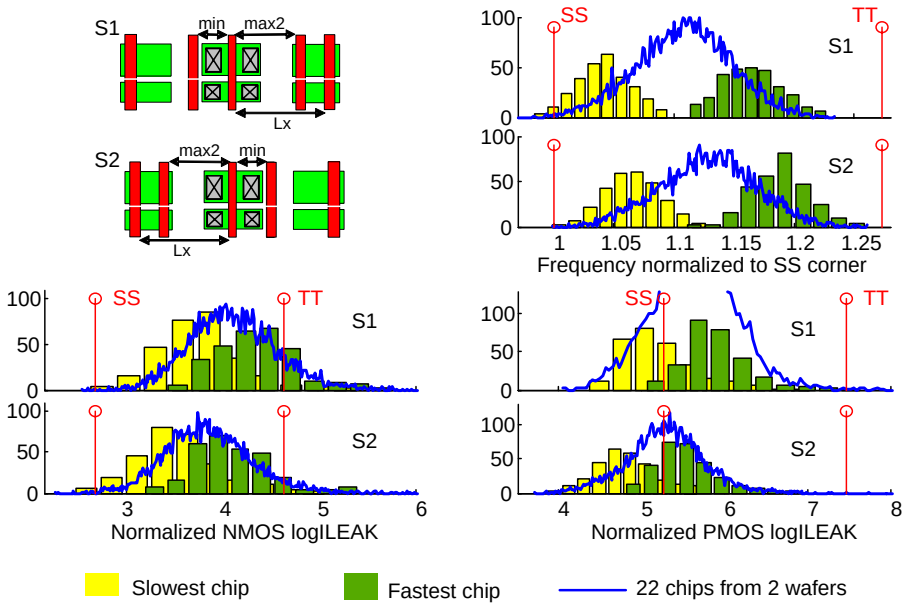


Figure 6.16: Effect of symmetrical layouts which are mirror images of each other in the 45nm test-chip.

of interface states because the metal coverage lowers the temperature on the gate during annealing, which resulted in V_{th} increase of $\approx 10\%$. In our layout, metal-1 is in close proximity to non metal-covered gate since the gate-length is small and the source and drain are fully covered with metal-1 contacts. Hence, it is likely that the temperature difference during annealing between metal-covered gates and non metal-covered gates is too small to create a significant shift in V_{th}. It is also possible that the process has sufficiently high annealing temperature to passivate the interface states of metal-covered gates resulting in a small V_{th} shift due to different annealing temperature. The effect of V_{th} shift on RO frequency at high voltage is less significant.

In the 45nm testchips, the gates were covered with M2 instead due to the space limitation. In this case the impact of metal coverage is negligible for both RO frequency and I_{LEAK} measurements (Figure 6.18).

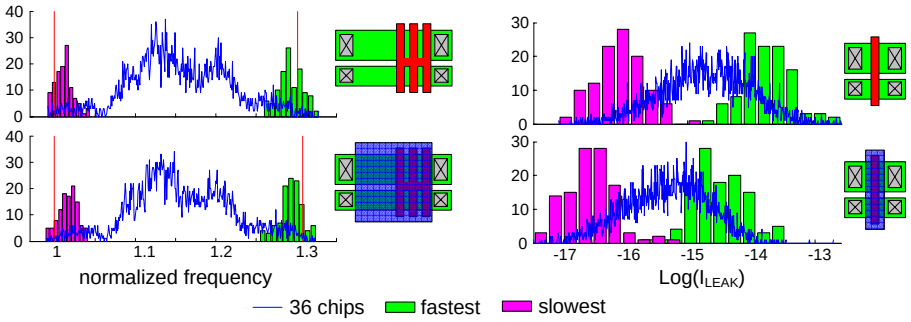


Figure 6.17: Effect of M1 coverage over gates in the 90nm test-chips.

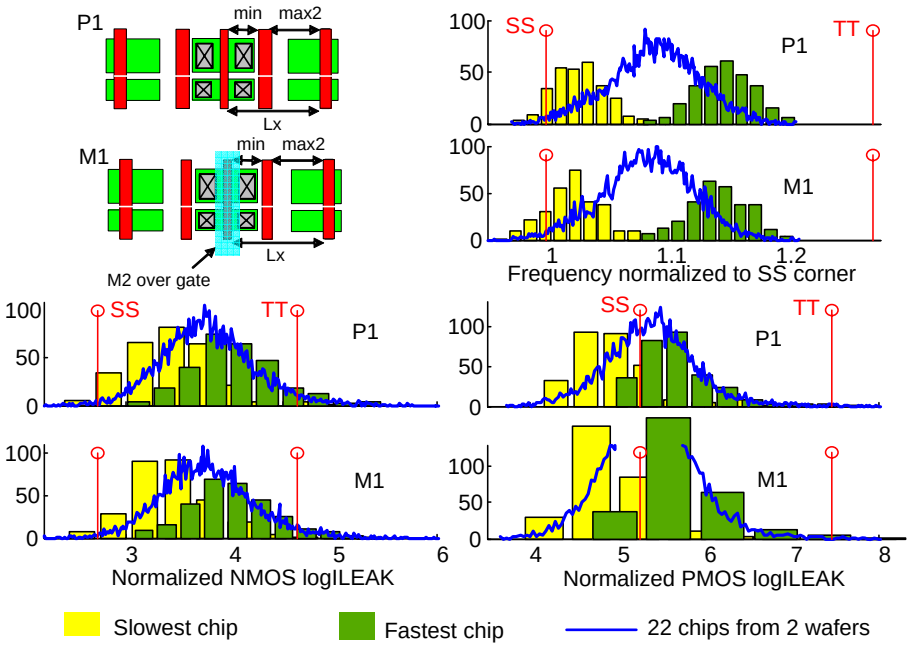


Figure 6.18: Effect of M1 coverage over gates in the 45nm test-chips.

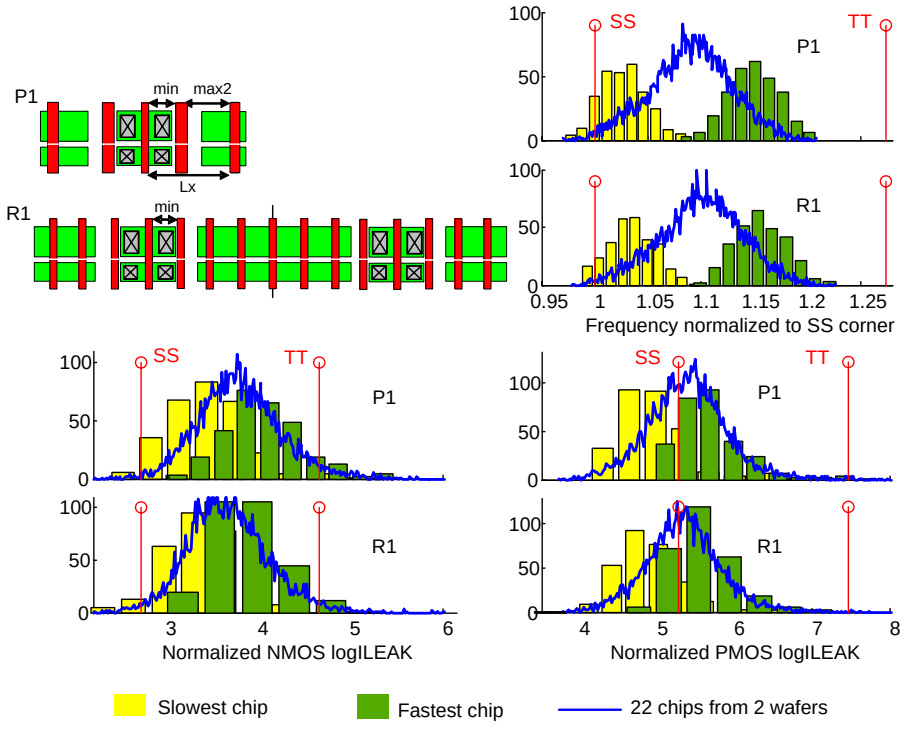


Figure 6.19: Effect of fixed poly-Si gate pitch in the 45nm test-chips.

6.2.5 Fixed Poly-Si Gate Pitch

The use of fixed gate pitch has negligible impact on variability as shown in Figure 6.19. RO frequency of fixed gate pitch layout(R1) has a σ/μ of 2.2% which is the same as for non-fixed gate pitch layout(P1). This shows that, as long as there is regularity, the use of fixed poly-Si pitches in a grid-like layout does not reduce variability significantly.

6.2.6 Shallow Trench Isolation (STI)

Figure 6.20 compares the impact of layouts with and without STI in the 45nm process. The layout without STI is slower by around 3% and has higher PMOS leakage current likely due to the effect of L variation on PMOS leakage. As this process uses SACVD trench oxide that generates a low tensile strain, STI stress increases the mobility of NMOS

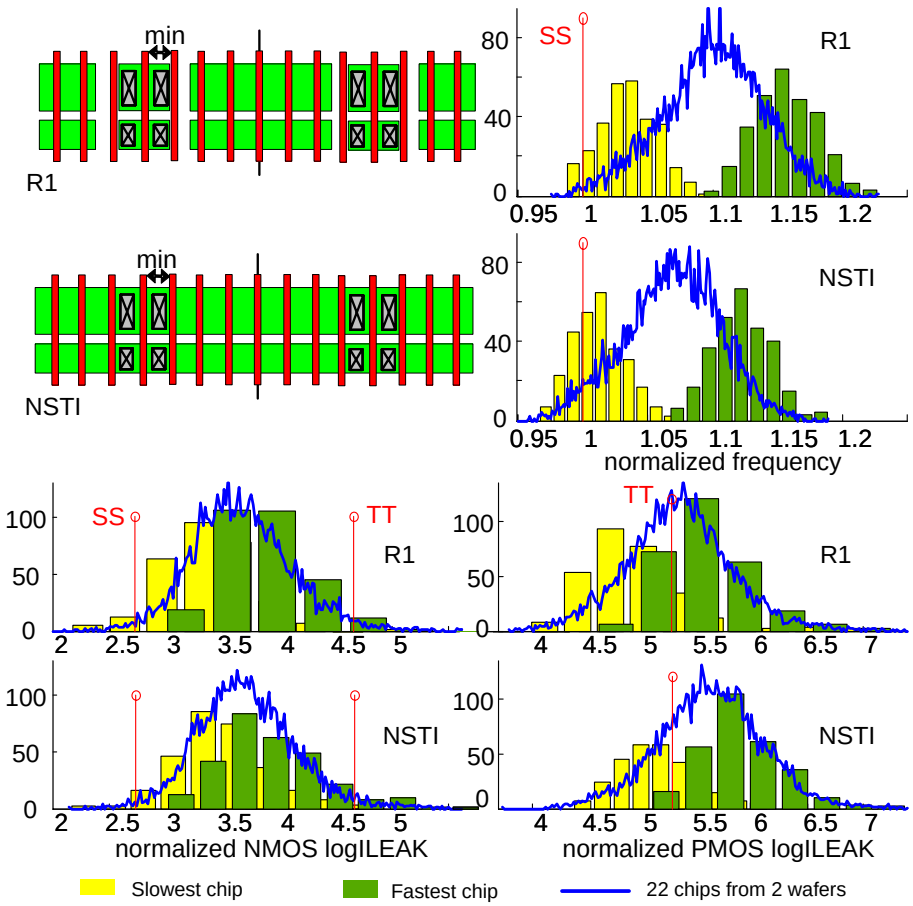


Figure 6.20: Effect of STI on RO frequency and transistor leakage currents in the 45nm test-chips. NSTI uses gate isolation [7] instead of STI.

transistors and causes the layout with STI to be faster, thereby compensating the effect of L. Variation in L could be due to the STI step that causes unevenness on the surface of the chip.

6.2.7 Source/Drain Diffusion area

Figure 6.21 studies the impact of a longer S/D diffusion in the 45nm process. Differences in S/D capacitance are captured by the layout extraction. After normalization, the layout with larger S/D area is 5% faster, while the leakage currents remain approximately

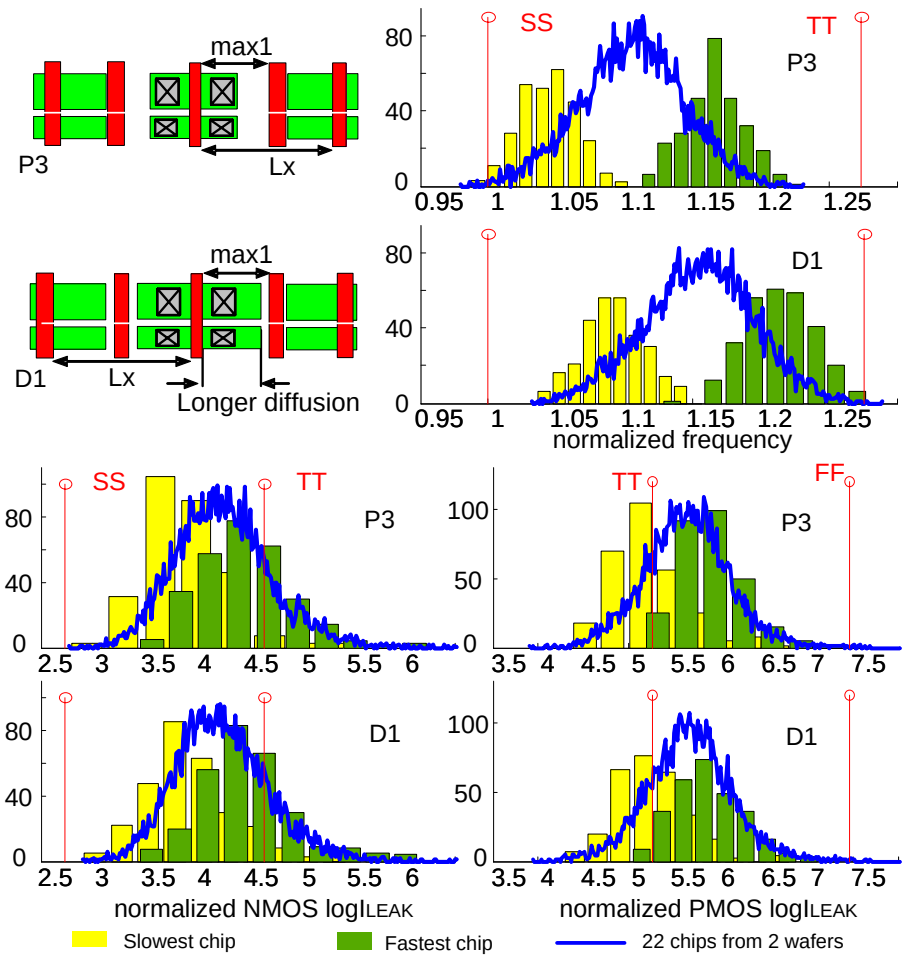


Figure 6.21: Effect of a longer S/D diffusion on RO frequency and transistor leakage currents in the 45nm test-chips.

unchanged. This can be explained by the fact that a larger S/D area will allow the CESL to exert more tensile strain on the transistors, thereby increasing the mobility of the NMOS transistors and resulting in faster RO. Since mobility varies linearly with leakage current, its effect on leakage current in the log scale is small.

6.3 Characteristics of D2D and WID Variations

In the 90nm process 36 chips from half a wafer were measured and in the 45nm process, 22 chips from 2 wafers were measured. Less systematic variations were observed in the 45nm test-chip for both D2D and WID variations. However, random WID variations increased in 45nm due to reduction in channel area. Spatial correlation in 90nm was significant due to systematic WID variations whereas no significant spatial correlation was found in 45nm.

6.3.1 Statistics of D2D and WID Variations

In the 90nm process, the wafer map of mean frequency and leakage of each die for the densest single gate layout configuration is shown in Figure 6.22. It shows a typical radial pattern that can be attributed to non-uniform resist development [11, 13] or systematic temperature gradient during RTA [71]. Faster and leakier chips are located at the center of the wafer. D2D variation is significant resulting in a $3\sigma/\mu$ of 15% in RO frequency over half a wafer for the densest single gate structure. For the other single gate structures, it increases slightly to around 17%.

In the 45nm process, measured results of 22 dies from two wafers are studied. For each layout, there is a shift of $\approx 2.5\%$ in frequency between the mean of the 2 wafers. D2D $3\sigma/\mu$ is shown in Figure 6.23. It varies from 6.5% to 7.5% for the 2 wafers and overall D2D variation was 8.5% for all the layouts except for the vertically placed layout V1. Extensive measurements on more chips even distributed over the wafer gives a D2D variation of $\approx$ 15%[1].

[1]Measurements by Kun Qian, U.C. Berkeley.

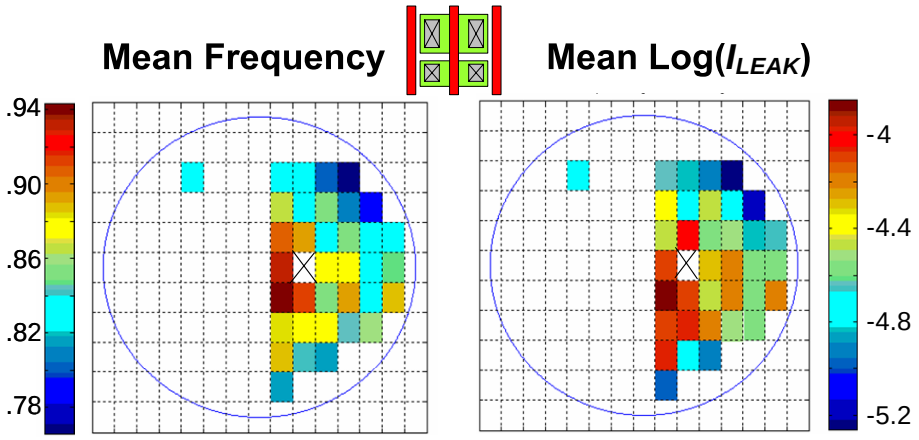


Figure 6.22: Wafer maps of mean RO frequency and mean Log(ILEAK) in the 90nm process. X marks a defective chip. Location of dies on only half the wafer are known.

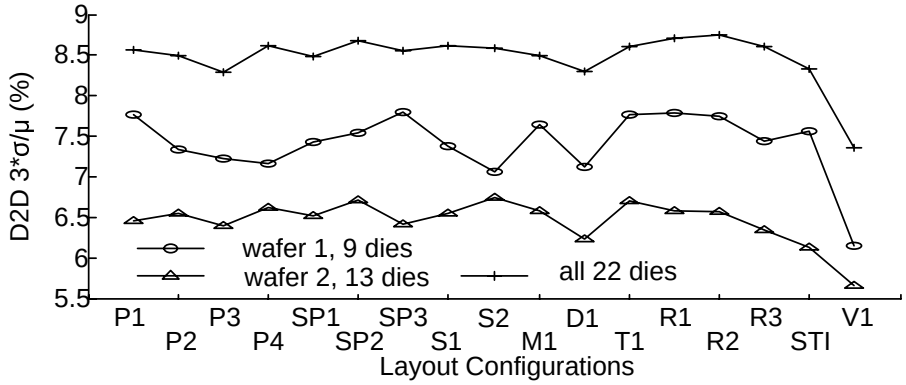


Figure 6.23: Comparison of RO frequency D2D variations for 2 wafers and the overall D2D variations in the 45nm process.

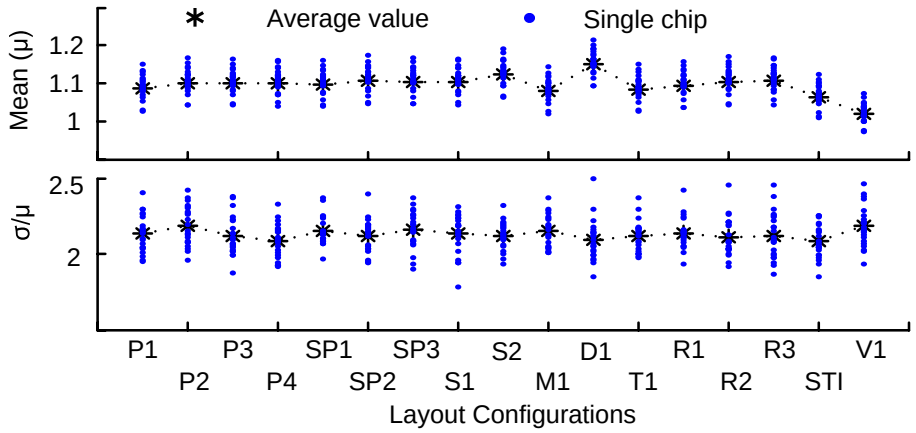


Figure 6.24: 45nm WID statistics of RO frequency for 22 dies. The frequency is normalized to the SS corner frequency.

Figure 6.24 plots the WID mean and σ/μ of the RO frequency for each of the 22 dies in the 45nm process as a function of the layout configuration. The mean frequency has been normalized to the SS corner in order to compare differences in layouts that have not been captured by the layout extraction. WID $3\sigma/\mu$ of RO frequency is approximately 6.6% which is more than twice that in 90nm ($3\sigma/\mu \approx 3.5\%$). This is consistent with the transistor channel area reduction by a factor of 4 between the two processes. No significant dependence of WID variation on layout was observed.

6.3.2 WID Spatial Correlation of ROs

In the 90nm process, WID variation of identical layout structures is small ($3\sigma/\mu \approx 3.5\%$) and weakly dependent on the layout. For each layout structure, the data is normalized to zero mean and unit variance for each chip before being used to compute its spatial correlation. Confidence intervals for the correlation are computed using Fisher's z-transformation to convert Pearson's correlation to a normally distributed random variable. In the fre-

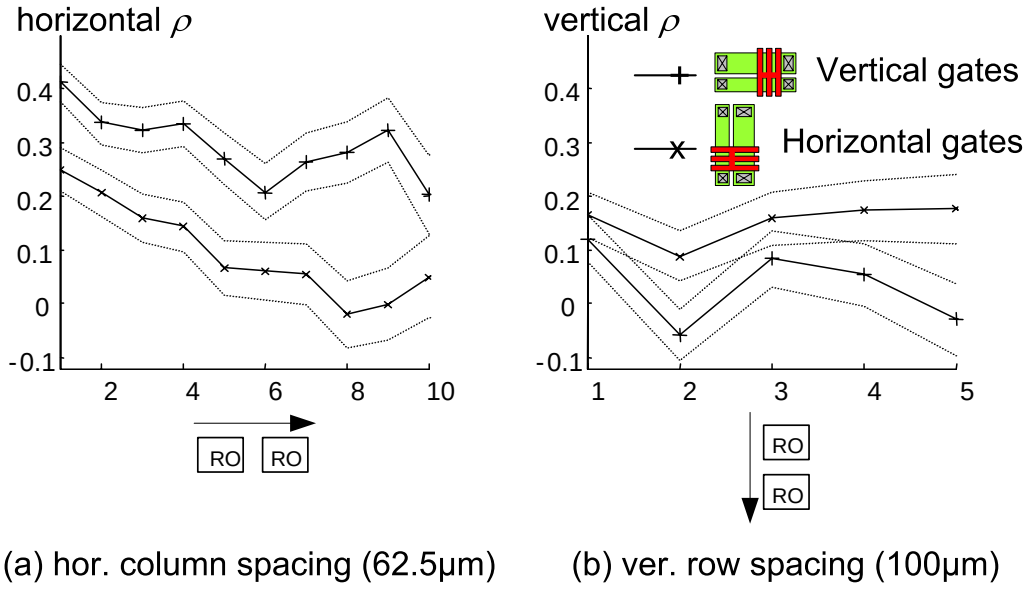


Figure 6.25: Spatial correlation coefficient (ρ) vs a) horizontal column spacing and b) vertical row spacing, for vertical and horizontal gates in the 90nm process. Dotted lines represent 99% confidence bounds for ρ.

quency measurements, the spatial correlation depends only on the direction of spacing and orientation of the gates.

Spatial correlation is higher in the horizontal direction of the chip than in the vertical. In particular, for horizontally spaced ROs (Figure 6.25a), correlation is higher for vertically oriented gates, whereas for vertically spaced ROs (Figure 6.25b), correlation is higher for horizontally oriented gates. This data is summarized in Table 6.5, showing that horizontally spaced ROs with vertically oriented gates have a spatial correlation of 0.4 between immediate neighbors whereas horizontally spaced ROs with horizontally oriented gates have a spatial correlation of 0.25.

This dependence on direction could be explained by the step-and-scan photolithography. The horizontal direction in this chip is along the slit of light in the stepper and is

Slit direction = horizontal Scan direction = vertical		Horizontal spacing	Vertical spacing
RO Freq.	Vertical Gates	0.4	0.1
	Horizontal Gates	0.25	0.2
I_{LEAK}		0	0

Table 6.5: Spatial correlation coefficient of neighboring ROs for different gate orientation and spacing direction.

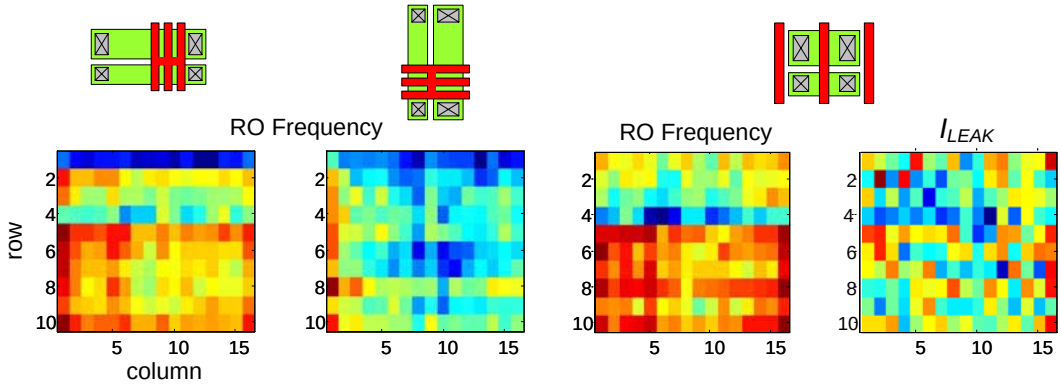


Figure 6.26: Surface plot of normalized data of 36 chips in the 90nm process. RO of layout with vertically oriented gates show a 1% frequency shift between rows 4 and 5.

subject to lens aberrations and curvature, resulting in more correlated features. The vertical direction is along the scan direction, which is subject to variation in scan speeds, stage vibration that affects focus, and light dosage, resulting in less correlated features.

Other manufacturing effects such as e-beam mask stitching [77] discontinuity or systematic shift in wafer stage scan speed could also account for the observed dependence on direction. In order to investigate the systematic reticle field effects, normalized data from 36 chips is averaged and Figure 6.26 shows the resulting surface plot of RO frequency and I_{LEAK}. There is a sharp systematic shift in frequency between the 4th and 5th rows for RO

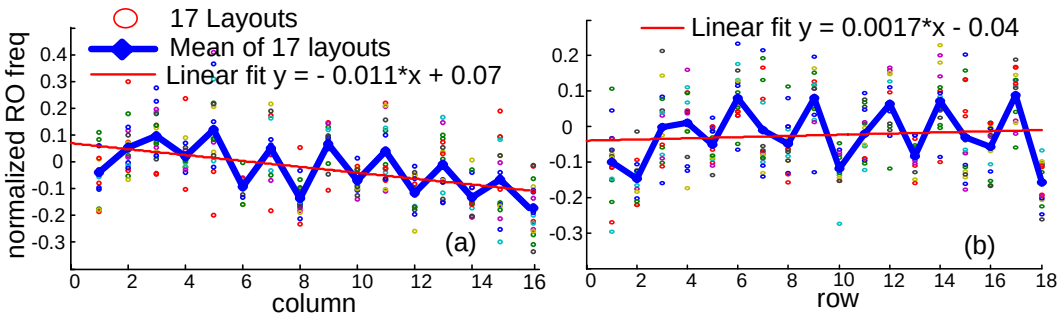


Figure 6.27: Plot of 45nm RO frequency normalized to zero mean and unit variance showing mask characteristics. Small systematic variation of $\approx 0.2 \times \sigma$.

structures with vertically oriented gates and a small shift in I_{LEAK}. This could be attributed to a discontinuity at the level of the mask reticle which suggests e-beam mask stitching defects. Another reason could be a shift in the scan speed of the wafer/mask stage during scanning of the reticle image onto the wafer. Since the scan speed is adjusted simultaneously during scan based on look-ahead sensors, it is possible that there is discontinuity in the sensor readings causing the scan speed to shift from the 4^{th} to 5^{th} row. Leakage current, which is less sensitive to gate-length (L) but more sensitive to threshold voltage than RO frequency, has much smaller spatial correlation.

In the 45nm process, systematic variations in the mask are also investigated by normalizing the data of each chip to zero mean and unity standard deviation and averaging the normalized data from 22 chips to remove random variations. Figures 6.27a and 6.27b plot the mean frequency for each column and each row respectively. No significant systematic variation was found, partially due to the fact that random WID variation has increased. Finally, there is no significant spatial correlation in the WID measurements.

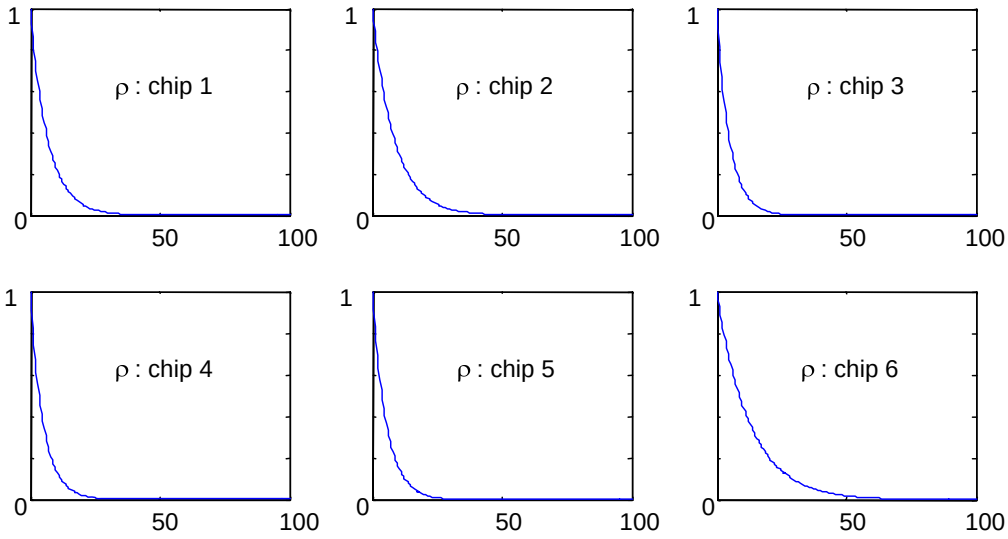


Figure 6.28: Plots of correlation coefficient (ρ) vs spacing between inverters in a RO obtained from fitting measurement data to a exponential model.

6.3.3 Spatial Correlation of Gates in a RO

The 90nm test-chip contains an array of VRO tiles as described in Section 6.1. Each tile contains 8 ROs with 13, 19, 25, 31, 37, 43, 49 and 55 number of stages for characterizing the correlation of delays between inverters that are connected to each other (Section 6.1).

In the analysis, systematic variations in the vertical direction are significant and are removed by shifting each measurement by the mean of the ROs of the same row. In this way the variances observed are due to random variations instead of systematic RO to RO variations that will render inverters within a RO strongly correlated. The measured data is fitted to the model and the extracted correlation plots for 6 chips are shown in Figure 6.28. Except for chip 6, correlation goes to almost zero after $\approx 40\mu m$.

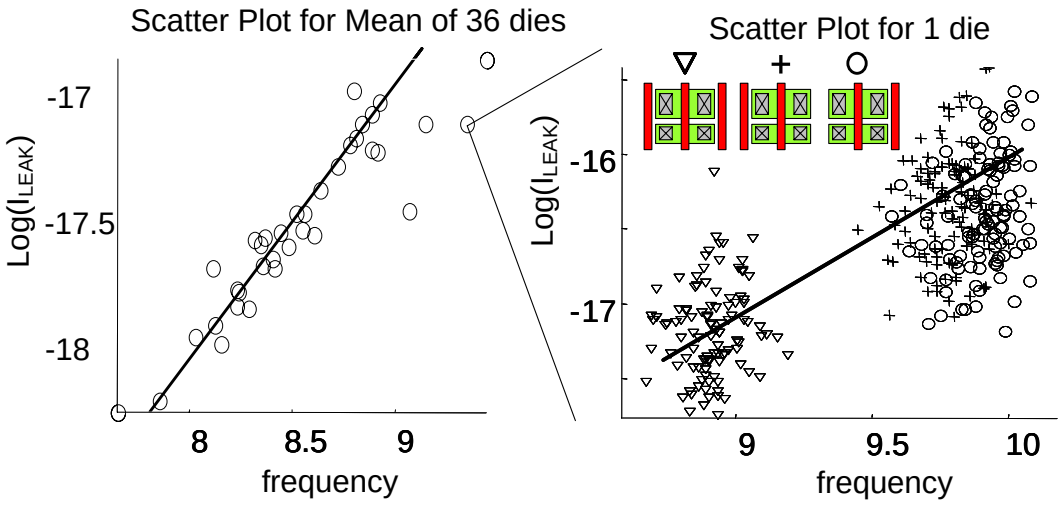


Figure 6.29: 90nm scatter plot of $Log(I_{LEAK})$ vs frequency showing a) mean $Log(I_{LEAK})$ and mean RO frequency for 36 dies, b) all $Log(I_{LEAK})$ and RO frequency for 3 layouts on 1 die.

6.4 Inferring Process Parameters (90nm)

In the 90nm process, the scatter plot of $Log(I_{LEAK})$ vs frequency shows a strong positive correlation between I_{LEAK} and frequency for D2D and layout-to-layout (L2L) variations (Figure 6.29). This means that these variations are dominated by process parameters which cause both $Log(I_{LEAK})$ and frequency to vary in the same manner. WID scatter plot shows no significant correlation, suggesting that it is caused by a combination of process parameters.

In order to relate these variations to process parameters, the least squares method and the Berkeley Short-channel IGFET Model (BSIM3) model are used to infer the variation in gate length (L), gate oxide thickness (Tox) and channel dopant concentration (Nch) from the frequency (F) and leakage current ($I = Log(I_{LEAK})$) measurements. Since the sensitivities of the measured data to process parameters vary with different condition of

No.	Vdd (V)	Vbs (V)	T (^oC)
1	1	0	25
2	**1.4**	0	25
3	1	**-0.2**	25
4	1	0	**60**

Table 6.6: Measurement conditions.

supply voltage (Vdd), substrate voltage (Vbs) and temperature (T), the accuracy of the inference has been improved by taking the measurements under the different conditions listed in Table 6.6.

If index i represents the four different measurement conditions, then:

- F_m, I_m: Estimated mean F and I
- L_m, Tox_m, Nch_m: Estimated mean process parameters
- F_0, I_0: Mean measurement data for 1 die and 1 layout
- L_{typ}, Tox_{typ}, Nch_{typ}: Typical values of L, Tox and Nch

$$Y_i = \begin{bmatrix} \frac{F-F_m}{F_0} \\ \\ \frac{I-I_m}{I_0} \end{bmatrix}_i , X = \begin{bmatrix} \frac{L-L_m}{L_{typ}} \\ \\ \frac{Tox-Tox_m}{Tox_{typ}} \\ \\ \frac{Nch-Nch_m}{Nch_{typ}} \end{bmatrix} , M_i = \begin{bmatrix} \frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{L}{L_{typ}}\right)} & \frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{Tox}{Tox_{typ}}\right)} & \frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{Nch}{Nch_{typ}}\right)} \\ \frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{L}{L_{typ}}\right)} & \frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{Tox}{Tox_{typ}}\right)} & \frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{Nch}{Nch_{typ}}\right)} \end{bmatrix}$$

$$\Rightarrow Y_i = M_i \cdot X$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{bmatrix} \cdot X \Rightarrow \hat{X} = M^+ \cdot Y$$

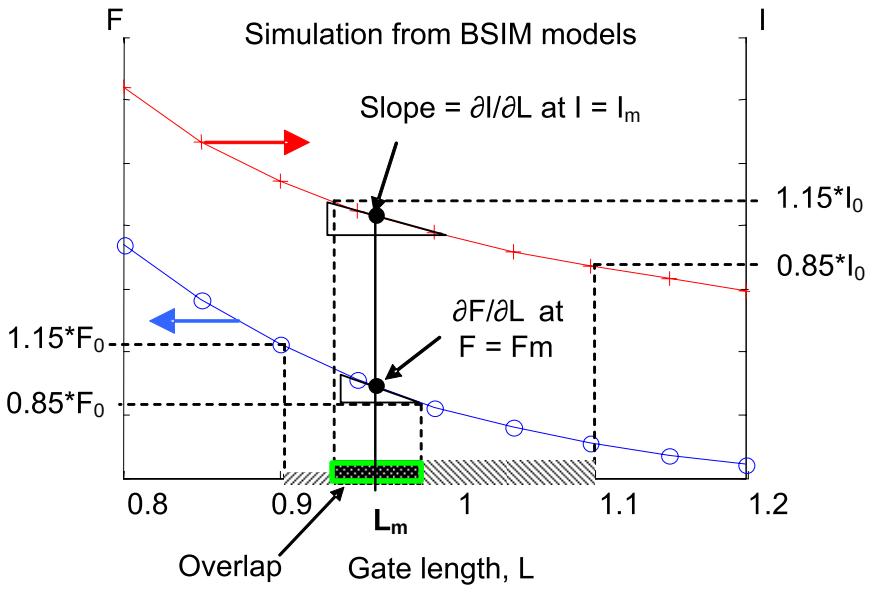


Figure 6.30: Illustrates the estimation of the nominal value of L (L_m) in the case where only one parameter (L) is inferred. I_0 and F_0 are the measurement data. L_m is used to obtain the gradients and to find I_m and F_m which are used to normalize the data in the least squares estimation.

where

M^+ = pseudo-inverse of M

$\hat{X}$ = Least squares estimate of X

By using SPICE (Simulation Program with Integrated Circuit Emphasis) simulations, F and I are plotted vs. L, and the overlap region of L which gives a $\pm 15\%$ change in the mean measured data F_0 and I_0 is found. The mean of this overlap region gives L_m which will be used to obtain the gradients and the normalization constants I_m and F_m. The parameters are then normalized and process parameters are inferred using least squares estimation. Figure 6.30 illustrates this procedure.

This method is used for three process parameters, L, Tox and Nch. By using

	Default	**Vdd = 1.4V**	**Vbs = -0.2V**	**T = 60^oC**
$\frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{Nch}{Nch_{typ}}\right)}$	-0.052	-0.053	-0.039	-0.049
$\frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{Tox}{Tox_{typ}}\right)}$	-0.35	-0.35	-0.14	-0.39
$\frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{L}{L_{typ}}\right)}$	-1.79	-1.83	-1.65	-1.86
$\frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{Nch}{Nch_{typ}}\right)}$	0.061	0.066	0.090	0.050
$\frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{Tox}{Tox_{typ}}\right)}$	-0.045	-0.056	-0.098	-0.028
$\frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{L}{L_{typ}}\right)}$	1.01	1.07	1.23	0.92

Table 6.7: Sensitivities used in the 90nm process parameters extraction for one chip and one layout structure.

SPICE simulations to obtain plots of F and I vs L, Tox and Nch over a small range around their typical values for the different conditions listed in Table 6.6. Using these plots, the measured F and I data are mapped to an overlapping range of values of the process parameters. The mean values of this range (L_m, Tox_m, Nch_m) are used as the nominal value of X to obtain the gradients in M and the normalization constants (F_m, I_m). For a typical chip with the densest single gate layout configuration, the gradients (M) are given in Table 6.7 for the four different conditions. The components in X and Y are then normalized. Finally, least squares estimation is applied to estimate X ($\hat{X}$). This is repeated for each die and each layout configuration since the small amount of variation assures a good linear approximation.

Figure 6.31 shows the distribution of normalized L, obtained from the inference process. The sensitivities of F and I to L are larger than that of other process parameters as shown in Table 6.7. This results in significantly more accuracy in the inferred L. The plots

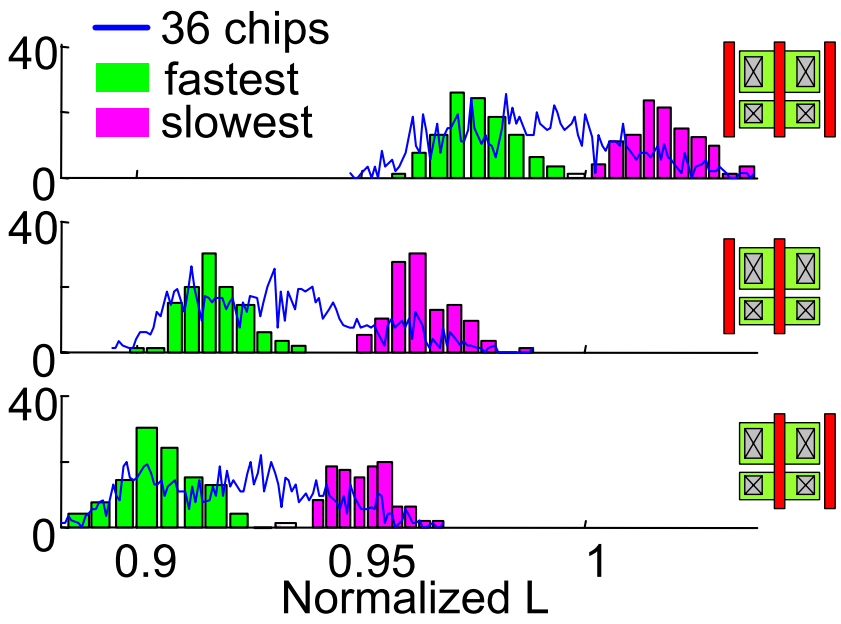


Figure 6.31: De-embedded distribution of L.

of normalized L show a strong correlation to the frequency and leakage plots in Figure 6.9 and 6.10. Faster and leakier ROs have shorter L. The mean values of the L distributions depend significantly on the layout, showing that systematic layout-to-layout variation is attributed primarily to L shifts. D2D spreads of L are significant as the bar-plots of the fastest and slowest chip are spaced far apart.

Figure 6.32 shows a typical surface plot of a die showing the WID variation of measured F and I. and the corresponding inferred normalized L. Systematic WID variation observed in F is strongly reflected in L and varies by $\approx 1\%$.

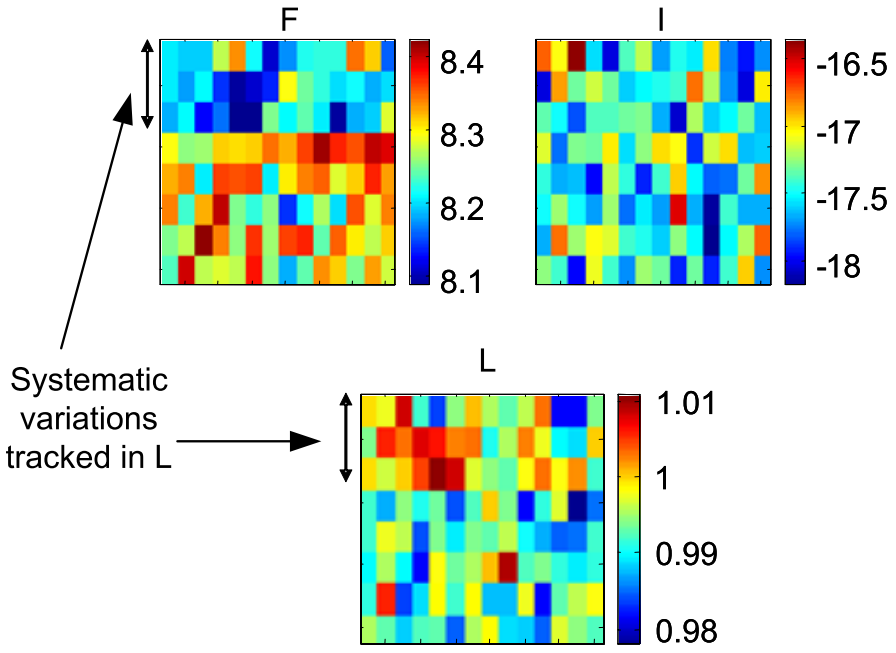


Figure 6.32: Top plots: surface plots of within-die RO frequency (F) and $Log(I_{LEAK})$ (I) of a typical die; Bottom plot: corresponding surface plots of estimated normalized L.

6.5 Summary

In 90nm, the largest impact of layout on performance comes from gate poly-Si density, which causes a systematic shift in frequency of up to 10%. D2D variation is significant resulting in a 3× standard-deviation/median ($3\sigma/\mu$) of 15% over half a wafer. Finally, WID variation for identical structures is relatively small ($3\sigma/\mu \approx 3.5\%$). WID spatial correlation of RO frequency is significant and shows a dependency on the direction of spacing and the orientation of the gates. Each of these observations can be attributed to a particular step in the manufacturing process.

In 45nm, systematic layout-induced variations, in particular those related to poly-Si density, are significantly reduced compared to the early 90nm process. Variations due to strain and other factors are now dominant. Finally, random WID variation has increased

	90nm	**45nm**
Proximity Effects	10%	2%
D2D ($3\sigma/\mu$)	15%	15%
WID ($3\sigma/\mu$)	3.5%	6.6%
WID Spatial Correlation	Significant	Insignificant

Table 6.8: Comparison of 45nm and 90nm technology results.

proportionally to transistor area reduction, while systematic D2D variation has remained the same. Table 6.8 compares the results of the two technologies.

Chapter 7

Conclusion

Two test-chips have been designed as part of a methodology to measure and analyze variability in CMOS processes. These chips contain circuits to measure an array of RO and leakage transistors. The array is designed to investigate the impacts of layout on transistor performance and allows for the measurement of systematic, random, D2D and WID variations. This methodology was successfully applied to a 90nm and a 45nm process. Several circuits for measuring SRAM performance variability have also been proposed. These circuits permit the measurement of static and dynamic performance metrics of the SRAM cell in an array with minimum perturbation to the array layout.

7.1 Key Contributions

The contributions of this work is two-fold. Firstly, we have developed a methodology to measure and analyze systematic layout dependent effects and both systematic and random WID and D2D variations in the early stages of a CMOS process. This was

achieved with a custom-designed test-chip containing an addressable array of devices under test (DUT) and accurate on-chip measurement circuits. The DUT consists of ring oscillators (RO) and transistors in the off-state with different layout configurations that investigate layout effects. The array is made up of small scalable modules that easily abut to form the big array. Scan chains are used to address the array, resulting in a small pad count. RO frequency is measured off-chip after dividing the frequency down with on-chip dividers. An on-chip current measurement circuit in the form of a single slope ADC was used to measure leakage currents of the off-transistors with good accuracy. Finally, a complex setup using a custom-designed PCB, easily available instruments, and GPIB connection was used to automate the data collection.

Secondly, analysis of the measurement data shows that systematic layout dependent effects are significant in 90nm. In 45nm, layout effects are weaker and dominated by strain. Other less known systematic effects were also revealed. In 90nm, rotated structures have smaller variation, and symmetrical structures that are mirror image of each other experience a small but systematic shift in frequency. Metal coverage over gate causes V_{th} to increase resulting in lower I_{LEAK}. In 45nm, strain effects due to an increase in source/drain area dominate the systematic effects but overall effects are small. Effects due to STI strain and poly-Si pitch are within 2%.

Characteristics of D2D and WID variations were also measured and analyzed. In 90nm, D2D variation was significant and random WID variation was small, whereas in 45nm, D2D variation has remained approximately the same and random WID variation has increased, likely due to the reduction in channel area. The systematic component of D2D

variation is important and exhibits a radial pattern on the wafer.

7.2 Methods to Mitigate Variability

Layout has a significant impact on lithography-induced variability in 90nm technology. Analysis of the inferred process parameters indicates that poly-Si gate pitch has the strongest impact on the effective transistor gate length, resulting in up to 10% variation in RO frequency for inverters laid out with single poly-Si lines. This systematic effect can be compensated by using layout extraction tools that account for proximity effects [76] or better OPC [92]. A simpler method would be to use regular layouts [93] or more restrictive layout rules that only allow a few possible gate pitches, together with an extractor that maps each gate pitch to its respective gate length.

The use of step-and-scan lithography induces stronger correlation between gates placed parallel to the direction of slit of light than those parallel to the direction of scan. This effect can be exploited in the layout of regular datapaths and memory [8, 94]. By placing gates of the same path in the low correlation direction and by placing parallel paths in the high correlation direction, a tighter performance spread can be obtained.

Analysis of the spatial correlation of ROs and inverters can be used to model the spatial correlation coefficient of circuit paths and interconnected gates, allowing for more accurate statistical timing analysis.

For the same layout, systematic D2D variations dominate the total variations for small chips. Knowing the contribution of WID and D2D variation to the total variation will allow circuit designers to design for WID variations and use die-level compensation schemes

[32] to reduce D2D variations. Process engineers will also be able to focus on reducing the dominant sources of variations.

More restricted design rules and likely better OPC in 45nm process reduce the layout induced performance variations by limiting variations in poly-Si density. Only 2% frequency shifts due to proximity effects were measured in the 45nm test-chips. However, other layout-induced variations due to the CESL and STI stress have become more significant and are added on to the total variation.

Measurements show a trend toward less systematic but more random variations. From 90nm to 45nm, random WID variation has more than doubled whereas systematic D2D variation has remained unchanged. One way to reduce random variations is to increase channel area. As this increases area and power, only critical gates should be wider. Also, averaging more gates in a critical path will reduce the uncertainty over the path delay.

7.3 Future Work

The existing methodology can be refined to measure a larger range of current. Instead of measuring leakage current only, the array can be modified to make I-V measurements which can be used to obtain more accurate parameter extraction.

Investigating other layout effects such as well-proximity effects, gate width variation and alignment errors can yield interesting results. TCAD tools that simulate the gate formation process can be used to match measurement results to predicted behavior. Layouts that are sensitive to certain aspects of lithography such as focus or aberrations can

be used to monitor these effects[1].

This work focus on the variability of the transistor. However, interconnects variability can also have a significant impact on performance. Variability of the backend process can be measured and analyzed using the RO array. New methods can also be developed to measure the capacitance and resistance of wires and vias.

A methodology for SRAM measurement can be developed based on the circuits presented in Chapter 5. These circuits measure the static and dynamic noise margins of the SRAM cell in the array with minimum perturbation of the array layout. Measurements from the SRAM write RO and the SRAM read RO correlate with access times measurements but are simpler, more accurate, and faster to measure[2].

7.4 Conclusion

In conclusion, a methodology has been successfully applied, to characterize variability in a 90nm and a 45nm CMOS process. Sources of systematic variations have been identified through analysis of the measurement results. Systematic and random components of D2D and WID variations have been measured. Several new layout-dependent effects and novel circuits for characterizing SRAM performance variability have been proposed for future investigation. This will setup a framework for characterizing variability and help in the mitigation of its effects.

[1]Work in progress by Lynn Wang, U.C. Berkeley.
[2]Work in progress by Jason Tsai, U.C. Berkeley.

Bibliography

[1] A. B. Kahng and Y. C. Pati, "Subwavelength lithography and its potential impact on design and eda," in *Proc. Design Automation Conference*, New Orleans, LA, USA, June 1999, pp. 799–804.

[2] E. Josse, S. Parihar, O. Callen, P. Ferreira, C. Monget, A. Farcy, M. Zaleski, D. Villanueva, R. Ranica, M. Bidaud, D. Barge, C. Laviron, N. Auriac, C. L. Cam, S. Harrison, S. Warrick, F. Leverd, P. Gouraud, S. Zoll, F. Guyader, E. Perrin, E. Baylac, J. Belledent, B. Icard, B. Minghetti, S. Manakli, L. Pain, V. Huard, G. Ribes, K. Rochereau, S. Bordez, C. Blanc, A. Margain, D. Delille, R. Pantel, K. Barla, N. Cave, and M. Haond, "A cost-effective low power platform for the 45-nm technology node," in *2006 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec 2006, pp. 1–4.

[3] C. Cam, F.Guyader, C. Buttet, P.Guyader, G.Ribes, M.Sardo, S.Vanbergue, F.Boeuf, F.Arnaud, E.Josse, and M.Haond, "A low cost drive current enhancement technique using shallow trench isolation induced stress for 45-nm node," in *2006 Symp. VLSI Technology Dig. of Tech. Papers*, 2006, pp. 82–83.

[4] B. L. Gratiet, P. Gouraud, E. Aparicio, L. Babaud, K. Dabertrand, M. Touchet, S. Kremer, C. Chaton, F. Foussadier, F. Sundermann, J. Massin, J.-D. Chapon, M. Gatefait, B. Minghetti, J. de Caunes, and D. Boutin, "Process control for 45nm cmos logic gate patterning," in *Metrology, Inspection, and Process Control for Microlithography XXII*, ser. Proc. SPIE, J. A. Allgair and C. J. Raymond, Eds., vol. 6922, March 2008.

[5] L.-T. Pang and B. Nikolić, "Impact of layout on 90nm process parameter fluctuations," in *2006 Symp. VLSI Circuits Dig. of Tech. Papers*, June 2006, pp. 69–70.

[6] B. Nikolić and L.-T. Pang, "Measurements and analysis of process variability in 90nm cmos," in *8th Int. Conf. Solid-State and Integrated Circuit Technology*, Shanghai, China, 2006, pp. 505–508.

[7] L.-T. Pang and B. Nikolić, "Measurement and analysis of variability in 45nm strained-Si CMOS technology," in *Custom Integrated Circuits Conference Dig. of Tech. Papers*, September 2008, accepted to CICC 2008.

[8] K. Bowman, S. Duvall, and J. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. of Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, February 2002.

[9] S. Nassif, "Delay variability: sources, impacts and trends," in *ISSCC 2000, Dig. of Tech. Papers*, February 2000, pp. 368 – 369.

[10] M. Orshansky, L. Milor, P. Chen, K. Keutzer, and C. Hu, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE Trans. on Computer-Aided Design of Circuits and Systems*, vol. 21, no. 5, pp. 544–553, May 2002.

[11] J. Cain and C. Spanos, "Electrical linewidth metrology for systematic cd variation characterization and causal analysis," in *Metrology, Inspection, and Process Control for Microlithography XVII*, ser. Proc. SPIE, D. J. Herr, Ed., vol. 5038, June 2003, pp. 350–361.

[12] M. Orshansky, L. Milor, and C. Hu, "Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial mask-level correction," *IEEE Trans. on Semiconductor Manufacturing*, vol. 17, no. 1, pp. 2–11, February 2004.

[13] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos, "Modeling within-field gate length spatial variation for process-design co-optimization," in *Design and Process Integration for Microelectronic Manufacturing III*, ser. Proc. SPIE, L. W. Liebmann, Ed., vol. 5756, May 2005, pp. 178–188.

[14] E. G. Colgan, R. J. Polastre, M. Takeichi, and R. L. Wisnieff, "Thin-film-transistor process-characterization test structures," *IBM J. of Research and Development*, vol. 42, no. 3/4, May/July 1998.

[15] D. Boning, J. Panganiban, K. Gonzalez-Valentin, C. MdCowell, A. Gattiker, and F. Liu, "Test structures for delay variability," in *2002 ACM/IEEE Tau Workshop*, Monterrey, CA, USA, 2002.

[16] K. M. G.-V. Gettings, "Extraction of variation sources due to layout practices," Master's thesis, Massachusetts Institute of Technology, Cambridge, Mass., 2002.

[17] J. S. Panganiban, "A ring oscillator based variation test chip," Master's thesis, Massachusetts Institute of Technology, Cambridge, Mass., 2002.

[18] X. Ouyang, C. N. B. Milor, and R. F. W. Pease, "High-throughput mapping of short-range spatial variations using active electrical metrology," *IEEE Trans. on Semiconductor Manufacturing*, vol. 15, no. 1, pp. 108–117, February 2002.

[19] S. Ohkawa, M. Aoki, and H. Masuda, "Analysis and characterization of device variations in an lsi chip using an integrated device matrix array," *IEEE Trans. on Semiconductor Manufacturing*, vol. 17, no. 2, pp. 155–165, May 2004.

[20] K. Agarwal, F. Liu, C. McDowell, S. Nassif, K. Nowka, M. Palmer, D. Acharyya, and J. Plusquellic, "A test structure for characterizing local device mismatches," in *2006 Symp. VLSI Circuits Dig. of Tech. Papers*, 2006, pp. 67 – 68.

[21] K. Agarwal and S. Nassif, "Characterizing process variation in nanometer CMOS," in *Proc. Design Automation Conference*, San Diego, CA, USA, June 2007, pp. 396–399.

[22] D. Levacq, T. Minakawa, M. Takamiya, and T. Sakurai, "A wide range spatial frequency analysis of intra-die variations with 4-mm 4000 1 transistor arrays in 90nm CMOS," in *Custom Integrated Circuits Conference Dig. of Tech. Papers*, September 2007, pp. 257 – 260.

[23] V. Wang and K. Shepard, "On-chip transistor characterisation arrays for variability analysis," *Electronics Letters*, vol. 43, pp. 806 – 807, July 2007.

[24] W. Zhao, Y. Cao, F. Liu, K. Agarwal, D. Acharyya, S. Nassif, and K. Nowka, "Rigorous extraction of process variations for 65nm CMOS design," in *37th European Solid State Device Research Conference, 2007*, September 2007, pp. 89 – 92.

[25] V. Wang, K. Agarwal, S. Nassif, K. Nowka, and D. Markovic, "A design model for random process variability," in *9th Int. Symp. on Quality Electronic Design, 2008*, March 2008, pp. 734 – 737.

[26] A. Bassi, A. Veggetti, L. Croce, and A. Bogliolo, "Measuring the effects of process variations on circuit performance by means of digitally-controllable ring oscillators," in *2003 Int. Conf. on Microelectronic Test Structures*, March 2003, pp. 214–217.

[27] B. Zhou and A. Khouas, "Measurement of delay mismatch due to process variations by means of modified ring oscillators," in *2005 IEEE Int. Symp. on Circuits and Systems*, vol. 5, May 2005, pp. 5246–5249.

[28] K. Tae-Hyoung, R. Persaud, and C. Kim, "Silicon odometer: An on-chip reliability monitor for measuring frequency degradation of digital circuits," *IEEE J. of Solid-State Circuits*, vol. 43, pp. 874 – 880, April 2008.

[29] X. Ouyang, T. Deeter, C. N. Berglund, R. F. W. Pease, J. Lee, , and M. A. McCord, "High-throughput high-density mapping and spectrum analysis of transistor gate length variations in sram circuits," *IEEE Trans. on Semiconductor Manufacturing*, vol. 14, no. 4, pp. 318–329, November 2001.

[30] J. B. Kuang, J. C. Kao, H. C. Ngo, and K. J. Nowka, "Cascaded test circuit with inter-bitline drive devices for evaluating memory cell performance," U.S. Patent 7,349,271, March, 2008.

[31] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM J. of Research and Development*, vol. 50, no. 4/5, July/September 2006.

[32] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. of Solid-State Circuits*, vol. 37, no. 11, pp. 1396 – 1402, November 2002.

[33] H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: variability characterization and modeling for 65- to 90-nm processes," in *Custom Integrated Circuits Conference Dig. of Tech. Papers*, 2005, pp. 593–599.

[34] M. Orshansky, S. Nassif, and D. Boning, *Design for Manufacturability and Statistical Design: A Constructive Approach*, ser. Integrated Circuits and Systems, A. Chandrakasan, Ed. Springer, 2007.

[35] C. Liu, F. Baumann, A. Ghetti, H. Vuong, C. Chang, K. Cheung, J. Colonell, W. Lai, E. Lloyd, J. Miner, C. Pai, H. Vaidya, R. Liu, and J. Clemens, "Severe thickness variation of sub-3nm gate oxide due to Si surface faceting, Poly-Si intrusion, and corner stress," in *1999 Symp. VLSI Technology Dig. of Tech. Papers*, 1999, pp. 75–76.

[36] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 112 – 119, January 2002.

[37] M. Ercken, L. Leunissen, I. Pollentier, G. P. Patsis, V. Constantoudis, and E. Gogolides, "Effects of different processing conditions on line edge roughness for 193nm and 157nm resists," in *Metrology, Inspection, and Process Control for Microlithography XVIII*, ser. Proc. SPIE, R. M. Silver, Ed., vol. 5375, February 2004, pp. 266–275.

[38] P. Oldiges, Q. Lin, K. Petrillo, M. Sanchez, and M. H. M. Ieong, "Modeling line edge roughness effects in sub 100 nanometer gate length devices," in *Proc. SISPAD*, Seatle, WA, USA, 2000, pp. 131–134.

[39] J. Croon, G. Storms, S. Winkelmeier, I. Pollentier, M. Ercken, S. Decoutere, W. Sansen, and H. Maes, "Line edge roughness: characterization, modeling and impact on device behavior," in *2002 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, 2002, pp. 307–310.

[40] R. W. Keyes, "Scaling, small numbers and randomness in semiconductors," *IEEE Circuits and Devices Magazine*, vol. 10, pp. 28 – 31, May 1994.

[41] M. Eisele, J. Berthold, D. Schmitt-Landsiedel, and R. Mahnkopf, "The impact of intra-die device parameter variations on path delays and on the design for yield of low voltage digital circuits," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 5, pp. 360 – 368, December 1997.

[42] P. Stolk, F. Widdershoven, and D. Klaassen, "Modeling statistical dopant fluctuations in mos transistors," *IEEE Trans. on Electron Devices*, vol. 45, no. 9, pp. 1960–1971, September 1998.

[43] D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *1999 Symp. VLSI Circuits Dig. of Tech. Papers*, 1999, pp. 171–172.

[44] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, "Quantum mechanical enhancement of the random dopant induced threshold voltage fluctuations and lowering in sub 0.1 micron MOSFETs," in *1999 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 1999, pp. 535–538.

[45] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, , and S. Saini, "Increase in the random dopant induced threshold fluctuations and lowering in sub-100 nm MOSFETs due to quantum effects: A 3-D density-gradient simulation study," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 722 – 729, April 2001.

[46] V.-Y. A. Thean, M. Sadd, and B. J. White, "Effects of dopant granularity on superhalo-channel mosfets according to two- and three-dimensional computer simulations," *IEEE Trans. on Nanotechnology*, vol. 2, pp. 97–101, June 2003.

[47] T. Mizuno, J. Okumtura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, pp. 2216 – 2221, November 1994.

[48] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1433 – 1439, 1989.

[49] C. S. Smith, "Piezoresistance effect in germanium and silicon," *Phys. Rev.*, vol. 94, no. 1, pp. 42 – 49, April 1954.

[50] P. Chidambaram, C. Bowen, S. Chakravarthi, and R. W. C. Machala, "Fundamentals of silicon material properties for successful exploitation of strain engineering in modern cmos manufacturing," *IEEE Trans. on Electron Devices*, vol. 53, pp. 994–964, May 2006.

[51] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr, "The effect of general strain on the band structure and electron mobility of silicon," *IEEE Trans. on Electron Devices*, vol. 54, pp. 2183 – 2190, September 2007.

[52] H. Park, K. Jones, J. Slinkman, and M. Law, "The effects of strain on dopant diffusion in silicon," in *1993 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 1993, pp. 303 – 306.

[53] S. Thompson, G. Sun, K. Wu, J. Lim, and T. Nishida, "Key differences for process-induced uniaxial vs. substrate-induced biaxial stressed Si and Ge channel MOSFETs," in *2004 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 2004, pp. 221 – 224.

[54] S. Mohta, N.; Thompson, "Mobility enhancement," *IEEE Circuits and Devices Magazine*, vol. 21, pp. 18 – 23, Sept.-Oct. 2005.

[55] Z.-Y. Cheng, M. Currie, C. Leitz, G. Taraschi, E. Fitzgerald, J. Hoyt, and D. Antoniadas, "Electron mobility enhancement in strained-Si n-MOSFETs fabricated on SiGe-on-insulator (SGOI) substrates," *IEEE Electron Device Letters*, vol. 22, pp. 321 – 323, July 2001.

[56] T. Komoda, A. Oishi, T. Sanuki, K. Kasai, H. Yoshimura, K. Ohno, A. Iwai, M. Saito, F. Matsuoka, N. Nagashima, and T. Noguchi, "Mobility improvement for 45nm node by combination of optimized stress and channel orientation design," in *2004 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 2004, pp. 217 – 220.

[57] J.-S. Lim, S. Thompson, and J. Fossum, "Comparison of threshold-voltage shifts for uniaxial and biaxial tensile-stressed n-MOSFETs," *IEEE Electron Device Letters*, vol. 25, pp. 731 – 733, November 2004.

[58] K. Mistry, M. Armstrong, C. Auth, S. Cea, T. Coan, T. Ghani, T. Hoffmann, A. Murthy, J. Sandford, R. Shaheed, K. Zawadzki, K. Zhang, S. Thompson, and M. Bohr, "Delaying forever: Uniaxial strained silicon transistors in a 90nm cmos technology," in *2004 Symp. VLSI Technology Dig. of Tech. Papers*, June 2004, pp. 50 – 51.

[59] C.-H. Ge, C.-C. Lin, C.-H. Ko, C.-C. Huang, Y.-C. Huang, B.-W. Chan, B.-C. Perng, C.-C. Sheu, P.-Y. Tsai, L.-G. Yao, C.-L. Wu, T.-L. Lee, C.-J. Chen, C.-T. Wang, S.-C. Lin, Y.-C. Yeo, and C. Hu, "Process-strained si (pss) cmos technology featuring 3d strain engineering," in *2003 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 2003, pp. 3.7.1 – 3.7.4.

[60] S. Pidin, T. Mori, R. Nakamura, T. Saiki, R. Tanabe, S. Satoh, M. Kase, K. Hashimoto, and T. Sugii, "MOSFET current drive optimization using silicon nitride capping layer for 65-nm technology node," in *2004 Symp. VLSI Technology Dig. of Tech. Papers*, June 2004, pp. 54 – 55.

[61] V. Chan, R. Rengarajan, N. Rovedo, W. Jin, T. Hook, P. Nguyen, J. Chen, E. Nowak, X.-D. Chen, D. Lea, A. Chakravarti, V. Ku, S. Yang, A. Steegen, C. Baiocco, P. Shafer, H. Ng, S.-F. Huang, and C. Wann, "High speed 45nm gate length cmosfets integrated into a 90nm bulk technology incorporating strain engineering," in *2003 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 2003, pp. 3.8.1 – 3.8.4.

[62] S. Pidin, T. Mori, K. Inoue, S. Fukuta, N. Itoh, E. Mutoh, K. Ohkoshi, R. Nakamura, K. Kobayashi, K. Kawamura, T. Saiki, S. Fukuyama, S. Satoh, M. Kase, and K. Hashimoto, "A novel strain enhanced CMOS architecture using selectively deposited high tensile and high compressive silicon nitride films," in *2004 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 2004, pp. 213 – 216.

[63] S. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, M. Zhiyong, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, and Y. El-Mansy, "A logic nanotechnology featuring strained-silicon," *IEEE Electron Device Letters*, vol. 25, pp. 191 – 193, April 2004.

[64] C. Mazure and I. Cayrefourcq, "Status of device mobility enhancement through strained silicon engineering," in *Proc. 2005 IEEE Int. SOI Conference*, October 2005, pp. 1 – 6.

[65] G. Giusi, F. Crupi, E. Simoen, G. Eneman, and M. Jurczak, "Performance and reliability of strained-silicon nmosfets with sin cap layer," *IEEE Trans. on Electron Devices*, vol. 54, pp. 78 – 82, Jan 2007.

[66] K. Ota, K. Sugihara, H. Sayama, T. Uchida, H. Oda, T. Eimori, H. Morimoto, and Y. Inoue, "Novel locally strained channel technique for high performance 55nm CMOS," in *2002 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, December 2002, pp. 27 – 30.

[67] R. W. Keyes, "Explaining strain - the positive and negative effects of elastic strain in n-silicon," *IEEE Circuits and Devices Magazine*, vol. 18, pp. 36 – 39, September 2002.

[68] X. Wang, B. Cheng, S. Roy, and A. Asenov, "Simulation of strain enhanced variability in nMOSFETs," in *9th Int. Conference on Ultimate Integration of Silicon, 2008*, March 2008, pp. 89 – 92.

[69] J.-S. Goo, Q. Xiang, Y. Takamura, F. Arasnia, E. Paton, P. Besser, J. Pan, and L. Ming-Ren, "Band offset induced threshold variation in strained-Si nMOSFETs," *IEEE Electron Device Letters*, vol. 24, pp. 568 – 570, September 2003.

[70] H. Tuinhout, M. Pelgrom, R. P. de Vries, and M. Vertregt, "Effects of metal coverage on mosfet matching," in *1996 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, 1996, pp. 735–738.

[71] J. D. Plummer, M. D. Deal, and P. B. Griffin, *Silicon VLSI Technology.* Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[72] A. K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography.* Bellingham, WA, USA: SPIE Press, 2001.

[73] I. Matthew, C. E. Tabery, T. Lukanc, M. Plat, M. Takahashi, and A. Wilkison, "Design restrictions for patterning with off-axis illumination," in *Optical Microlithography XVIII*, ser. Proc. SPIE, B. W. Smith, Ed., vol. 5754, May 2004, pp. 1574–1585.

[74] T. A. Brunner, "Impact of lens aberrations on optical lithography," *IBM J. of Research and Development*, vol. 41, pp. 57–67, Jan/Mar 1997.

[75] T. Brunner, D. Corliss, S. Butt, T. Wiltshire, C. P. Ausschnitt, and M. Smith, "Laser bandwidth and other sources of focus blur in lithography," in *Optical Microlithography XIX*, ser. Proc. SPIE, D. G. Flagello, Ed., vol. 6154, March 2006, pp. 1574–1585.

[76] M. Choi and L. Milor, "Impact on circuit performance of deterministic within-die variation in nanoscale semiconductor manufacturing," *IEEE Trans. on Computer-Aided Design of Circuits and Systems*, vol. 25, no. 7, pp. 1350–1367, July 2006.

[77] T. Kjellberg and R. Schatz, "The effect of stitching errors on the spectral characteristics of dfb lasers fabricated using electron beam lithography," *J. Lightwave Technology*, vol. 10, no. 9, pp. 1256–1266, September 1992.

[78] D. Steele, A. Coniglio, C. Tang, and B. Singh, "Characterizing post exposure bake processing for transient and steady state conditions, in the context of critical dimension control," in *Metrology, Inspection and Process Control for Microlithography XVI*, ser. Proc. SPIE, D. Herr, Ed., vol. 4689, July 2002, pp. 517–530.

[79] I. Ahsan, O. Glushchenkov, R. Logan, E. J. Nowak, H. Kimura, G. Berg, J. Herman, E. Maciejewski, A. Chan, A. Azuma, S. Deshpande, B. Dirahoui, G. Freeman, A. Gabor, M. Gribelyuk, S. Huang, M. Kumar, K. Miyamoto, D. Mocuta, A. Mahorowala, E. Leobandung, and H. Utomo, "Rta-driven intra-die variations in stage delay, and parametric sensitivities for 65nm technology," in *2006 Symp. VLSI Technology Dig. of Tech. Papers*, 2006, pp. 170–171.

[80] Y.-M. Sheu, K.-W. Su, S. Tian, S.-J. Yang, C.-C. Wang, M.-J. Chen, and S. Liu, "Modeling the well-edge proximity effect in highly scaled MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 11, pp. 2792–279, November 2006.

[81] G. Eneman, M. Jurczak, P. Verheyen, T. Hoffmann, A. D. Keersgieter, and K. D. Meyer, "Scalability of strained nitride capping layers for future cmos generations," in *35th European Solid State Device Research Conference, 2005*, September 2005, pp. 449 – 452.

[82] S. Eneman, P. Verheyen, R. Rooyackers, F. Nouri, L. Washington, R. Degraeve, B. Kaczer, V. Moroz, A. D. Keersgieter, R. Schreutelkamp, M. Kawaguchi, Y. Kim, A. Samoilov, L. Smith, P. Absil, K. D. Meyer, M. Jurczak, and S. Biesemans, "Layout impact on the performance of a locally strained PMOSFET," in *2005 Symp. VLSI Technology Dig. of Tech. Papers*, June 2005, pp. 22 – 23.

[83] G. Eneman, P. Verheyen, R. Rooyackers, F. Nouri, L. Washington, R. Schreutelkamp, V. Moroz, L. Smith, A. D. Keersgieter, M. Jurczak, and K. D. Meyer, "Scalability of the si/sub 1-x/ge/sub x/ source/drain technology for the 45-nm technology node and beyond," *IEEE Trans. on Electron Devices*, vol. 53, pp. 1647 – 1656, July 2006.

[84] T. Sato, T. Uezono, S. Hagiwara, K. Okada, S. Amakawa, N. Nakayama, and K. Masu, "A MOS transistor-array for accurate measurement of subthreshold leakage variation," in *8th Int. Symp. on Quality Electronic Design, 2007*, March 2007, pp. 21 – 26.

[85] K. M. G.-V. Gettings, "Study of cmos process variation by multiplexing analog characteristics," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Mass., 2007.

[86] L.-T. Pang, J. B. Kuang, J. Sivagnaname, T. Nguyen, F. Gebara, J. Kao, G. Carpenter, , and K. Nowka, "45nm low voltage SRAM characterization," in *2007 Austin Conference on Energy Efficient Design*, 2007.

[87] A. Bhavnagarwala, S. Kosonocky, C. Radens, K. Stawiasz, R. Mann, Q. Ye, and K. Chin, "Fluctuation limits & scaling opportunities for CMOS SRAM cells," in *2005 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec 2005, pp. 659–662.

[88] Zhengguo, A. Carlson, L.-T. Pang, K. Duong, T.-J. K. Liu, and B. Nikolić, "Large-scale read/write margin measurement in 45nm CMOS SRAM arrays," in *2008 Symp. VLSI Circuits Dig. of Tech. Papers*, June 2008.

[89] J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, T. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, and D. Wendel, "Implementation of the CELL Broadband Engine in a 65nm SOI technology featuring dual-supply SRAM arrays supporting 6GHz at 1.3V," in *ISSCC 2007, Dig. of Tech. Papers*, February 2007, pp. 322–606.

[90] R. A. Bianchi, G. Bouche, and O. R. dit Buisson, "Accurate modeling of trench isolation induced mechanical stress effects on MOSFET electrical performance," in *2002 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec 2002, pp. 117–120.

[91] I. Ohkura, T. Noguchi, K. Sakashita, H. Ishida, T. Ichiyama, and T. Enomoto, "Gate isolation - a novel basic cell configuration for CMOS gate arrays," in *Custom Integrated Circuits Conference Dig. of Tech. Papers*, May 1982, pp. 307–310.

[92] D. Beale, J. Shiely, L. M. III, and M. Rieger, "Advanced model formulations for optical and process proximity correction," in *Optical Microlithography XVII*, ser. Proc. SPIE, B. W. Smith, Ed., vol. 5377, no. 12, May 2004, pp. 721–729.

[93] V. Kheterpal, V. Rovner, T. G. Hersan, D. Motiani, Y. Takegawa, A. J. Strojwas, and L. Pileggi, "Design methodology for ic manufacturability based on regular logic-bricks," in *Proc. Design Automation Conference*, Anaheim, CA, USA, June 2005, pp. 353–358.

[94] H. Onodera, "Variability: Modeling and its impact on design," *IEICE Trans. Electron.*, vol. E89-C, no. 3, pp. 342 – 348, March 2006.

Appendix A

Mathematics

A.1 Integrator equation derivations.

From Figure A.1, the following equations are obtained:

$$\frac{V_i}{Z_{Cp1}} + \frac{V_i - V_o}{Z_C} = I \tag{A.1}$$

$$G_m \cdot V_i + \frac{V_o}{R_o} + \frac{V_o}{Z_{Cp2}} + \frac{V_o - V_i}{Z_C} = 0 \tag{A.2}$$

From A.2

$$V_i \cdot (G_m - \frac{1}{Z_C}) = -V_o(\frac{1}{R_o} + \frac{1}{Z_{Cp2}} + \frac{1}{Z_C}) \tag{A.3}$$

From A.1

$$I = V_i(\frac{1}{Z_{Cp1}} + \frac{1}{Z_C}) - \frac{V_o}{Z_C} \tag{A.4}$$

Substitute A.3 into A.4

$$\frac{I}{V_o} = \left[\frac{\frac{1}{R_o} + \frac{1}{Z_{Cp2}} + \frac{1}{Z_C}}{G_m - \frac{1}{Z_C}}\right](\frac{1}{Z_{Cp1}} + \frac{1}{Z_C}) - \frac{1}{Z_C} \tag{A.5}$$

After simplification, we get

$$\frac{V_o}{I} = \frac{\frac{1}{G_m} - \frac{1}{jwC}}{1 + \frac{C+C_{p1}}{G_m \cdot R_o \cdot C} + jw \cdot \left[\frac{C_{p1} \cdot C_{p2} + C(C_{p1}+C_{p2})}{C \cdot G_m}\right]}$$

$$\Leftrightarrow \frac{V_o}{I} = \frac{\frac{1}{G_m} - \frac{1}{jwC}}{1 + A + jw \cdot B} \tag{A.6}$$

$$\tag{A.7}$$

where

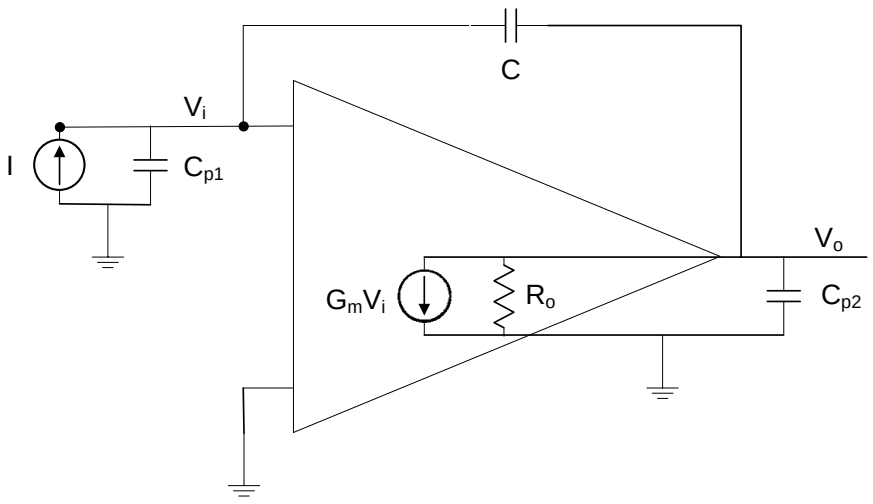


Figure A.1: Small signal model of the integrator.

$A = \frac{C+C_{p1}}{G_m \cdot R_o \cdot C}$

$B = \frac{C_{p1} \cdot C_{p2} + C(C_{p1}+C_{p2})}{C \cdot G_m}$

$G_m \approx 10mS$

$R_o \approx 1e5$

$C = 10pF \geq C_{p1}, C_{p2}$

Thus,

$A, B \leq 1e-3$

and,

$$V_o = \frac{1}{G_m} \cdot I - \frac{1}{C} \cdot \int I \cdot dt \tag{A.8}$$

where

V_o = integrator output voltage

G_m = transconductance gain of the amplifier for amplifier output voltage at V_o

I = integrated current

C = integrating capacitance

t = time

A.2 Least squares estimate, error estimate.

In this section, error analysis for the least squares estimate(LSM) method to infer the process parameters from measurement data is described.

In the LSM method,

$Y = M \cdot X + E$

where

$$Y = \begin{bmatrix} F_1 \\ I_1 \\ F_2 \\ I_2 \\ \vdots \end{bmatrix}_i, X = \begin{bmatrix} L \\ Tox \\ Nch \end{bmatrix}, M = \begin{bmatrix} \frac{\partial F_1}{\partial L} & \frac{\partial F_1}{\partial Tox} & \frac{\partial F_1}{\partial Nch} \\ \frac{\partial I_1}{\partial L} & \frac{\partial I_1}{\partial Tox} & \frac{\partial I_1}{\partial Nch} \\ \frac{\partial F_2}{\partial L} & \frac{\partial F_2}{\partial Tox} & \frac{\partial F_2}{\partial Nch} \\ \frac{\partial I_2}{\partial L} & \frac{\partial I_2}{\partial Tox} & \frac{\partial I_2}{\partial Nch} \\ \vdots & \vdots & \vdots \end{bmatrix},$$

$$E = \begin{bmatrix} eF_1 \\ eI_1 \\ eF_2 \\ eI_2 \\ \vdots \end{bmatrix} \sim N(0, \Sigma)$$

Y = Frequency and leakage current measurement data

X = Process parameters

M = Sensitivities

E = Errors associated with measurement. Assumed to be normally distributed with zero mean and covariance matrix Σ

Given the measurement data Y and simulation data M, X is estimated with minimum error. The following steps describes the estimation process and how the estimation errors can be obtained.

1. Whiten the noise

In the general case, the noise may not be white which means that Σ may not be the identity matrix(I_D). In order to minimize the error, the noise function E has to be converted to white Gaussian noise. This is easily done using well-established methods such as Graham-Schmitt decomposition.

If $\Sigma = \Sigma^{1/2} \cdot \Sigma^{1/2}$, $\widetilde{Y} = \Sigma^{-1/2} \cdot Y$, $\widetilde{M} = \Sigma^{-1/2} \cdot M$, $\widetilde{W} = \Sigma^{-1/2} \cdot E$,

We now have $\widetilde{Y} = \widetilde{M} \cdot X + W$ where $W \sim N(0, I_D)$

2. Least means squares estimate

   Using minimum least squares, an estimate of X is obtained.

$$\widehat{X} = \begin{bmatrix} \widehat{L} \\ \widehat{Tox} \\ \widehat{Nch} \end{bmatrix} = \left(\widetilde{M}' \cdot \widetilde{M}\right)^{-1} \cdot \widetilde{M}' \cdot \widetilde{Y}$$

3. Covariance matrix of $\widehat{X}$

   The error for each of the estimated process parameter is reflected in the covariance matrix of $\widehat{X}$. This is obtained as shown below.

$$Cov(\widehat{X}) = \left(\widetilde{M}' \cdot \widetilde{M}\right)^{-1} = \begin{bmatrix} Var(\widehat{L}) & Cov(\widehat{L}, \widehat{Tox}) & Cov(\widehat{L}, \widehat{Nch}) \\ Cov(\widehat{Tox}, \widehat{L}) & Var(\widehat{Tox}) & Cov(\widehat{Tox}, \widehat{Nch}) \\ Cov(\widehat{Nch}, \widehat{L}) & Cov(\widehat{Nch}, \widehat{Tox}) & Var(\widehat{Nch}) \end{bmatrix},$$

In the case of the 90nm measurements, we have

$$Y = M \cdot X + E$$

where E $\sim$ N(0, σ^2) and σ^2 is the variance of the white measurement noise.

We get

$$\frac{cov(X)}{\sigma^2} = \left(\widetilde{M}' \cdot \widetilde{M}\right)^{-1} = \begin{bmatrix} 2.64 & -7.44 & -54.3 \\ -7.44 & 23.5 & 149 \\ -54.3 & 149 & 1155 \end{bmatrix},$$

We see that the error is smallest for L, $\approx 10 \times$ bigger for Tox and $\approx 50 \times$ bigger for Nch.

Appendix B

Layouts

This appendix aims to provide information on the 90nm and 45nm test-chip designs that would enable a circuit designer to examine the cadence libraries and understand the structure of the designs.

B.1 Arrangements of the Layouts in a Tile.

Please refer to the cells testchip/bigtile for 90nm layouts and 45nmRO_leak_ST/bigtile2 for 45nm layouts. For testchip/bigtile (90nm), the floor-plan is given in Figure B.1. For 45nmRO_leak_ST/bigtile2 (45nm), the floor-plan is given in Figure B.2.

B.2 Library cells in 90nm.

This section describes the setup for the 90nm design kit and the location of the 90nm test-chip library. Several cells in the higher levels of hierarchy are described.

B.2.1 90nm Library Setup.

Sourced setup file: /vol/hitz/home/ltpang/cshrc.st90nm
Library paths are located in: /tools/designs/Dcdg/users/ltpang/project/st90nm/cds.lib

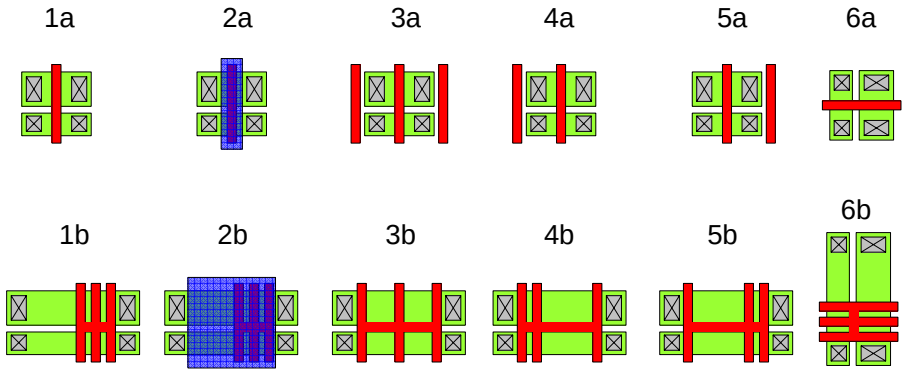
B.2.2 90nm Library.

Library Name: testchip

fullchip: Full chip minus the dummies.

padring2: Cell fullchip minus the sealring. This cell contains the I/O cells and their routing.

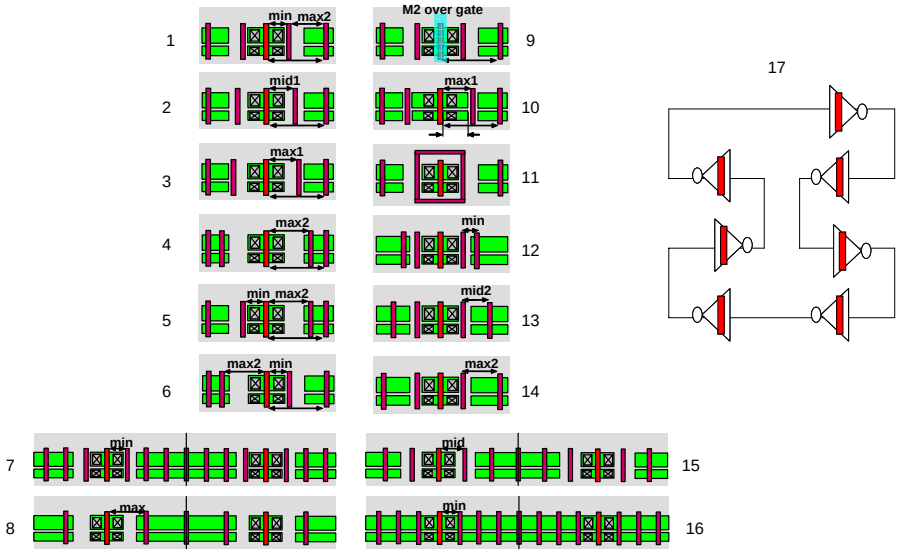
topcore: Contains the core of the chip. Main cells are tile_array and topintbig2. Also contains dummies.



bigtile (90nm)

RO 1a	Leakage NMOS 1a	RO 1b	Leakage NMOS 1b
RO 2a	Leakage NMOS 2a	RO 2b	Leakage NMOS 2b
RO 3a	Leakage NMOS 3a	RO 3b	Leakage NMOS 3b
RO 4a	Leakage NMOS 4a	RO 4b	Leakage NMOS 4b
RO 5a	Leakage NMOS 5a	RO 5b	Leakage NMOS 5b
RO 6a	Leakage NMOS 6a	RO 6b	Leakage NMOS 6b

Figure B.1: Arrangement of the 12 layouts in a tile for the 90nm test-chip.



Bigtile2 (45nm)

RO 1	Leakage 1	RO 9	Leakage 9	RO 17	Leakage 17
RO 2	Leakage 2	RO 10	Leakage 10		
RO 3	Leakage 3	RO 11	Leakage 11		
RO 4	Leakage 4	RO 12	Leakage 12		
RO 5	Leakage 5	RO 13	Leakage 13		
RO 6	Leakage 6	RO 14	Leakage 14		
RO 7	Leakage 7	RO 15	Leakage 15		
RO 8	Leakage 8	RO 16	Leakage 16		

Figure B.2: Arrangement of the 17 layouts in a tile for the 45nm test-chip.

tile_array: Contains array_row.

array_row: Contains a row of 4bigtiles and row dividers.

4bigtiles: Contains 4 tiles of RO and I_{LEAK} transistors and a VRO tile.

bigtile: Contains tile1, tile2, tile3, tile4, tile5, tile8. tile1-5 correspond to layouts 1a/b to 5a/b and tile8 corresponds to layout 6a/b.

vro_all: VRO tile.

topintbig2: Contains the integrators and logic for current measurement.

integrp: Integrator for PMOS current measurement.

integrn: Integrator for NMOS current measurement.

Library Name: finalchip

DUM_fullchip: Full chip with dummy fills. This is the one that is taped out.

fullchip_DUMMIES: Dummy fills for the whole chip.

fullchip: The full 90nm test-chip without dummy fills

B.3 Library cells in 45nm.

This section describes the setup for the 45nm design kit and the location of the 45nm test-chip library for the RO and leakage transistor array and the leakage measurement circuits. Several cells in the higher levels of hierarchy are described.

B.3.1 45nm Library Setup.

sourced setup file: /tools/designs/Dcdg/users/ltpang/st45nm2/st45nm2/.cshrc_cmos045
Library paths are located in: /tools/designs/Dcdg/users/ltpang/st45nm2/st45nm2/cds.lib

B.3.2 45nm Library.

Library Name: 45nmRO_leak_ST

top_array_leakmeas: Contains the RO and leakage transistor array with the leakage current measurement circuits.

top_array_leakmeas_lvs: Contains the leakage measurement circuit and routing.

Ileakmeas: Contains the 2 integrators, multiplexers and routing.

integrp_final: Integrator for PMOS current measurement.

integrn_final: Integrator for NMOS current measurement.

row18x16_20: Contains the RO and leakage array made up of 18 rows of row1x16_20 and the row multiplexer.

row1x16_20: Contains the a row of tiles and scan chain flip-flops at one end of the row and multiplexers at the other end.

tile1x16_20: Contains 16 bigtile2 in a row.

bigtile2: contains tile1-9, tile12-14, tile20-23. tile1-6 correspond to layout1-6 in Figure **??**, tile20-21 to layout7-8, tile7-9 to layout9-11, tile12-14 to layout12-14, tile22-23 to layout15-16, and tile_vert to layout17.

Library Name: st45_padring

padring4: The 45nm padring.

Appendix C

Test Setup

The chips are packaged in ceramic PGA packages and measured through a socket mounted on a custom PCB. The PCB contains components that generate regulated supply voltages and analog bias currents from external 5V and 10V supplies. Control signals are sourced from a pattern generator(Tektronix DG2020A, Figure C.1) and accurate reference currents are generated with a source-meter (Keithley 2400, Figure C.2). A 20 GS/s oscilloscope (Agilent Infiniium, Figure C.3) is used to make frequency measurements and a 150ps resolution logic analyzer(Tektronix TLA5202, Figure C.4) is used to measure the delay between the start and stop signals of the ADC. In order to improve accuracy and reduce the impact of noise, several hundred measurements are averaged for each data point.

Perl scripts and small programs in C are written to automate the process of data collection. These scripts send commands to the instruments using a LAN/GPIB interface (National Instruments GPIB-ENET/100, Figure C.5) and write measurement data onto a file.

This appendix will briefly describe a few instruments used to carry out the measurement procedure. Section C.1 gives a few pointers to how the GPIB-ENET/100 box is set up. Section C.2 describes how the pattern generator is programmed to automate data collection and section C.3 describes the procedures and set-up for data collection.

C.1 GPIB-ENET/100 Guide

Several instruments connect to each other via GPIB interface to automate the measurement process. A computer sends out commands to the GPIB interface via the LAN network. This is done via a GPIB-ENET/100 box from National Instruments shown in Figure C.5. This section describes how the GPIB-ENET/100 box is set up.

1. Register the GPIB-ENET/100 box with Brad or Kevin. Obtain the following information from them.

   hostname: nigpib-3.EECS.Berkeley.EDU
   IP address: 128.32.62.232
   netmask: 255.255.255.0

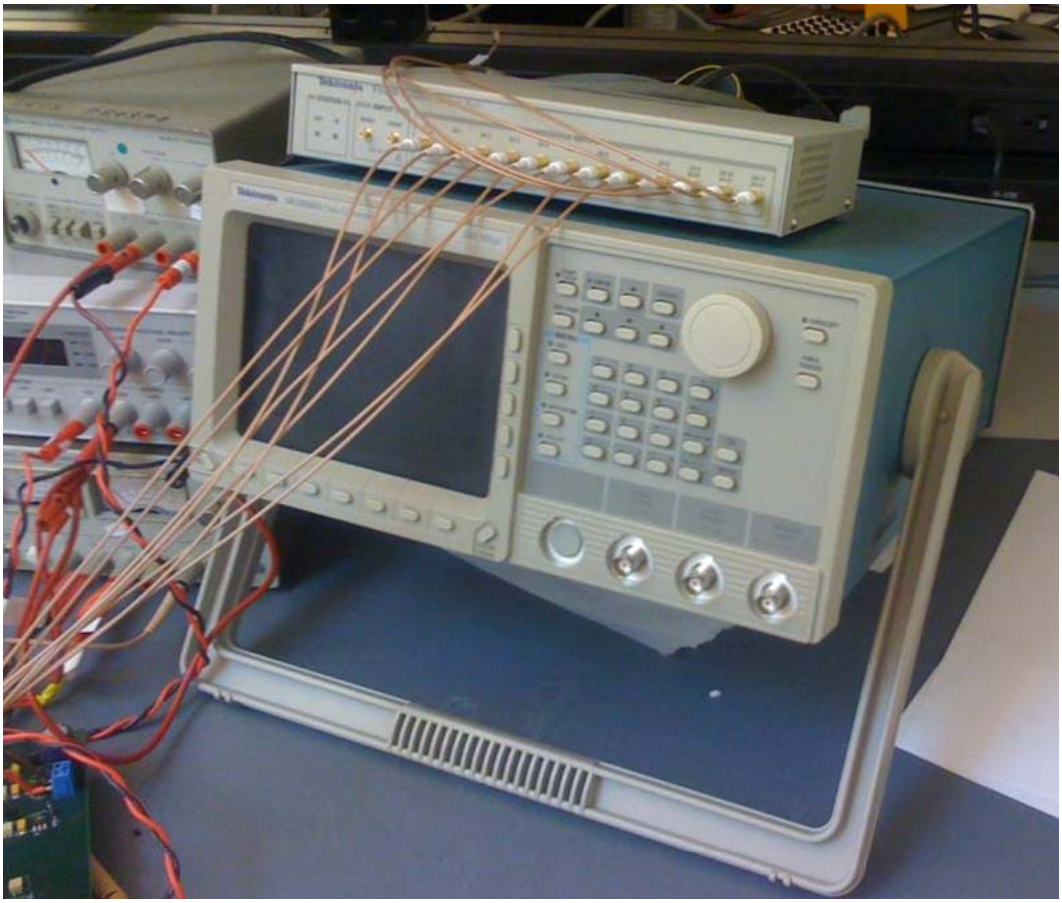


Figure C.1: DG2020A data generator from Tektronix.

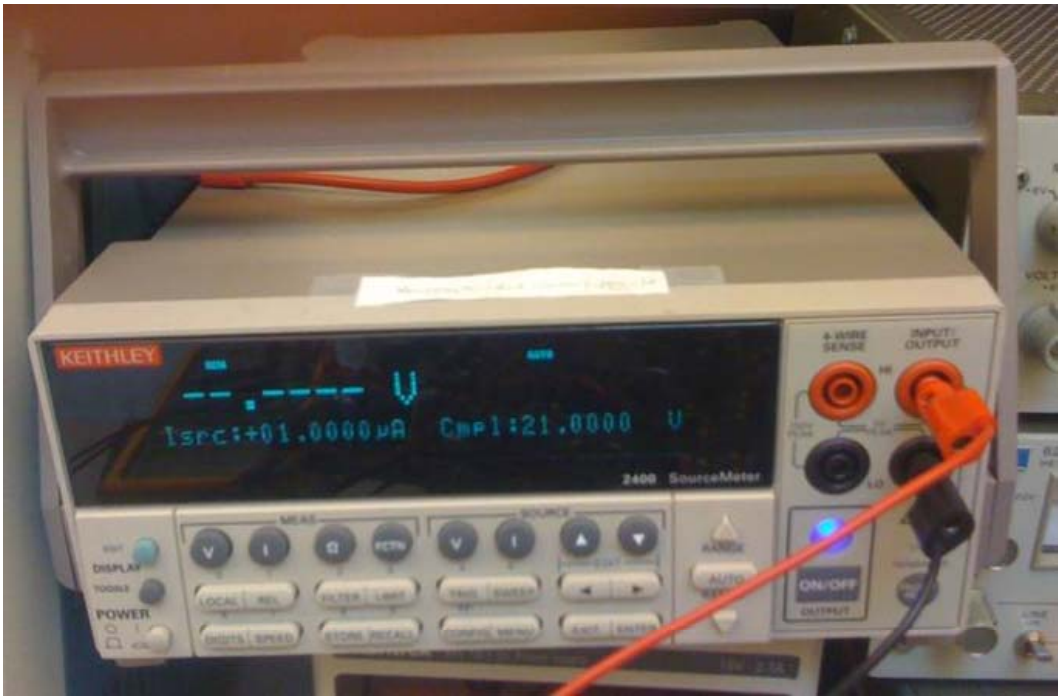


Figure C.2: Keithley 2400 source meter.

Figure C.3: Infiniium 54855A DSO 20GSPS oscilloscope from Agilent.

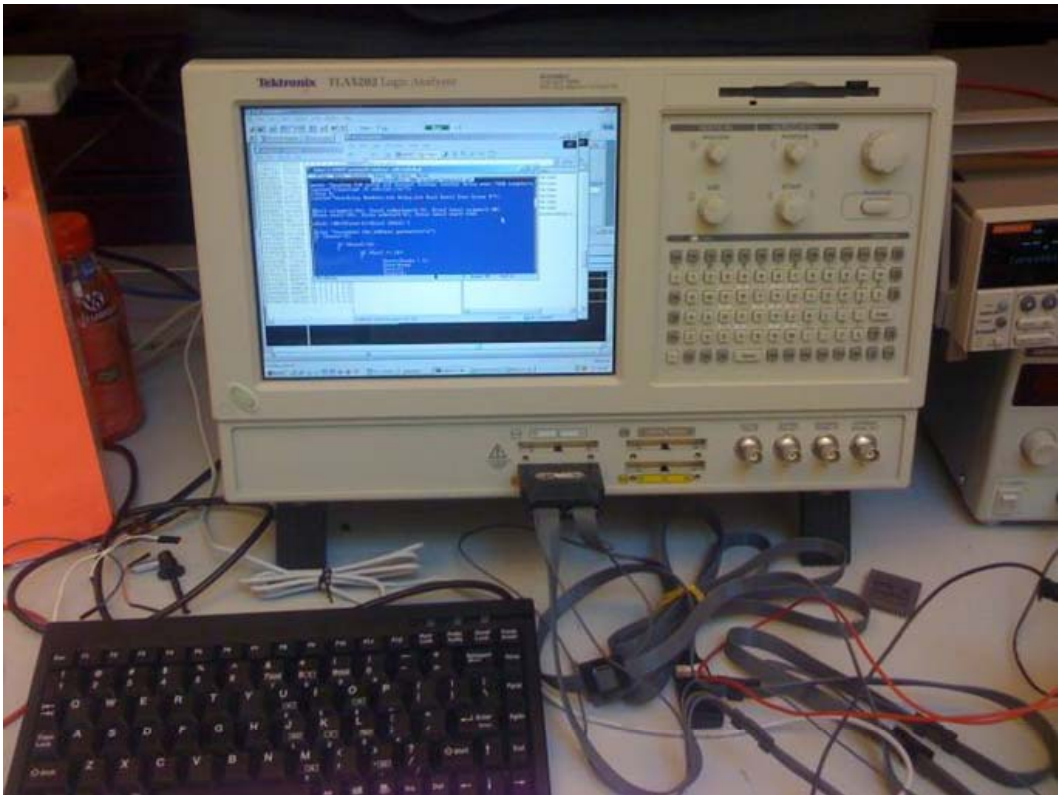


Figure C.4: TLA5202 logic analyzer from Tektronix.

Figure C.5: GPIB-ENET/100 box from National Instruments.

gateway: 128.32.62.2
DNS: 128.32.62.21

2. See chapter 3 of National Instruments GPIB Manual 'Getting Started with Your GPIB-ENET/100 and NI-488.2 for Windows 2000/NT'.
   Select 'Start>>Programs>>NI-488.2M for Windows NT>>
   GPIB-ENET100 Utilities>>Device Configuration'

   If the device is registered, you'll see it (verify Ethernet address). It should have an IP address/hostname. Currently the devices that have been registered are

   128.32.62.232 00:80:2f:10:0b:5f 010663FE 1st setup
   128.32.62.245 00:80:2f:10:0a:75 01065A05 Louis using
   128.32.62.250 00:80:2f:10:0a:70 0106599A 2nd setup

   If device is not configured, click properties and fill in the information according to what is provided in 1.

3. See chapter 4. Select 'Start >>Settings>>Control Panel' and double click the GPIB icon. The GPIB configuration dialog box appears. Select GPIB0, click Board Type and select GPIB-ENET/100, click 'ok', Click 'Configure' and fill in the IP address of the device you want to control (eg. 128.32.62.250). Click 'ok'.

4. See chapter 5. Select 'Diagnostic' item under 'Start>>Programs>>NI-488.2M' for Windows NT. Click 'Test All'. If all is good, it will indicate 'passed' under 'Status'.

C.2 DG2020A Pattern Generator

The DG2020A data generator from Tektronix (Figure C.1) is used to provide the digital input signals for the inputs and clocks of the scan chains, the control signals for the current integrator and a few other control signals. This is described in Section C.3. This section will briefly describe how the DG2020A is set up and programmed to automate the collection of data. More information can be found in the DG2020A manual.

C.2.1 Configuring the DG2020A

1. Set to enhanced mode
2. Set voltage levels to the chip's I/O pads
3. Set pod assignments
4. Set oscillation frequency

C.2.2 Programming the DG2020A

1. **putgrp** GRPDEF.DAT

   Defines the groups of signals and assigns them to the output pins. GRPDEF.DAT is a text file that contains the assignments. An example of the contents of GRPDEF.DAT is given below. p1a is assigned to pins 0 - 0, sclk to 7-7, etc.

   sclk,7,7
   nsel,6,6
   incol,5,5
   p2,4,4
   inrow,3,3
   p1b,2,2
   p1,1,1
   p1a,0,0

2. **putword** shiftblk.dat

   Defines the bit pattern. An example of the contents of shiftblk.dat is given below. The first line starts with a # symbol and indicate the number of lines in the bit pattern. The second line also starts with a # symbol and indicate the starting address as 0. The next 64 lines show how the bit pattern changes over 64 cycles. There are 40 channels by default in the DG2020A and they are represented by 10 hexadecimal numbers. Each hexadecimal number represents 4 bits of digital output. The example below shows the last 8 outputs changing which is sufficient for the 45nm test-chip.

   # size 64
   # start 0
   00,00,00,00,45
   00,00,00,00,C5
   00,00,00,00,45
   00,00,00,00,45
   |
   |
   00,00,00,00,45

3. **putblk** BLKDEF.DAT

   Names the portion of waveform (block). In eg below, 0-63 time steps form a block called SHIFT, 64-127 form INCOL1 and so on.

   0,SHIFT
   64,INCOL1
   128,INROW2
   192,COL2ROW
   256,COL1ROW
   320,REST

384,NONE
448,MNMOS
512,MBNMOS
576,INROW2CK
640,INROW1
704,COL1ROW1

4. **putseq** file.dat

   Defines the pattern that is to be used. In eg below, start by sending the block SHIFT 400 times, followed by INCOL1 1 time, and so on until MNMOS 1 time and that is repeated until the DG2020A is stopped.

   SHIFT,400,0,0,0,0
   INCOL1,1,0,0,0,0
   INROW1,1,0,0,0,0
   MNMOS,1,0,0,0,1

   The contents of the input file must express the following in ASCII with one step per line.

   - block name
   - repeat count (1 to 65536)
   - line number of the event jump destination
   - trigger wait on/off setting(ON=1, OFF=0)
   - event jump on/off setting (ON=1, OFF=0)
   - infinite loop on/off setting (ON=1, OFF=0)

   This format corresponds to the parameter block supplied to the DATA:SEQUENCE:DEFINE command with the delimiter codes replaced by the newline code used in normal text files. The first line in this file corresponds to sequence line number 0, and each following line to the sequence line number incremented by 1. See the seqdef.dat sample data file below.

   (seqdef.dat)

   LEAD,1,5,OFF,0,0
   LOOP1,256,5,OFF,ON,0
   LOOP2,64,5,0,ON,0
   LOOP1,256,5,ON,1,0
   SUB1,1,5,ON,1,0
   SUB2,1,5,ON,1,0
   TRAIL,1,5,ON,0,0
   TRAIL,1,5,0,0,0

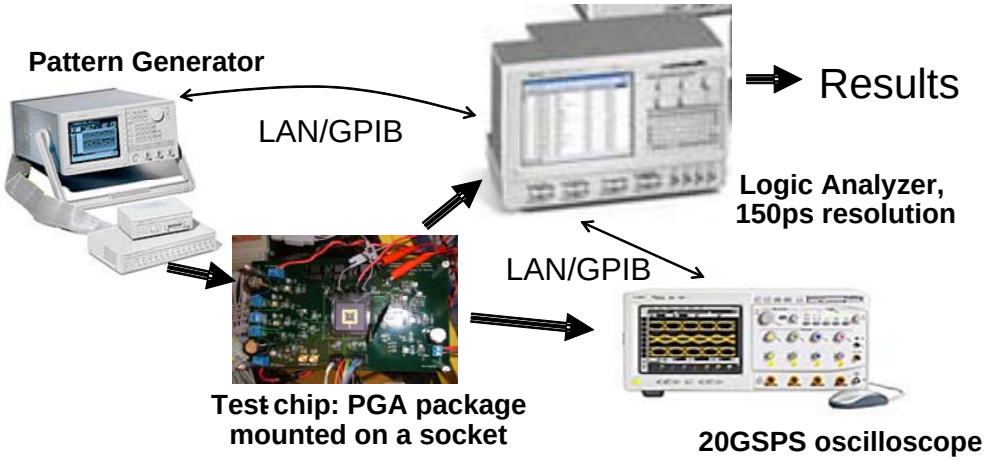


Figure C.6: Test setup for collecting data.

C.3 Measurement Setup and Procedures

Figure C.6 shows the setup for data collection. The pattern generator generates the address bits and the control signals while the logic analyzer measures the leakage current integration time and the oscilloscope measures the frequency of the ROs. Commands are sent to the instruments via the GPIB and LAN interfaces.

The PCB boards for measuring the 90nm and the 45nm test-chips are shown in figure C.7 and figure C.8 respectively. The packaged chip is mounted on a socket which is soldered onto the PCB. The socket has a lever that locks the packaged chip in place or releases it. This allows for easy changing of chips and facilitates the measurement of many chips. Components on the board include voltage regulators, current sources, capacitances, resistances, inductances, connectors, pins, switches, etc.

In the BWRC laboratory, two work spaces are used to shorten measurement time for the 45nm test-chips. One is used for leakage current measurement and the other for RO frequency measurement. Figure C.9 shows the test-bench for leakage current measurement. The TLA5202 is also a PC with its own hard disk and LAN connections. The Perl script is executed from it and the data and setup files are stored in its disk space. In the setup for RO frequency measurement, a PC is used instead and only the oscilloscope, the DG2020A and the supplies are required.

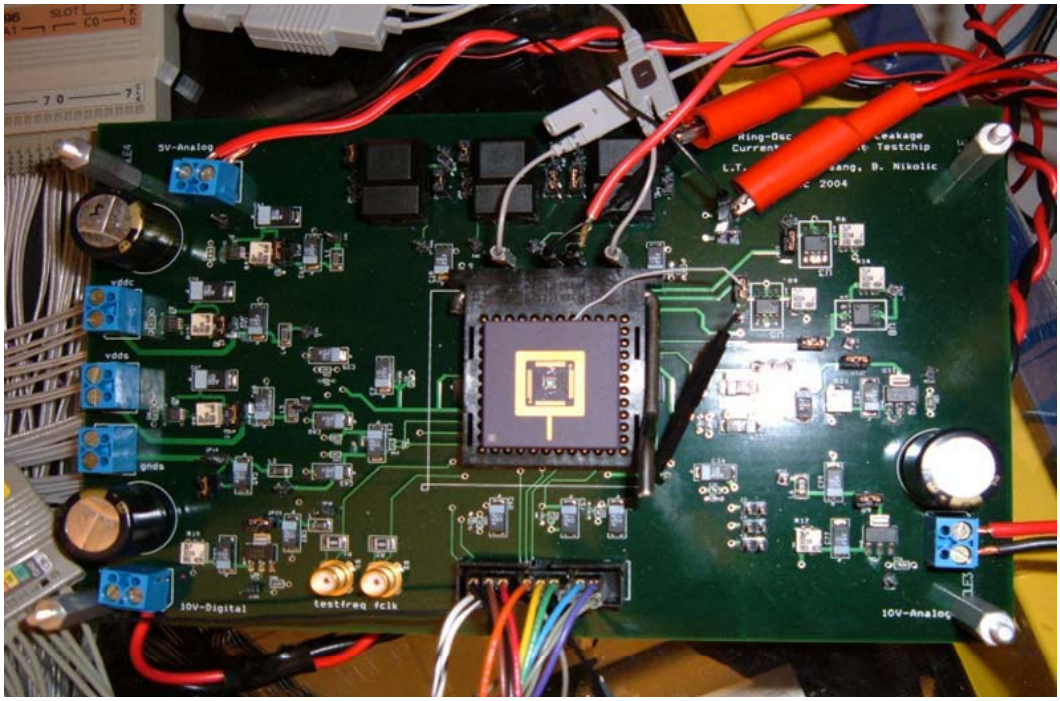


Figure C.7: PCB board for measuring the 90nm test-chips.

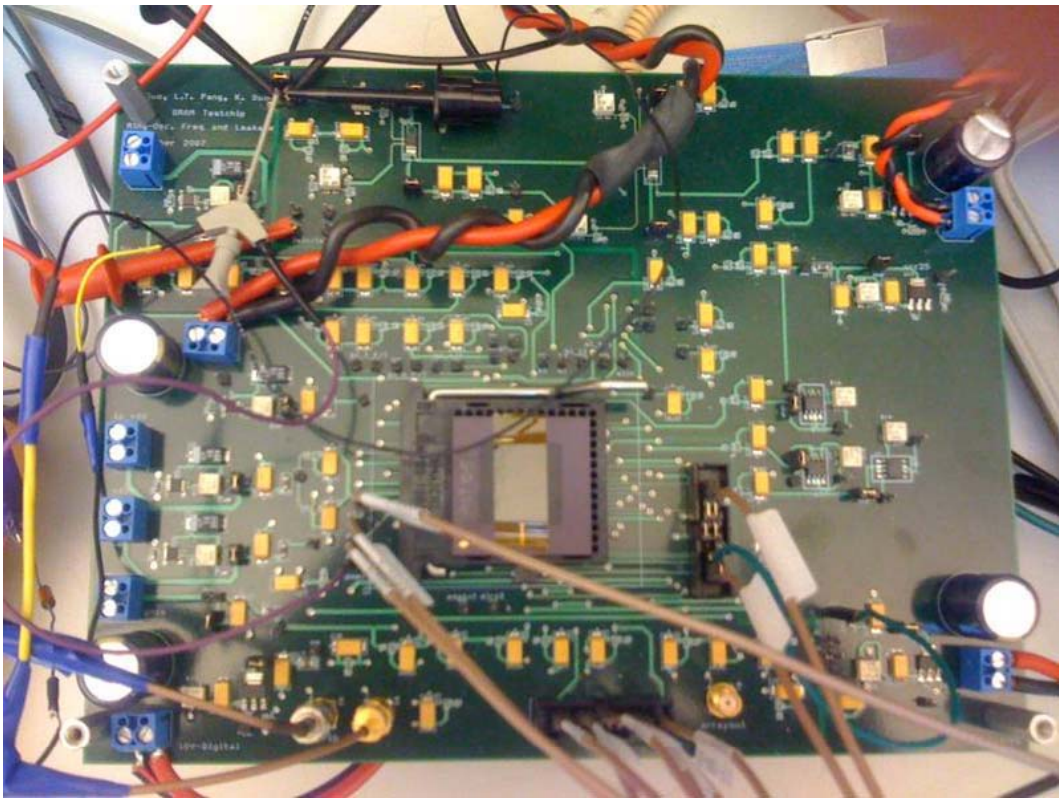


Figure C.8: PCB board for measuring the 45nm test-chips.

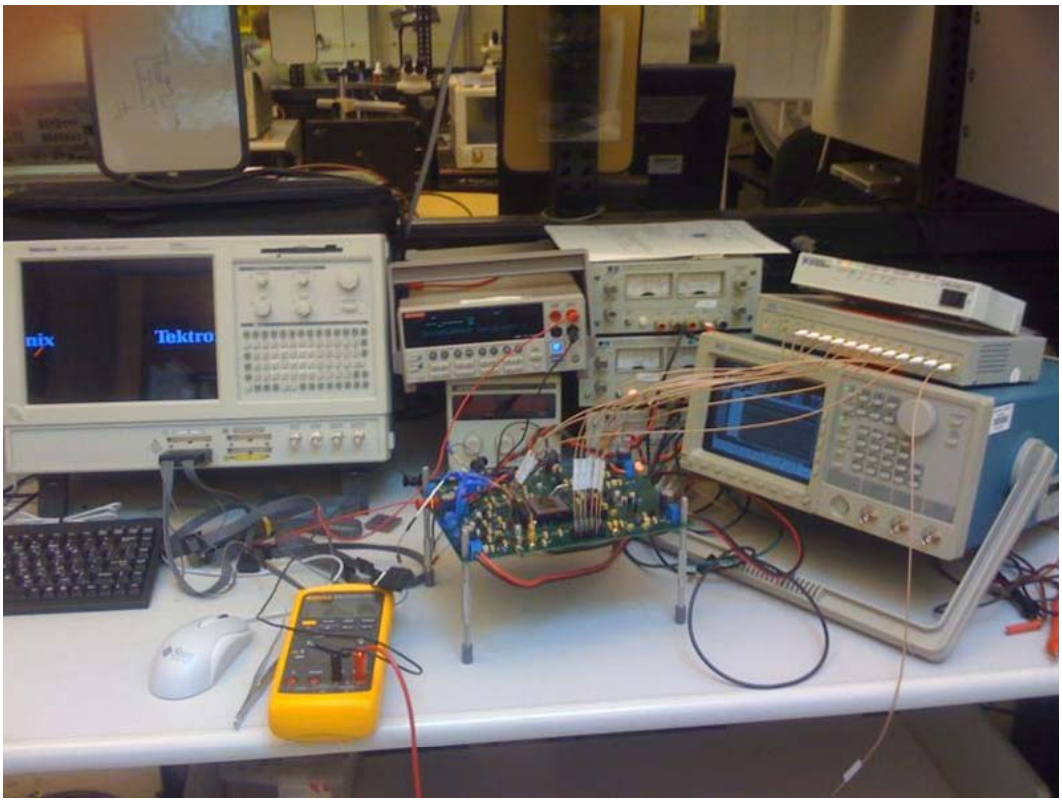


Figure C.9: Test setup for collecting leakage measurement data for the 45nm test-chips.