UNIVERSITY OF CALIFORNIA

Los Angeles

Optimization of CMOS for High Performance, Low Temperature Operation

A thesis submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical Engineering

by

Clifford Y.C. Hwang

1999

© Copyright by

Clifford Y.C. Hwang

1999

The thesis of Clifford Y.C. Hwang is approved.

Ian McLean

Dee-Son Pan

Jason C.S. Woo, Committee Chair

University of California, Los Angeles

1999

Contents

1 Introduction **1**

1.1 Background 1

1.2 Motivation 2

1.3 Organization 3

2 Performance and Reliability of Conventional CMOS at Low Temperatures **5**

2.1 Introduction 5

2.2 Low Temperature Performance 6

2.2.1 Device and Setup Details 6

2.2.2 Measurement Results 6

2.2.3 Projected Circuit Performance 9

2.3 Performance Limitations at Low Temperature 11

2.3.1 Threshold Voltage Shift 11

2.3.2 PMOS Carrier Freezeout 14

2.3.3 LDD Freezeout 17

2.4 Low Temperature Reliability 18

2.4.1 Stress Conditions 20

2.4.2 Lifetime Results 20

2.5 Conclusion 22

3 Calibration of TCAD Device Simulation Tools for Low Temperature CMOS 24

3.1 Introduction 24

3.2 Required Pre-Calibration Data 25

3.2.1 Device Measurements 25

3.2.2 Doping Profiles 25

3.3 NMOS Fit 27

3.3.1 I_{ds} vs. V_{gs} At Low V_{ds} 28

3.3.2 I_{ds} vs. V_{ds} With V_{gs} Varied 31

3.3.3 I_{ds} vs. V_{gs} With V_{bs} Varied 36

3.4 PMOS Fit 38

3.4.1 I_{ds} vs. V_{ds} With V_{gs} Varied 38

3.4.2 I_{ds} vs. V_{ds} With V_{gs} Varied 42

3.4.3 I_{ds} vs. V_{gs} With V_{bs} Varied 42

3.5 Conclusion 45

4 Redesigning CMOS for Improved Low Temperature Performance 48

4.1 Introduction 48

4.2 Design Methodology 49

4.3 Threshold Voltage Adjustment 49

4.3.1 To Room Temperature Value 49

4.3.2 Below Room Temperature Value 55

4.4 Retrograde Channel Profiling 58
4.5 Source/Drain Engineering 61
4.5.1 Halo Implant 61
4.5.2 LDD Implant 63
4.6 Gate Engineering 65
4.7 Conclusion 67

5 Performance of Optimized Low Temperature CMOS 69

5.1 Introduction 69
5.2 Test Chip Design 70
5.2.1 Process Test Structures 70
5.2.2 Transistors 73
5.2.3 Circuits 74
5.3 Process Flow 75
5.4 Process Splits 78
5.5 Low Temperature NMOS Performance 81
5.5.1 Low Power Performance 85
5.5.2 Retrograde Profile Performance 87
5.5.3 Halo Implant Performance 88
5.5.4 Threshold Voltage Considerations 93
5.6 Low Temperature PMOS Performance 101
5.6.1 Boron Penetration Effects 104
5.6.2 Freezeout Design Issues 105
5.7 Low Temperature Ring Oscillator Performance 108
5.8 Conclusion 111

6 Conclusion **113**

6.1 Summary 113

6.2 Suggestions for Future Work 114

List of Figures

2.1 I_{ds}-V_{gs} characteristics of a 25 μm x 25 μm NMOS device at 300 K, 218 K, 150 K, and 78 K. 7

2.2 I_{ds}-V_{ds} characteristics of a 25 μm x 0.9 μm NMOS device at 343 K, 300 K, 218 K, 150 K, and 78 K. 7

2.3 Improvement in drain saturation current versus drawn channel length for different temperatures. 8

2.4 Layout of a single CMOS inverter using MAGIC. 9

2.5 Simulated waveform for a 41 stage, 1 μm ring oscillator at 300 K. . . 10

2.6 Simulated propagation delay per stage versus power supply of a 41 stage, 1 μm ring oscillator at 300 K and 78 K. 11

2.7 Simulated propagation delay improvement per stage versus power supply of a 1 μm ring oscillator at different temperatures. 12

2.8 Threshold voltage versus temperature for a 25 μm x 1 μm device. . . 13

2.9 Energy band diagram for an n^+-poly PMOS device biased near the threshold voltage. 15

2.10 Energy band diagram for an n^+-poly NMOS device biased near the threshold voltage. 16

2.11 Subthreshold slope versus temperature for a 25 μm x 1 μm device. . . 16

2.12 Energy band diagram in the LDD region of an NMOS device. 17
2.13 Linear drain current versus temperature for a 25 μm x 1 μm device. . 19
2.14 Parasitic source/drain resistance used in HSPICE ring oscillator simulations for different temperatures. 19
2.15 Lifetime versus $1/V_{ds}$ for 25 μm x 1 μm NMOS devices at 78 K. . . . 21
2.16 Maximum V_{dd} at different temperatures so that a 10 year device lifetime is maintained and corresponding circuit-level performance improvement. 22

3.1 SUPREM-III doping profiles underneath the gate, LDD, and source/drain regions of an NMOS device. 26
3.2 SUPREM-III doping profiles underneath the gate, LDD, and source/drain regions of a PMOS device. 27
3.3 Parasitic source/drain resistance used in MEDICI to fit the measured data at different temperatures. 30
3.4 Simulated and measured I_{ds}-V_{gs} characteristics of NMOS transistors at 78 K (linear scale). 32
3.5 Simulated and measured I_{ds}-V_{gs} characteristics of NMOS transistors at 78 K (log scale). 33
3.6 Simulated and measured I_{ds}-V_{gs} characteristics of a 25 μm x 1 μm NMOS transistor at various temperatures. 33
3.7 Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm NMOS transistor at 78 K. 34
3.8 Simulated and measured I_{ds}-V_{ds} characteristics of NMOS transistors at 78 K. 35

3.9 Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm NMOS transistor at various temperatures. 35
3.10 I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 300 K. 36
3.11 I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 218 K. 37
3.12 I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 150 K. 37
3.13 I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 78 K. 38
3.14 Simulated and measured I_{ds}-V_{gs} characteristics of PMOS devices at 300 K (linear scale). 39
3.15 Simulated and measured I_{ds}-V_{gs} characteristics of a 25 μm x 1 μm PMOS transistor at different temperatures. 40
3.16 Simulated and measured I_{ds}-V_{gs} characteristics of a 25 μm x 1 μm PMOS transistor at different temperatures (log scale). 41
3.17 Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm PMOS transistor at 300 K. 43
3.18 Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm PMOS transistor at 300 K and 78 K. 43
3.19 Simulated and measured I_{ds}-V_{ds} characteristics of PMOS transistors at 218 K. 44
3.20 I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm PMOS device at 300 K. 44
3.21 I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm PMOS device at 78 K. 45

4.1 Illustration of how TCAD process (SUPREM), device (MEDICI), and circuit (HSPICE) simulation tools are used together to improve the performance of low temperature CMOS. 50

4.2 Threshold adjustment implant doses (boron) necessary to maintain a 300 K V_{th} for a given operating temperature. 50
4.3 Improvement in linear drain current of a 25 μm x 1 μm NMOS device at 78 K after threshold adjustment. 52
4.4 Improvement in saturation drain current of a 25 μm x 1 μm NMOS device at 78 K after threshold adjustment. 52
4.5 Propagation delay versus power supply voltage of a 41 stage ring oscillator at 78 K after threshold implant adjustment. 53
4.6 Improvement in ring oscillator delay versus power supply voltage as a result of threshold implant adjustment. 54
4.7 Improvement in delay at 78 K versus power supply voltage for a lower than room temperature threshold voltage. 56
4.8 Improvement in delay at 218 K versus power supply voltage for a lower than room temperature threshold voltage. 57
4.9 Optimized threshold voltage for NMOS transistors at 78 K. 57
4.10 Effective electron mobility versus field at 300 K and 78 K (from ref. 16). 60
4.11 SUPREM-III doping profiles underneath the gate for different implant species and anneals. 61
4.12 Illustration of a halo implant for an NMOS device. 62
4.13 Improvement in CMOS ring oscillator performance at 78 K if no freezeout occurs. 64

5.1 Test chip used to fabricate devices and circuits for low temperature operation. 71

5.2 Process test structure portion of the low temperature CMOS test chip. 72

5.3 n^+-poly NMOS transistor array and SEM bar. 73

5.4 Circuit blocks designed for the low temperature CMOS test chip. . . 75

5.5 Cross section of fabricated NMOS and PMOS devices. 79

5.6 Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.25 μm NMOS transistor at different temperatures (linear scale). 82

5.7 Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.25 μm NMOS transistor at different temperatures (log scale). 82

5.8 Measured threshold voltage and subthreshold slope of NMOS transistors at 300 K and 78 K. 83

5.9 Saturation current versus drawn channel length for NMOS transistors at 300 K and 78 K. 83

5.10 I_{ds}-V_{ds} characteristics of a 10 μm x 0.25 μm NMOS transistor at 300 K and 78 K. 84

5.11 Improvement in saturation current for NMOS transistors (normalized to 300 K) 85

5.12 Improvement in saturation current at different temperatures for NMOS transistors of varying channel lengths. 86

5.13 Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.52 μm NMOS transistor with indium implanted channel at different temperatures (log scale). 88

5.14 Measured I_{ds}-V_{ds} characteristics of a 10 μm x 0.24 μm NMOS transistor with indium implanted channel at 300 K and 78 K. 89

5.15 Measured threshold voltage and subthreshold slope of NMOS transistors with different threshold implants at 300 K. 90
5.16 Improvement in saturation current for NMOS transistors with different threshold implants at 78 K. 90
5.17 Improvement in saturation current for NMOS transistors with different retrograded profiles at 78 K. 91
5.18 Threshold voltage rolloff for NMOS transistors with different threshold implants at 78 K. 92
5.19 Improvement in saturation current for NMOS transistors with different halo implants at 78 K. 93
5.20 Improvement in saturation current for NMOS transistors with different halo implants at 78 K. 94
5.21 Improvement in saturation current for NMOS transistors with different halo implants at 78 K. 95
5.22 Threshold voltage and subthreshold slope versus channel length for NMOS transistors with different halo implants at 300 K. 96
5.23 Improvement in saturation current for NMOS transistors at gate biases above the threshold voltage. 97
5.24 I_{ds}-V_{gs} characteristics (log scale) of two 10 μm x 0.44 μm NMOS transistors at 300 K and 78 K. 98
5.25 Improvement in saturation current at 78 K for NMOS transistors optimized for 300 K and 78 K operation. 99
5.26 I_{ds}-V_{gs} characteristics (log scale) of a 10 μm x 0.44 μm NMOS transistor (optimized for 300 K operation) at 300 K and 78 K. 100

5.27 I_{ds}-V_{gs} characteristics (log scale) of a 10 μm x 0.5 μm PMOS transistor at different temperatures. 101
5.28 I_{ds}-V_{ds} characteristics of a 10 μm x 0.5 μm PMOS transistor at different temperatures. 102
5.29 Improvement in saturation current versus channel length for PMOS transistors at different temperatures. 103
5.30 I_{ds}-V_{gs} characteristics (log scale) of 10 μm x 0.5 μm PMOS transistors at 300 K using different poly-gate structures. 104
5.31 Improvement in saturation current versus channel length for PMOS transistors with N_2 implanted poly at different temperatures. 106
5.32 Improvement in saturation current versus channel length for PMOS transistors with nitrous oxide gates (N_2O) at 78 K. 107
5.33 I_{ds}-V_{gs} characteristics (log scale) of a 10 μm x 0.44 μm PMOS transistor with N_2O oxide at different temperatures. 107
5.34 Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.48 μm PMOS transistor at different temperatures (linear scale). 108
5.35 Propagation delay per stage of a 0.5 μm 61 stage ring oscillator at 300 K, 218 K, and 78 K. 109
5.36 Measured and predicted propagation delay per stage of a 0.5 μm 61 stage ring oscillator at 78 K. 110
5.37 Improvement in propagation delay versus power supply of 0.5 μm 83 stage, 2-input NOR ring oscillators with different channel profiles at 78 K. 110

5.38 Improvement in propagation delay versus power supply of a 0.5 μm 83 stage, 2-input NOR ring oscillator with retrograded channel at 78 K. 111

List of Tables

3.1 Parameters used in NMOS MEDICI simulations. 46

3.2 Parameters used in PMOS MEDICI simulations. 47

5.1 Masks layers used in the low temperature CMOS test chip. 71

5.2 Process flow for low temperature CMOS devices and circuits. 76

5.3 Process splits for low temperature CMOS devices and circuits. 80

5.4 Process splits for low temperature CMOS devices and circuits (continued). 81

5.5 Parasitic source/drain resistance at different temperatures for different NMOS and PMOS extension implant doses. 106

5.6 Breakdown of expected improvements and their sources for 0.5 μm CMOS at 78 K and a power supply of 3 v. 112

Acknowledgements

Much of this work would not have been possible without the help of numerous individuals. First and foremost, I want to thank Professor Jason Woo for his advice and assistance over these past six years. I think have become a better engineer and researcher through his leadership and guidance. I wish to also thank Professors Dee-Son Pan, Ian McLean, and Ming Wu for serving on my Ph.D. committee and finding the time to critique this work.

A great deal of my appreciation goes out to the “first generation” of students who went through Jason’s lab. They are Dr’s. Chi-Hung Lin, Tommy ”Siskel and Ebert” Hsiao, Neal ”World Traveler” Kistler, Vincent Chen, Janet Wang, Miryeong Song, and Deepak Nayak. Everyone made me feel welcome and wanted, and I’m glad that we’ve been able to keep in touch even long after serving your time here at UCLA. I also wish to acknowledge those who started in this group more recently: Tsung-Chia Kuo, Steve ”Citibank” Chin, Jason Gillick, Peng-Fei Zhang, Zhimin Zhou, Baohong Cheng, Ping Liu, Tony Tseng, Debbie Fixel, Danny Leung, Brad Ikegami, Guangzhao Pan, Jin-Ho Seo, Ivan To, Dawei Guo, Ciby Thuruthiyll, Vincent ”Lunchtime!” Gau, and Anand Inani. Special thanks go out to Brad, Debbie, Tony, Baohong, Steve, Anand, and Ivan (as well as Jeannie Alt and Dorothy Tarkington) for helping me take care of administrative details while I was up at Stanford

doing my processing. I also appreciate the interaction I had with other UCLA graduate students. They include: Martin Tanner, Shawn Thomas, Alvin Ching, Jeff Oh, Venkatraman Prabhaka, and Eric Bensen. It was nice to bear one another's burdens during our long stay here at UCLA.

For two and a half years, I served my grad school time fabricating several CMOS wafer lots up at Stanford University. I appreciate the help, assistance, and comradery from all of the staff at the Center for Integrated Systems. I want to thank John Shott for giving me the opportunity to use the facilities and Mary Donogohue for her administrative assistance. Special thanks go to the process engineers (Nancy Latta, Margaret Prisbe, Gladys Sarmiento, Marnel King, Robin King, Mary Martinez, Bob Wheeler, and Paul Jerabek) and process technicians (Len Booth, Karl Brandt, Bill Martin, Keith Gaul, and Pat Burke) who kept the lab running and were always easy to talk to, a key necessity when you're stuck in a bunny suit for 8 to 12 hours each day. Graduate students and staff who were especially helpful while I was up at Stanford include Dr. Mark McCord, Steve Kuehne, Dr. Anthony McCarthy, Vivek Subramanian, and Alvin Loke.

Finally, I want to thank Bob Hammond and Superconductor Technologies Incorporated for their financial support.

VITA

September 26, 1970	Born, Pasadena, California
1990–1992	Teaching Assistant University of California, San Diego La Jolla, California
1992	B.S., Electrical Engineering (Summa Cum Laude) University of California, San Diego La Jolla, California
1994	M.S., Electrical Engineering University of California, Los Angeles Los Angeles, California
1993–1995	Teaching Fellow University of California, Los Angeles Los Angeles, California
1992–1999	Research Assistant University of California, Los Angeles Los Angeles, California

PUBLICATIONS AND PRESENTATIONS

C. Hwang, B. Hammond, J. Gillick, C. Jenq, and J. Woo, "Process Modification for Improved Low Temperature CMOS Performance," *2nd European Workshop on Low Temperature Electronics (WOLTE II)*, p. 193-197, 1996.

C. Hwang, J. Gillick, B. Hammond, C. Jenq, and J. Woo, "Bias Dependent Hot-Carrier Reliability and Lifetime Over a Wide Temperature Range," *2nd European*

Workshop on Low Temperature Electronics (WOLTE II), p. 25-28, 1996.

C. Hwang, T. Kuo, and J. Woo, "Extraction of Gate Dependent Source/Drain Resistance and Effective Channel Length in MOS Devices at 300 and 77 K," *IEEE Trans. Electron Devices*, vol. 42, p. 1863-1865, 1995.

I. Groves, G. Brown, G. Pollack, K. Green, L. Dawson, A. D'Souza, C. Lin, M. Song, C. Hwang, J. Woo, and K. MacWilliams, "One-micrometer Radiation-Hardened Complementary Metal Oxide Semiconductor for Cryogenic Analog Applications," *Infrared Readout Electronics: Proceedings of the SPIE*, vol. 2226, p. 72-84, 1994.

C. Hwang and C. Tu, "User-Friendly Software for Molecular Beam Epitaxy," UCSD Undergraduate Research Conference, 1992.

ABSTRACT OF THE DISSERTATION

Optimization of CMOS for High Performance,

Low Temperature Operation

by

Clifford Y.C. Hwang

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 1999

Professor Jason C.S. Woo, Chair

Interest in the operation of CMOS at low temperature has grown steadily in recent years, especially in light of the increasing challenges in device design and manufacturing at deep-submicron dimensions. Numerous improvements in device and circuit performance are possible at lower temperatures. However, several obstacles stand in the way of implementing low temperature CMOS for mainstream applications. In this dissertation, design methodologies for maximizing the performance of low temperature CMOS will be presented. First, the limits of conventional, room temperature CMOS at low temperatures will be identified. Both carrier freeze-out and threshold voltage shifts significantly reduce the expected improvement from 300 K to 78 K.

After identifying the problems with conventional CMOS at low temperatures, TCAD simulation tools will be used to redesign the devices for better performance. A step by step procedure will be presented to fit measured I-V characteristics to MEDICI, a device simulator. Proper calibration can be easily achieved by changing

some of the modeling parameters to reflect the change in operating temperature. Less than 5% error between simulation and measurement results is obtained for a 1 μm CMOS process.

With properly calibrated simulation programs, new device designs for low temperature operation can be tested and evaluated. Simply readjusting the threshold voltage will lead to significant improvement on the device and circuit level without an increase in off-state leakage. Freezeout effects can also be eliminated by modifying a few process steps used to make conventional CMOS. Additional re-engineering of the channel and gate will further improve performance. In order to demonstrate the effectiveness of these changes, transistors and circuits have been fabricated. When a proper threshold voltage is used, the current drive doubles at 78 K. By retrograding the profile in the channel region, additional improvement in mobility at low temperature can be achieved. Carrier freezeout can be minimized by increasing the dose of the LDD implant and switching to p^+-poly PMOS. A 2x to 3x drop in propagation delay is demonstrated when all of these changes are implemented. With such improvements even at deep sub-micron dimensions, temperature should be considered as an additional factor in scaling CMOS for high performance applications.

Chapter 1

Introduction

1.1 Background

The electronics industry today continues to advance at an incredible rate. Almost every month, newer and faster computer chips are introduced to the consumer. To obtain higher performance, smaller MOS devices are needed. However, as minimum device dimensions approach the deep submicrometer regime, greater challenges have been encountered. Manufacturing limits are being approached in linewidth resolution, gate oxide thickness, and junction depth. In addition, the cost of developing each successive generation of CMOS continues to increase. Instead of scaling down the channel length, the operating temperature can be reduced to improve the performance of an already existing technology.

There has been ongoing interest in low temperature CMOS since the mid-1970's. By lowering the operating temperature of a device, numerous performance improvements are possible. Carrier mobility, current drive, metal conductivity, and latch-up suppression increase, while subthreshold slope and parasitic capacitances decrease [1]-[2]. A 3x improvement in digital circuit performance at 77 K has been shown in the past [3]. Since thermal noise is lower at low temperatures, improved analog RF performance is expected. In fact, recent advances in the performance and reliability in cryogenic cooling systems have opened up the possibility of using low temperature electronics in cellular base stations [4]. Having the freedom to reduce the operating temperature allows for greater flexibility in designing high performance CMOS for analog or digital applications.

1.2 Motivation

Although low temperature CMOS shows better performance versus room temperature operation, several key issues remain unresolved. The amount of improvement at short channel lengths can be minimal. In addition, carrier freezeout and the increase in threshold voltage at low temperatures limit the increase in current drive. The necessity of a cooling unit prevents the use of low temperature CMOS in portable, low-power applications. Furthermore, device reliability is known to worsen as the temperature decreases [5]-[6]. Design methodologies for obtaining high performance, low temperature CMOS have been proposed [7], but actually implementing the strategies is unclear.

Clearly, a better understanding of the advantages and limits of low temperature CMOS is needed before temperature can be accepted as a means of enhancing device performance. The objectives of this work are summarized below:

1. To quantify CMOS device and circuit improvement versus temperature down to deep-submicron dimensions. A wide range of temperatures (343 K to 78 K) will be used. Calculation of the maximum improvement achievable at low temperature is highly desirable, especially for the very small dimensions which will be used to design next-generation computer circuits.

2. To understand what limits low temperature performance and propose ways of minimizing their effect. The factors that limit device performance at room temperature are not necessarily similar at low temperatures.

3. To determine what changes need to made to an existing, room temperature CMOS technology so that maximum low temperature performance is achieved. Keeping a similar process flow reduces the introduction of significant processing or design problems.

1.3 Organization

This dissertation is organized into the following chapters. The performance and reliability of conventional, room temperature CMOS at low temperature is shown in Chapter 2. Although there is better device performance, improvements on the circuit level are very poor. Due to enhanced hot carrier degradation, lower operating voltages are required at lower temperatures, further reducing the performance improvement possible.

After noting the limitations of room temperature CMOS at low temperature, Chapter 3 presents a step-by-step procedure for calibrating TCAD device simulation tools to measurement results at low temperatures. This curve fitting is necessary so that accurate redesign of CMOS for high performance, low temperature operation is possible. By adjusting the temperature dependent modeling parameters, a very good fit in the I-V characteristics across channel length and temperature can be obtained.

If low temperature CMOS is to be implemented for mainstream applications, the most logical and economical approach will be to modify an existing, room temperature CMOS technology. The steps in the process flow that need to be modified for maximum performance at low temperature are addressed in Chapter 4. Although only a few steps are modified, simulations show significant performance improvement at low temperatures.

Chapter 5 shows results for devices and circuits designed for low temperature operation. Using retrograded profiling in the channel, a significant improvement in performance is possible at low temperatures due to the further enhancement in mobility. Along with proper threshold voltage design, a reduction of 2x in delay is easily achieved at sub-micron dimensions.

Finally, in Chapter 6, the results of this dissertation are summarized and suggestions for future work are presented.

Chapter 2

Performance and Reliability of Conventional CMOS at Low Temperatures

2.1 Introduction

Lowering the temperature of an MOS device results in better current drive and subthreshold slope. To illustrate these enhancements, measurements made on devices designed for room temperature operation will be presented in this chapter. Although improvement can be seen, the actual amount is minimal, especially at the circuit level. The factors that lead to this poorer than expected performance will be explained in detail.

Along with the performance data, reliability results will be shown for the same devices at temperatures of 78 K, 150 K, and 218 K. Due to enhanced hot carrier

degradation, a lower supply voltage will be required at lower temperatures, further limiting the improvement.

2.2 Low Temperature Performance

2.2.1 Device and Setup Details

NMOS and PMOS transistors fabricated from a 1 μm CMOS process were used in this study. The gate oxide thickness was 150Å, and the devices were designed for room temperature operation. Using a computer controlled HP 4142B connected to a temperature controlled Lakeshore Cryotronics dewar, I-V characteristics were measured at temperatures of 343 K (70^oC), 300 K, 218 K (the freezing point of CO_2), 150 K, and 78 K (the boiling point of N_2).

2.2.2 Measurement Results

Figure 2.1 shows the linear I_{ds}-V_{gs} characteristics of a 25 μm x 25 μm NMOS transistor at different temperatures. Higher drain current and more rapid turn-off characteristics are evident as the temperature is reduced. The improvement in subthreshold slope allows for a lower threshold voltage without an increase in off-state current.

Short channel, saturation current characteristics for NMOS and PMOS devices are shown in Figure 2.2. Although current drive increases as the temperature is reduced, the amount of improvement reduces for each successive drop. This is especially true for the PMOS devices. The improvement in saturation current across

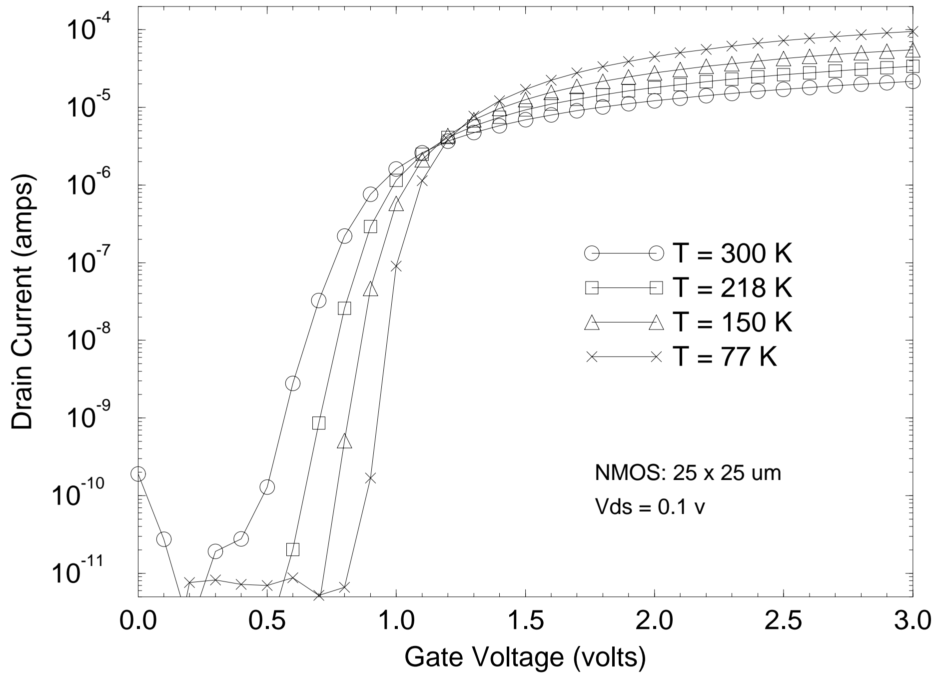

Figure 2.1: I_{ds}-V_{gs} characteristics of a 25 μm x 25 μm NMOS device at 300 K, 218 K, 150 K, and 78 K.

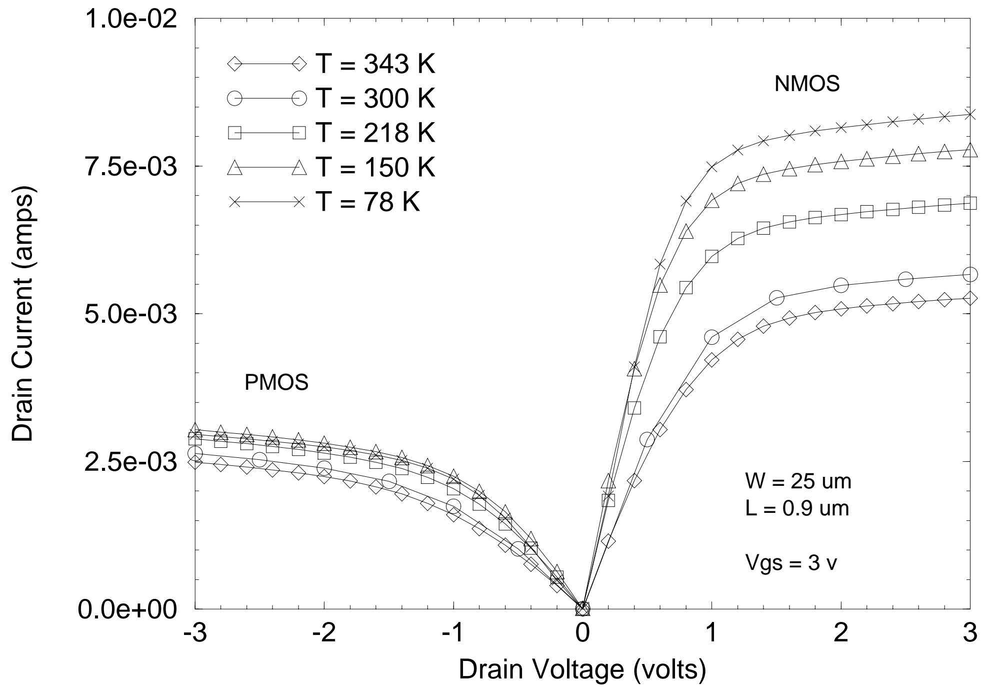

Figure 2.2: I_{ds}-V_{ds} characteristics of a 25 μm x 0.9 μm NMOS device at 343 K, 300 K, 218 K, 150 K, and 78 K.

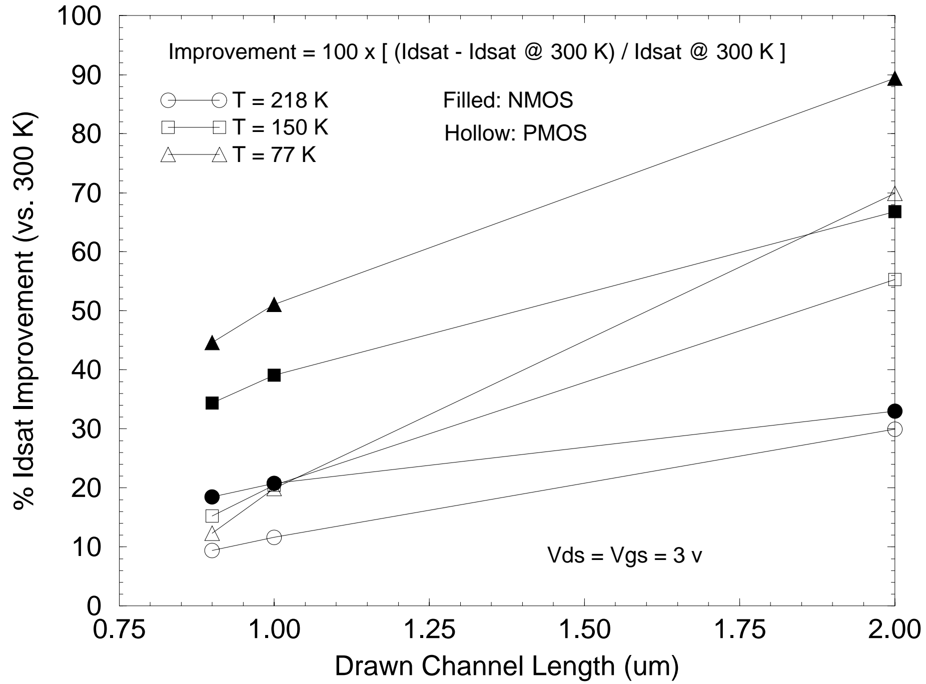

Figure 2.3: Improvement in drain saturation current versus drawn channel length for different temperatures. The data has been normalized to measurements at room temperature.

channel lengths is illustrated in Figure 2.3. Percentage improvement in saturation current is defined as:

$$Improvement = 100\,(\frac{I_{dsat}\ at\ \ ow\ temperature}{I_{dsat}\ at\ 300K} - 1) \tag{2.1}$$

Thus a 100% improvement corresponds to a 2x increase in current drive. At 78 K, the short channel PMOS saturation current improves by only 12% compared to room temperature results. In fact, the improvement at 150 K is higher (16%). Since most circuitry uses short channel lengths for high current drive and low input (gate) capacitance, low temperature operation appears to be an unattractive means of improving performance.

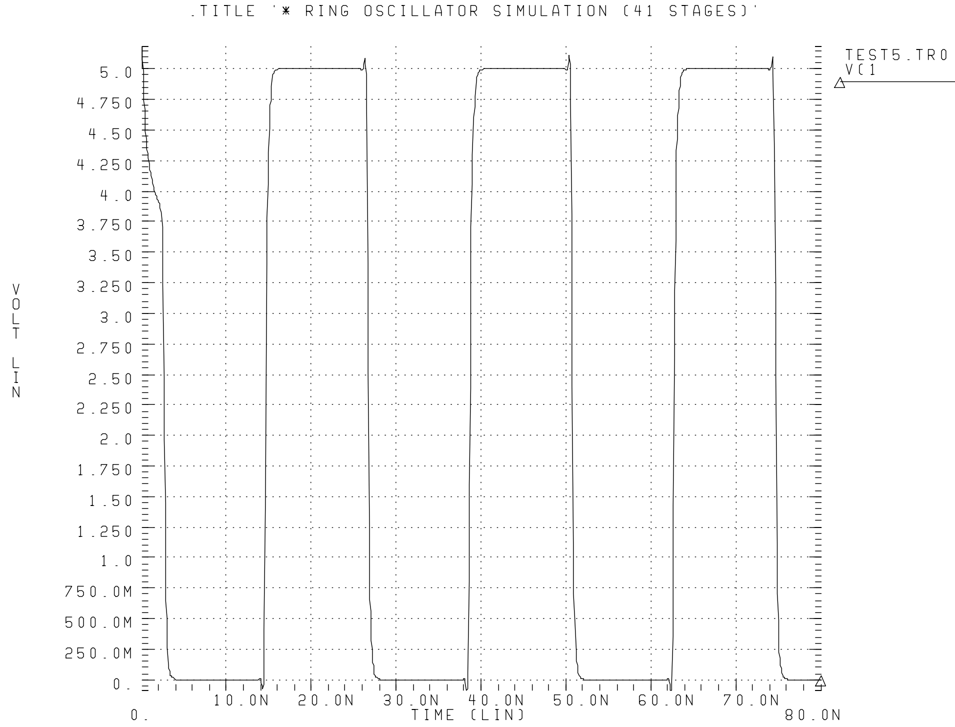

Figure 2.5: Simulated waveform for a 41 stage, 1 μm ring oscillator at 300 K. The power supply voltage is 5 v.

waveform and dividing by twice the number of stages. Propagation delays at 300 K and 78 K are shown in Figure 2.6. Defining percentage delay improvement as

$$Improvement = 100\,\left(\frac{de\ ay\ at\ 300K}{de\ ay\ at\ \ ow\ temperature} - 1\right) \qquad (2.2)$$

the improvement in propagation delay can be studied at different operating temperatures (Figure 2.7). A 100% improvement corresponds to a 2x reduction in delay. Looking at Figure 2.7, only a 1.5x improvement in delay is possible at best. This amount is very optimistic because in order to avoid enhanced hot-carrier degradation and to ease cooling requirements, a low power supply voltage is necessary at low temperatures.

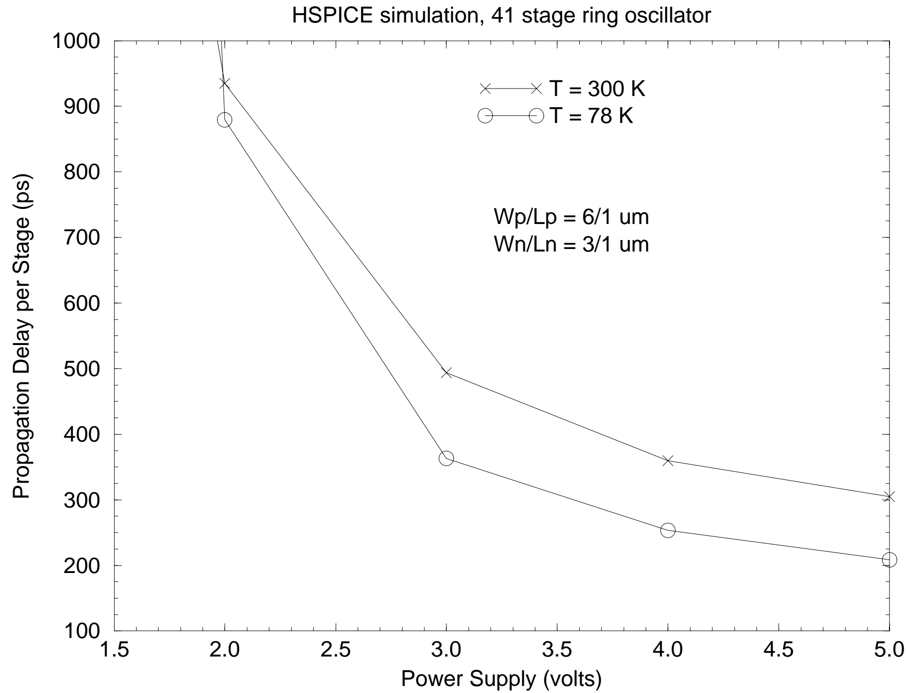

Figure 2.6: Simulated propagation delay per stage versus power supply of a 41 stage, 1 μm ring oscillator at 300 K and 78 K.

2.3 Performance Limitations at Low Temperature

The factors that hinder performance with reduced operating temperature will be discussed in this section. By minimizing these effects, high performance, low temperature CMOS devices and circuits can be realized.

2.3.1 Threshold Voltage Shift

Figure 2.8 shows the threshold voltage as a function of temperature. The increase in threshold voltage at lower temperatures results from the shift in the Fermi level. Assuming a non-degenerate p-type semiconductor, the Fermi energy is given by

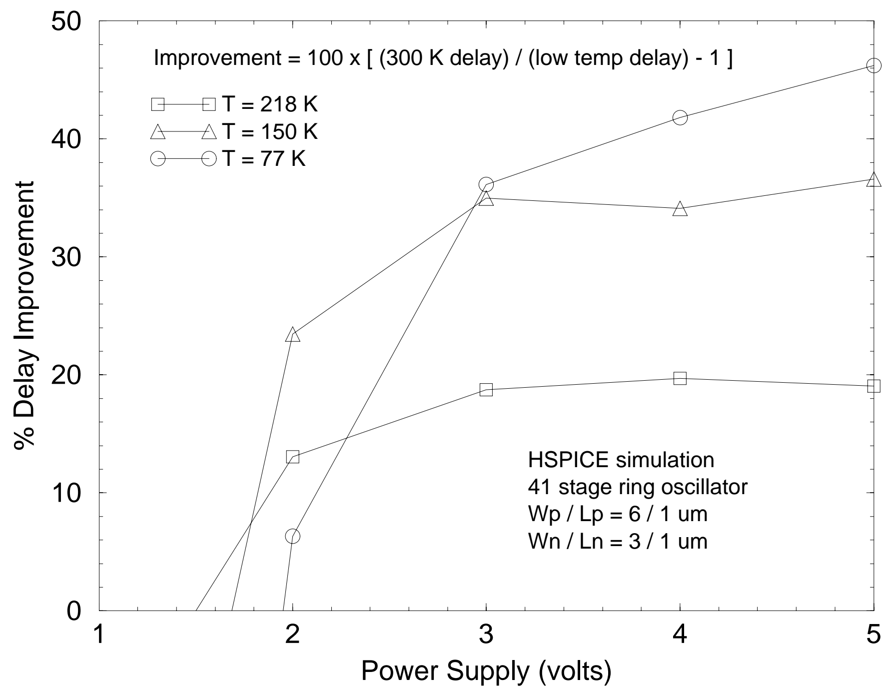

Figure 2.7: Simulated propagation delay improvement per stage versus power supply of a 1 μm ring oscillator at different temperatures. The results are normalized to 300 K.

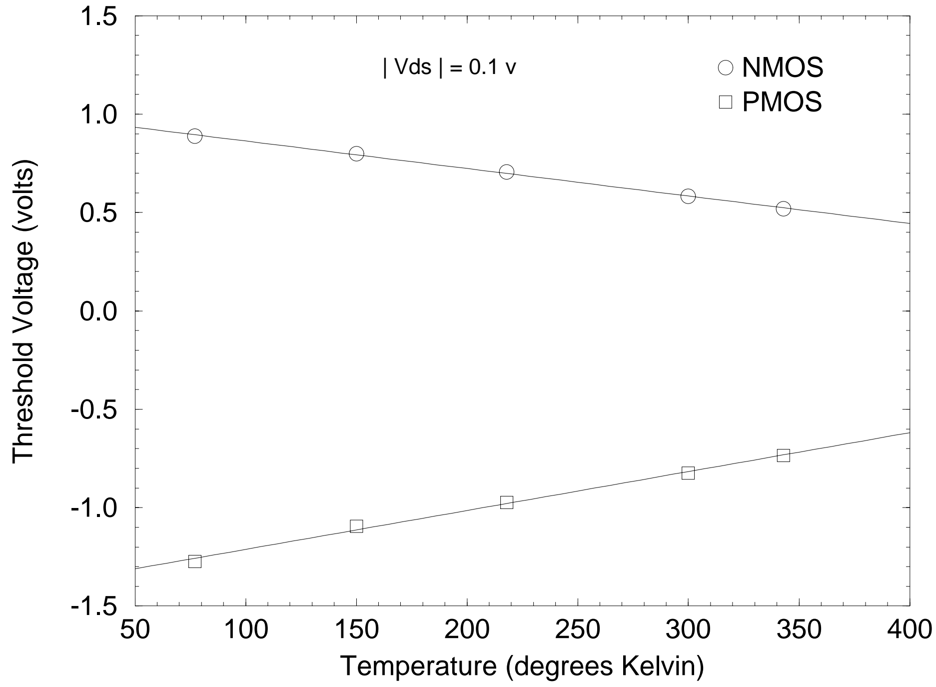

Figure 2.8: Threshold voltage versus temperature for a 25 μm x 1 μm device. A constant 250 nA drain current is used to define V_{th}.

$$E_f = \frac{kT}{q} \ln \frac{N_v}{N_A^-} + E_v \tag{2.3}$$

where N_v is the density of states in the valence band, N_A^- is the ionized impurity density, and E_v is the valence band energy. At lower temperatures, the first term of (2.3) is lower. Therefore, the Fermi energy shifts closer to the valence band and away from the midgap, or intrinsic energy, E_i. Since the threshold voltage requires that the bands bend by $2\phi_b$ where

$$\phi_b = |E_i - E_f| \tag{2.4}$$

more band bending is necessary at lower temperature. Therefore, the applied gate bias, or threshold voltage in this case, must increase. A similar argument holds if the semiconductor is n-type.

The increase in threshold voltage at low temperatures limits the improvement in current drive under the same bias conditions. This shift is especially undesirable because of the reduction in subthreshold slope. By taking the V_{th} increase into account in device design, better performance can be achieved.

2.3.2 PMOS Carrier Freezeout

Looking at Figure 2.8, the threshold voltage for the PMOS devices increases at a faster rate than the NMOS devices. This can be attributed to freezeout of the boron dopants in the channel. Figure 2.9 shows the band diagram for an n^+-poly PMOS device with a gate bias near the threshold voltage. The amount of dopants that ionize is given by

$$N_A^- = \frac{N_A}{1 + e^{\frac{q(E_a - E_f)}{kT}}} \tag{2.5}$$

where N_A is the concentration of dopants in the channel, and E_a is the acceptor level for boron. At lower temperatures, the exponential term becomes larger, which reduces the effective concentration of the channel (N_A^-). As a result, the threshold voltage increases at a greater rate as the temperature is lowered. This phenomena does not occur for a boron doped NMOS device because the bands bend in the opposite direction (Figure 2.10), and the difference between E_f and E_a is large.

Freezeout in the PMOS channel also degrades the subthreshold slope (Figure 2.11). This is because the exponential term in (2.5) becomes more sensitive

Figure 2.9: Energy band diagram for an n^+-poly PMOS device biased near the threshold voltage. Boron is implanted into the channel in order to have a proper threshold voltage.

Figure 2.10: Energy band diagram for an n^{+}-poly NMOS device biased near the threshold voltage.

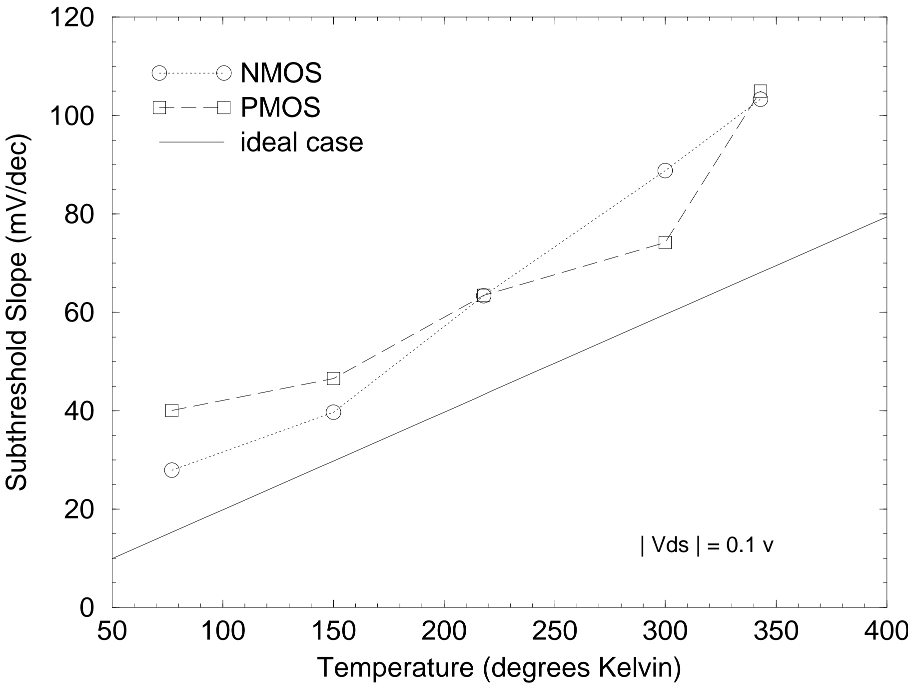

Figure 2.11: Subthreshold slope versus temperature for a 25 μm x 1 μm device.

Figure 2.12: Energy band diagram in the LDD region of an NMOS device.

to changes in $E_f - E_a$ as the temperature decreases. The ionized concentration begins to vary, depending on applied gate bias or amount of band bending (i.e. $E_f - E_a$). And since subthreshold slope is a function of the doping concentration, the subthreshold slope degrades.

2.3.3 LDD Freezeout

Freezeout not only worsens the PMOS characteristics near threshold, but also degrades both NMOS and PMOS current drive. Figure 2.12 shows the band diagram along the lightly doped source/drain (LDD) region of an NMOS device. The concentration of dopants that will ionize and give up free electrons is given by

$$N_D^+ = \frac{N_D}{1 + e^{\frac{E_f - E_d}{kT}}} \tag{2.6}$$

where N_D is the concentration of n-type dopants in the LDD region, and E_d is the donor level. At room temperature, the Fermi level is typically below the donor level so that full-ionization occurs. As the temperature decreases, E_f moves toward the conduction band, E_c. When the Fermi level crosses over E_d, the exponential term becomes significant and freezeout occurs. A similar argument holds in the PMOS LDD region.

For high doping concentrations, freezeout does not occur due to bandgap narrowing. The conduction band will drop for high levels of dopants. When E_c crosses over E_d, full-ionization occurs independent of E_f or the operating temperature. This typically occurs around concentrations of $4x10^{18}cm^{-3}$ (Mott's transition). Unfortunately, typical LDD concentrations are below this level, resulting in carrier freezeout. The active, electrical carrier concentration can be reduced by as much as 50% due to freezeout. Because the doping concentration in the LDD regions of both NMOS and PMOS devices is significantly lower at low temperature, the source/drain resistance increases and limits current drive. This is confirmed by looking at the measured linear drain current (Figure 2.13) and the parasitic source/drain resistance used in HSPICE circuit simulations (Figure 2.14). It is clear from both figures that freezeout does not play a role until the temperature drops below 150 K. For long channel lengths, the device resistance is dominated by the resistance of the channel, and improvement is more significant (see Figure 2.3).

2.4 Low Temperature Reliability

Performance is not the only critical factor in determining the value of a particular technology. Devices and circuits must also demonstrate good reliability. In this

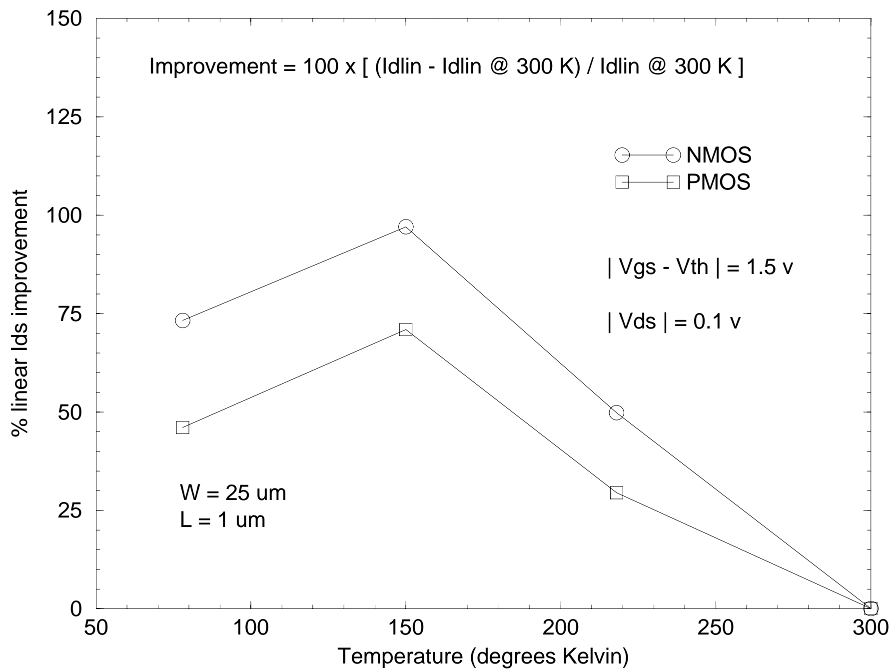

Figure 2.13: Linear drain current versus temperature for a 25 μm x 1 μm device. Freezeout reduces the expected linear drain current at 78 K by roughly 100%.

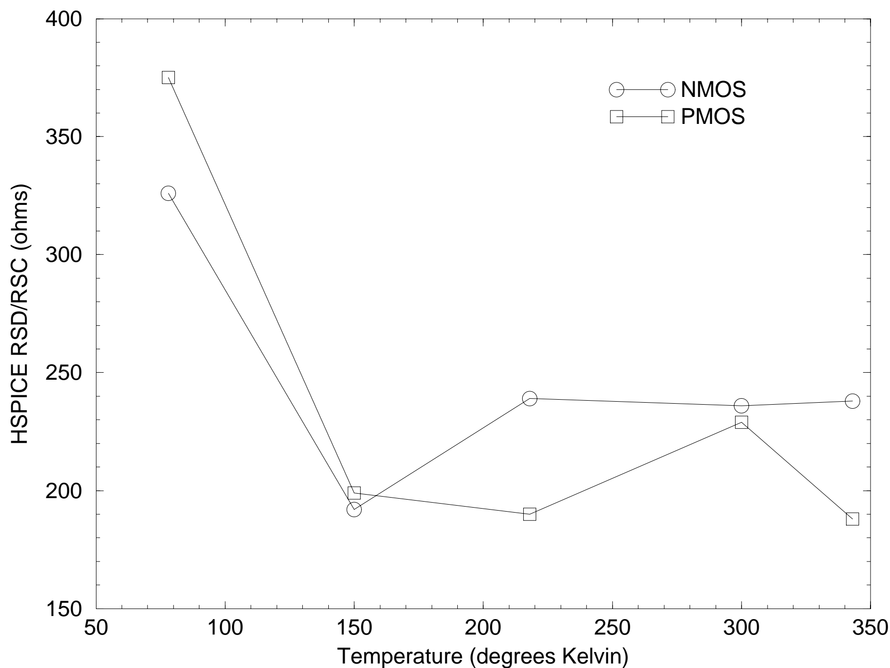

Figure 2.14: Parasitic source/drain resistance used in HSPICE ring oscillator simulations for different temperatures. Due to freezeout, the 78 K resistance is much higher than at any other temperature.

section, lifetime data for the aforementioned devices will be summarized. Additional details can be found in [9]- [10].

2.4.1 Stress Conditions

NMOS transistors were DC stressed at drain biases ranging from 3 v to 6.5 v for 4000 seconds. The gate bias used during the stress was either $V_{gs} = V_{ds}$ (maximum electron injection) or $V_{gs} = \frac{1}{2}V_{ds}$ (I_{submax}: maximum interface state generation). These stress conditions have been shown to cause maximum degradation at 78 K [6] and 300 K [11], respectively.

To improve reliability at room temperature, an LDD implant of phosphorous was used with a dose of $5x10^{13}cm^2$ at an energy of 75 keV. As shown in Section 2.3.3, the resulting doping profile freezes out at 78 K, resulting in degraded current drive.

2.4.2 Lifetime Results

Figure 2.15 shows the 10 year lifetime data for the two stress conditions. Using linear extrapolation, the time it takes for the linear transconductance (g_m) to degrade by 10% can be determined. This time is defined as the 10-year lifetime. As expected, for any V_{ds}, the $V_{gs} = V_{ds}$ stress condition results in the shortest lifetime. In order to have a 10-year lifetime at 78 K, the power supply must be reduced by half from the original 5 v design down to 2.7 v. Unfortunately, by operating circuits at 2.7 v, only a 25% improvement in performance is possible at 78 K (see Figure 2.7).

Figure 2.15: Lifetime versus $1/V_{ds}$ for 25 μm x 1 μm NMOS devices at 78 K. The two-slope behavior in the $V_{gs} = V_{ds}$ stress condition may be due to different damage locations or mechanisms for high versus moderate drain biases. In any case, the results indicate that stressing should be done near the power supply voltage of interest. Otherwise, the extrapolated 10-year lifetime can be too optimistic.

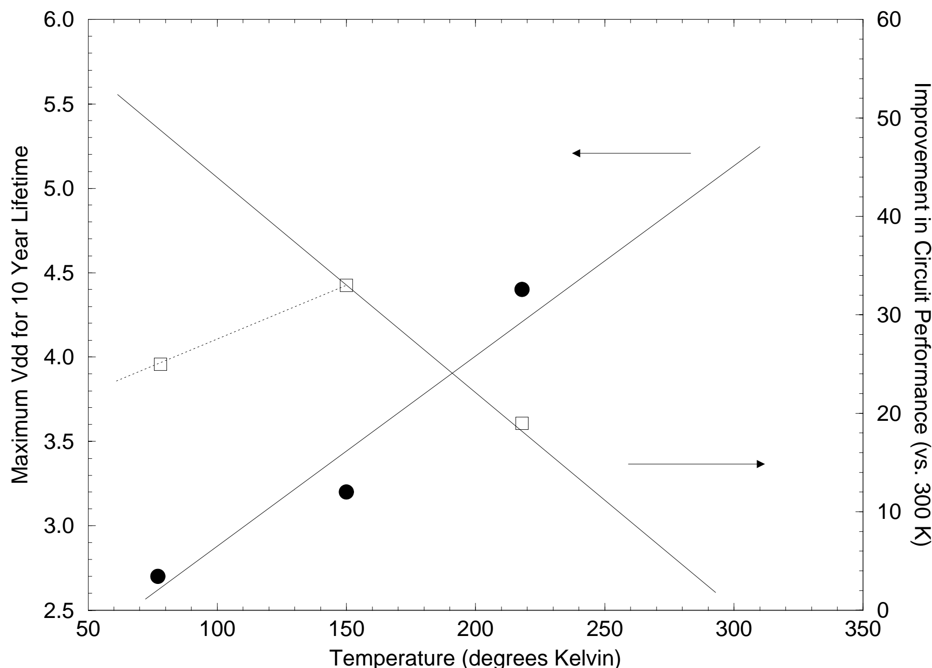

Figure 2.16: Maximum V_{dd} at different temperatures so that a 10 year device lifetime is maintained (solid circles) and corresponding circuit-level performance improvement (hollow squares). A linear trend is evident in both cases, although freezeout limits the performance improvement at 78 K.

At intermediate temperatures, the power supply voltage corresponding to a 10-year lifetime is higher (Figure 2.16), but the maximum amount of performance improvement possible (from Figure 2.7) remains below 35%. The limit in performance improvement at 78 K is primarily due to carrier freezeout in the LDD regions.

2.5 Conclusion

The performance limitations of conventional, room temperature CMOS at low temperature have been studied at length in this chapter. Although the process technology is old (1 μm CMOS) by today's standards, the challenges of low temperature CMOS are still evident. The increase in threshold voltage limits the improvement

in current drive that could've been possible had there been no shift at all. The remaining improvement in current is further reduced because power supply voltages need to be lowered to maintain a 10-year lifetime.

Carrier freezeout places additional limits on CMOS performance improvement. The effectiveness of the PMOS threshold implant is reduced, leading to poorer subthreshold behavior. As channel lengths are scaled into the submicron regime, the parasitic source/drain resistance plays a greater role in the overall device resistance. Due to freezeout in the LDD regions at 78 K, there is poor improvement in device and circuit performance from 150 K to 78 K. With a maximum improvement of only 35%, high performance, low temperature CMOS cannot be realized by simply cooling a conventional, room temperature technology. To warrant using a low operating temperature, the devices must be redesigned in order to have a more reasonable threshold voltage and minimized freezeout effects.

Chapter 3

Calibration of TCAD Device Simulation Tools for Low Temperature CMOS

3.1 Introduction

In order to gain a better understanding of low temperature device physics and ways of improving its performance, TCAD simulation tools can be used to predict device behavior. However, in order to obtain accurate results, the tools should be initially calibrated to actual measurement results. This chapter will present a series of steps used to fit data from MEDICI [12], a TCAD device simulator, to measured I-V characteristics. The key temperature dependent parameters that need to be modified will be summarized. After adjusting these values to better fit the data, the advantages and limitations of low temperature CMOS will become more evident.

3.2 Required Pre-Calibration Data

3.2.1 Device Measurements

Before running any simulations, I-V measurements at the desired operating temperature need to be made. To obtain a good fit across the entire dynamic range of the device, the following I-V characteristics are needed:

- Linear characteristics: I_{ds} vs. V_{gs} with low V_{ds}.
- Saturation characteristics: I_{ds} vs. V_{ds} with variable V_{gs}.
- Substrate bias characteristics: I_{ds} vs. V_{gs} with variable V_{sb}. These curves are important because most circuits contain transistors stacked in series.

The actual voltage ranges used will depend on the particular technology generation of the devices. Here, measurements were made up to 5 v at temperatures of 300 K, 218 K, 150 K, and 78 K.

3.2.2 Doping Profiles

Along with the I-V characteristics, the doping profiles for the entire transistor are needed. These can be obtained by running TCAD process simulations, which take as input the steps which are used to manufacture the device. If the process flow is unavailable, a "best-guess" profile can be constructed using for example, Gaussian distributions. In this case, the process flow was entered into SUPREM-III [13], a one-dimensional process simulator. Separate input decks were created for the gate, LDD, and source/drain regions of the device. The final doping profiles for both

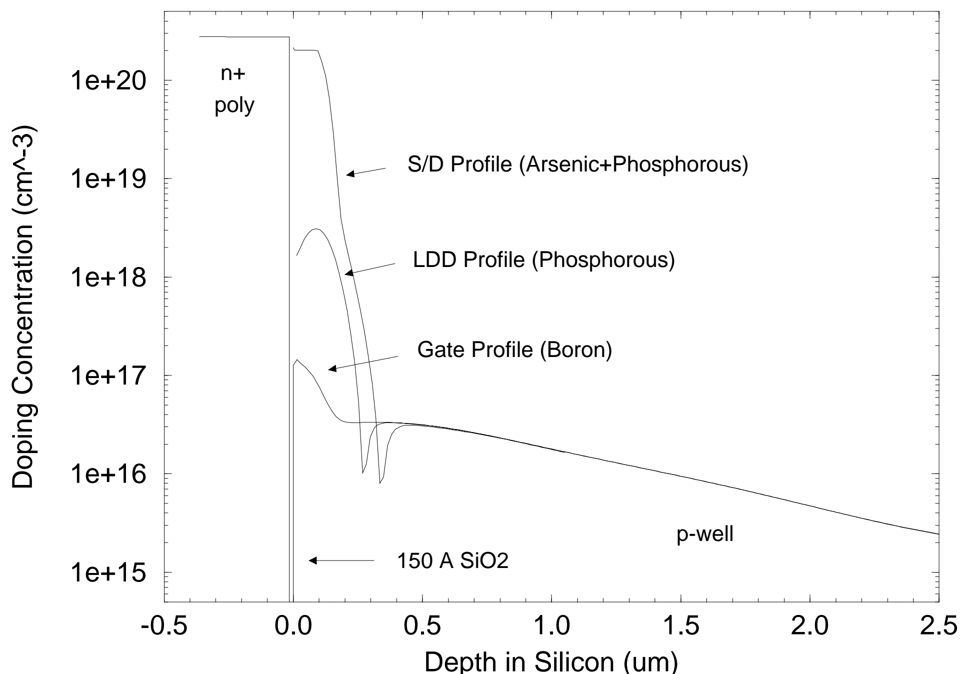

Figure 3.1: SUPREM-III doping profiles underneath the gate, LDD, and source/drain regions of an NMOS device.

NMOS and PMOS devices are shown in Figures 3.1 and 3.2. These profiles serve as input for the MEDICI simulations.

To transfer the vertical, one-dimensional NMOS profiles to the two-dimensional grid used for MEDICI simulations, the net active doping in the gate region is first applied uniformly over the entire grid. Then, the active concentration from the LDD implant (phosphorous) is inserted into the source/drain and LDD regions. Finally, the active concentration from the source/drain implant (arsenic) is added to the source/drain region of the structure. A similar approach is used to form the PMOS structure. The resulting doping profiles from the grid agree well with Figures 3.1 and 3.2.

To account for two-dimensional effects and the lateral spreading of dopants from one region into another, a lateral diffusion constant is used which multiplies the

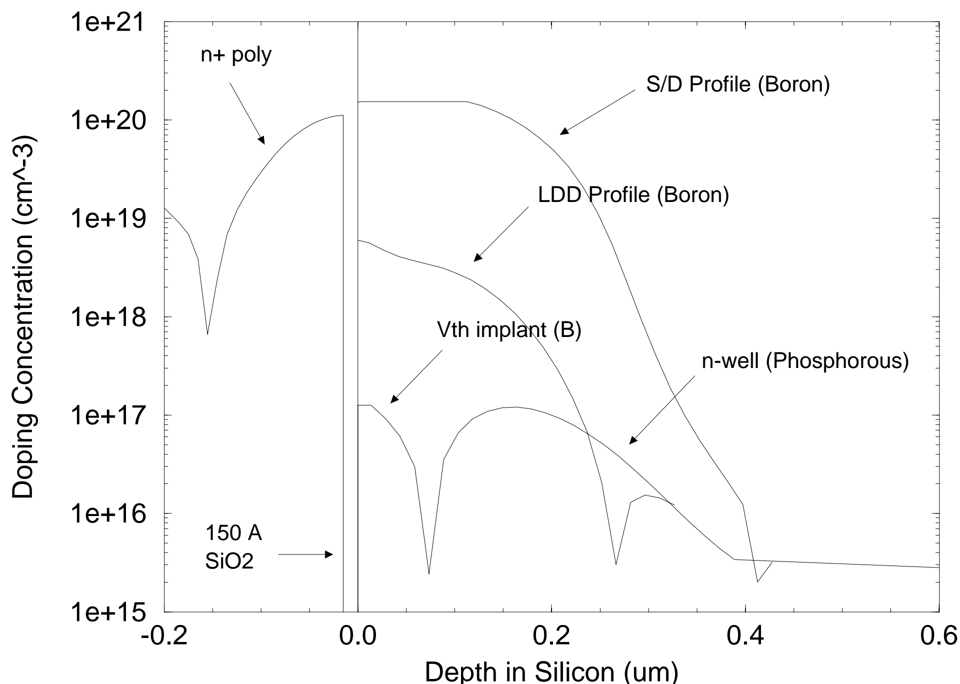

Figure 3.2: SUPREM-III doping profiles underneath the gate, LDD, and source/drain regions of a PMOS device.

extent of the original profile as it is rotated horizontally. A value of 0.8 is assumed, which is typical for standard CMOS processing. More accurate two-dimensional profiles can be obtained by running a 2-D process simulator, such as TSUPREM-4 [14]. By using this tool, the transfer of the doping profile to MEDICI is more direct. However, simulation time and complexity increase with the use of this tool.

3.3 NMOS Fit

After obtaining the measured data and doping profiles, calibration of the device simulator can begin. The I-V curves are fitted one set at a time across all channel lengths, starting with the long channel devices. By properly adjusting the temperature dependent terms, a good initial fit can be obtained. Further refinement can

be done by adjusting process related parameters. Care must be taken not to alter any of the values beyond reason.

3.3.1 I_{ds} vs. V_{gs} At Low V_{ds}

The first set of curves to fit are the drain current versus gate voltage curves at low drain bias. Initially, the following temperature dependent parameters are modified in the input deck:

- The temperature is set to the desired value in the MODELS card.
- Full Fermi-Dirac statistics are set in the MODELS card for all simulations below 300 K. This is due to the fact that the Boltzmann approximation, which requires that $E_c - E_f \gg kT$ and $E_f - Ev \gg kT$, fails at low temperature because the Fermi energy moves closer to the band edges.

In addition, the parameters that model mobility degradation due to the vertical electric field are adjusted. Here, an enhanced, temperature dependent surface mobility model [15] is used (SRFMOB2). This model consists of three terms.

$$\begin{aligned} \mu_{eff,n} = & \frac{1}{MUN1.SM}(\frac{E_{eff\perp,n}}{10^6})^{EXN1.SM} + \frac{1}{MUN2.SM}(\frac{E_{eff\perp,n}}{10^6})^{EXN2.SM} \\ & + \frac{1}{MUN3.SM}(\frac{N_B}{10^{18}})(\frac{10^{12}}{N_{inv}})^{EXN3.SM} \end{aligned} \tag{3.1}$$

The parameters which can be modified for MEDICI simulations (via the MOBILITY card) are indicated using all capital letters. The three terms above account for mobility degradation due to phonon scattering, surface roughness scattering, and ionized impurity scattering, respectively. The dominant mechanism depends on the operating temperature. The vertical electric field, $E_{eff\perp,n}$, is equal to

$$E_{eff\perp,n} = \frac{(Q_{depl} + ETAN\, Q_{inv})}{\epsilon_{si}} \tag{3.2}$$

At 78 K, parameters from [16] are chosen. For intermediate temperatures, a linear extrapolation between the 78 K values and the default 300 K values is used. A better approach would be to use the equations extracted from mobility measurements at intermediate temperatures in [18].

These initial modifications should result in a reasonable fit on the first try. Starting with the long channel devices, adjustments can be made to the fixed charge density (QF: INTERFACE card) to better fit the threshold voltage. At low temperatures, the bulk mobility (MUN0, MUP0: MOBILITY card) should also be increased, within reason, so that a good fit is obtained in the I_{ds}-V_{gs} characteristics at low to moderate gate biases. An initial value can be obtained by extracting and fitting the bulk mobility from a long channel device. Rearranging the long channel, linear I_{ds}-V_{gs} equation, yields

$$\mu_{eff} = \frac{I_{ds}\, L_{eff}}{W_{eff}\, C_{ox}\, (V_{gs} - V_{th} - \frac{V_{ds}}{2})V_{ds}} \tag{3.3}$$

where L_{eff} is the effective channel length, W_{eff} is the effective channel width, C_{ox} is the oxide capacitance, and V_{th} is the threshold voltage. The resulting curve can then be fitted to a simple model of the mobility:

$$\mu_{eff} = \frac{MUN0}{1 + F1(V_{gs} - V_{bi} - F3\, V_{ds})} \tag{3.4}$$

where V_{bi} is the built-in voltage, and F1 and F3 are fitting factors.

At short channel lengths, the simulated current will tend to overestimate the measurement results at high gate biases. To obtain a better fit, parasitic

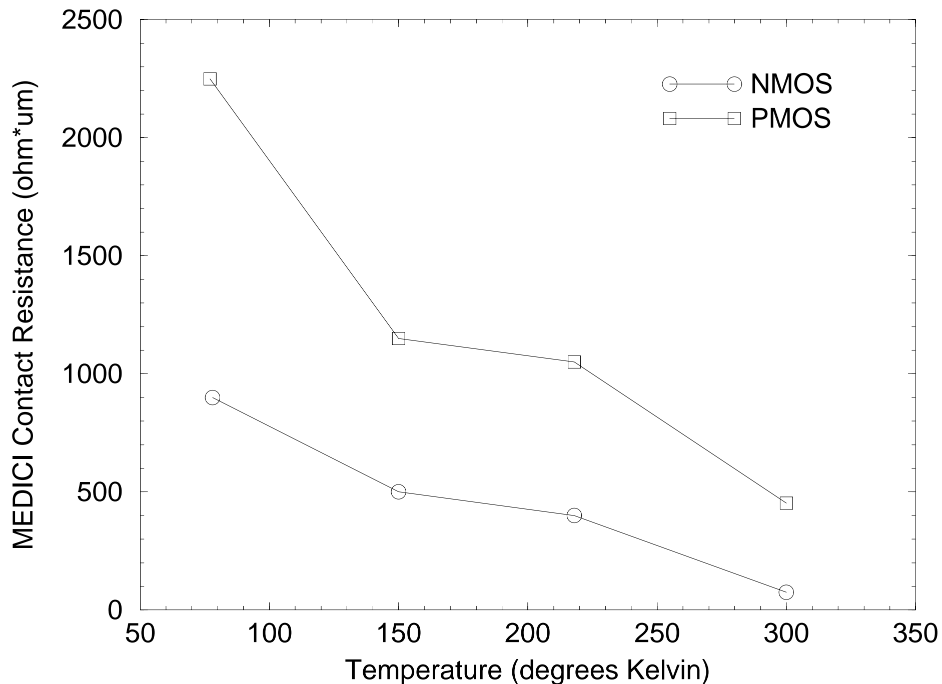

Figure 3.3: Parasitic source/drain resistance used in MEDICI to fit the measured data at different temperatures. The increase in resistance at lower temperatures is due to LDD freezeout.

source/drain resistance can be added in the input deck using the CONTACT card. Adding too much resistance will also lower the long channel simulation current. Figure 3.3 shows the source/drain resistance that was needed to obtain a good fit across all channel lengths at different temperatures. Similar to the HSPICE results (Figure 2.14), the source/drain resistance increases dramatically at 78 K, indicating freezeout in the LDD regions. Although MEDICI has an incomplete ionization feature, it cannot handle degenerate conditions. As a result, freezeout is still assume for doping concentrations well above Mott's transition. The entire source/drain region is treated as a freezeout region, resulting in extremely low currents and a poor I-V fits. Therefore, additional source/drain resistance is used instead of the freezeout option (in the MODELS card) to model freezeout effects here.

If an unacceptable fit still results in the linear I-V characteristics, process dependent parameters can be adjusted. These include the lateral diffusion (LATD) and the spacer thickness (which affects the LDD length and source/drain resistance). In addition, the drawn channel length can be altered, to account for poor dry etching of the poly gate or lithography errors. As a last resort, the doping profiles can be altered by either changing the process simulation parameters or numerically shifting the profile if it was created using Gaussian or uniform distributions.

Figure 3.4 shows the simulated and measured linear I_{ds}-V_{gs} characteristics at 78 K for NMOS transistors of different channel lengths. Since MEDICI is only a two-dimensional simulator, the currents are first multiplied by the device width (25 μm) before comparing with the measured results. In order to simulate narrow width effects, a three-dimensional device simulator, such as DAVINCI, is required. A good fit is obtained in the both the linear and subthreshold (Figure 3.5) regimes. At intermediate temperatures, linear extrapolation of the mobility degradation parameters still results in a good fit (Figure 3.6).

3.3.2 I_{ds} vs. V_{ds} With V_{gs} Varied

After obtaining a good fit in the linear I-V curves, the saturation characteristics should be calibrated. In order to avoid shifting the previous simulation results, only parameters relevant to high horizontal electric field effects should be adjusted. These include the saturation velocity (VSATN) and BETAN, which are set using the MOBILITY card and affect the mobility as follows:

$$\mu_n = \frac{\mu_{eff,n}}{[1 + (\frac{\mu_{eff,n} E_{||,n}}{VSATN})^{BETAN}]^{1/BETAN}} \tag{3.5}$$

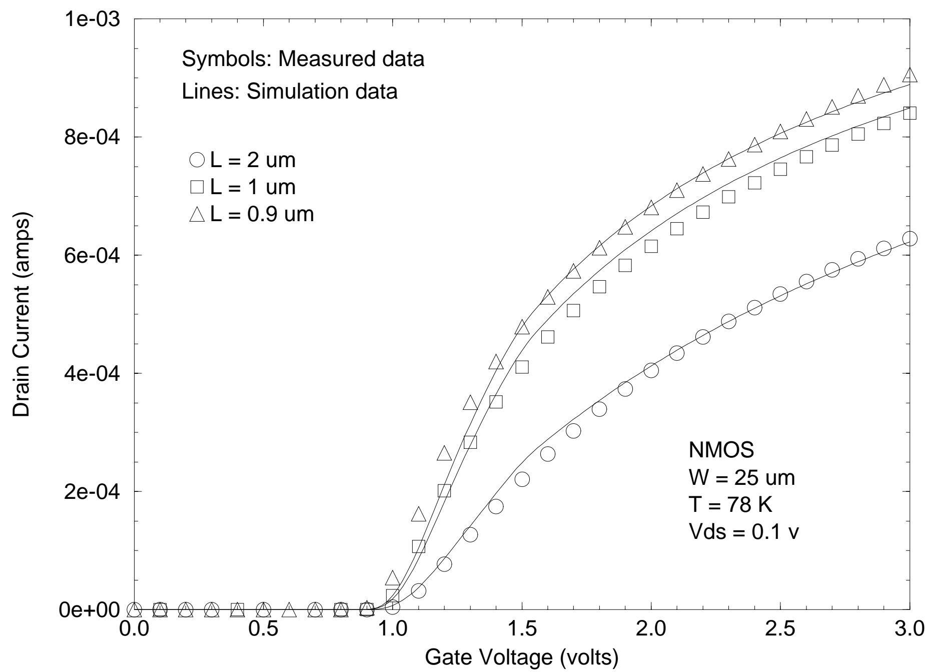

Figure 3.4: Simulated and measured I_{ds}-V_{gs} characteristics of NMOS transistors at 78 K (linear scale). The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

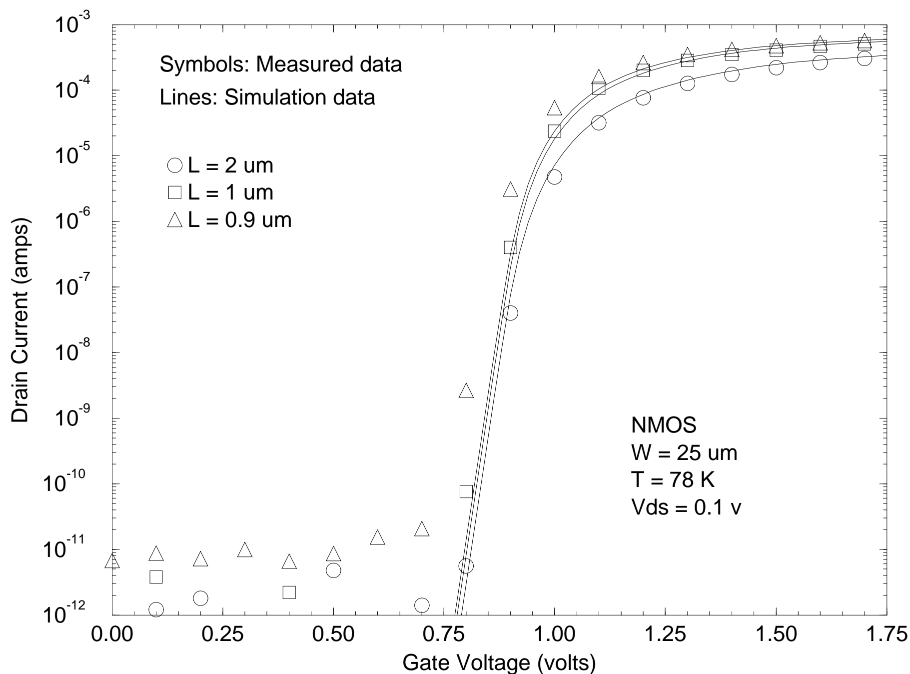

Figure 3.5: Simulated and measured I_{ds}-V_{gs} characteristics of NMOS transistors at 78 K (log scale). The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

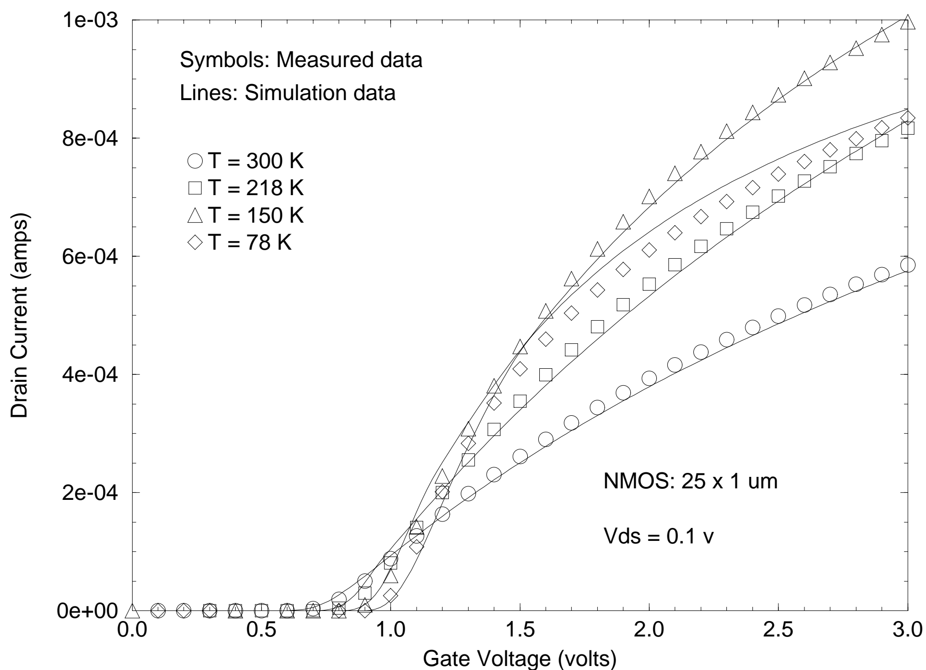

Figure 3.6: Simulated and measured I_{ds}-V_{gs} characteristics of a 25 μm x 1 μm NMOS transistor at various temperatures. Less than 5% error is achieved over a large range of temperatures. Due to the V_{th} shift and freezeout effects, the linear drain current at 78 K is lower than at 150 K.

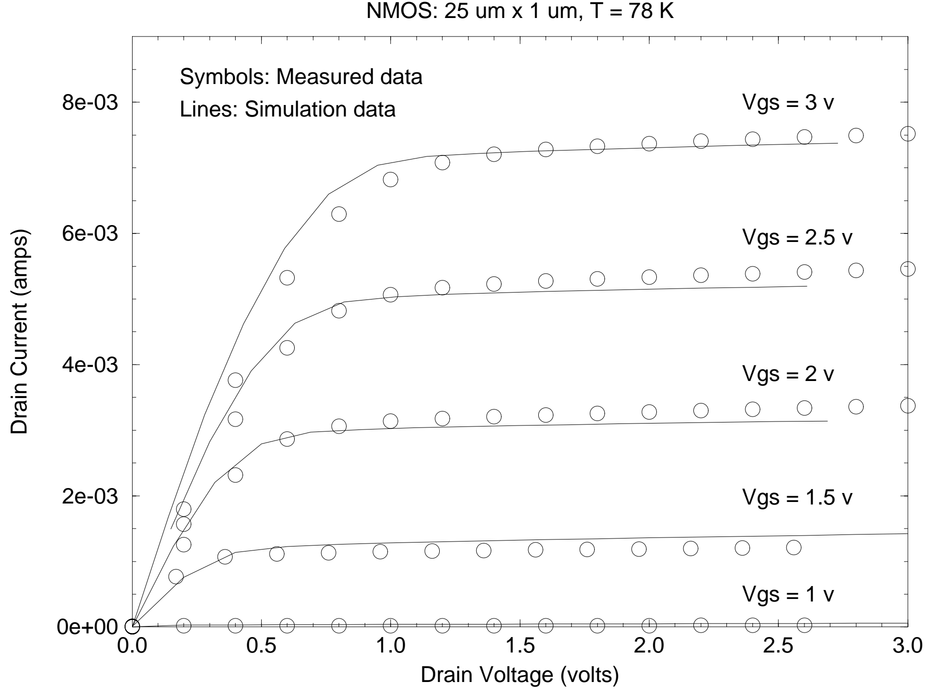

Figure 3.7: Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm NMOS transistor at 78 K. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

In the above equation, $\mu_{eff,n}$ is given by equation 3.1 and $E_{||,n}$ is the horizontal electric field along the Si-SiO_2 interface. Experimental work has shown that at lower temperatures, the saturation velocity increases, resulting in higher current drive. The value of BETAN can be treated as a fitting factor and increased to keep the change in saturation velocity reasonable. For the NMOS devices, BETAN was left at 2 at all temperatures, and the saturation velocity was increased up to $1.22x10^7$ cm/s at 78 K.

Figure 3.7 shows the simulated and measured I_{ds}-V_{ds} characteristics for a 25 x 1 μm NMOS device at 78 K. Again, good agreement is obtained across channel lengths (Figure 3.8) and temperatures (Figure 3.9). Since these curves cover a larger range

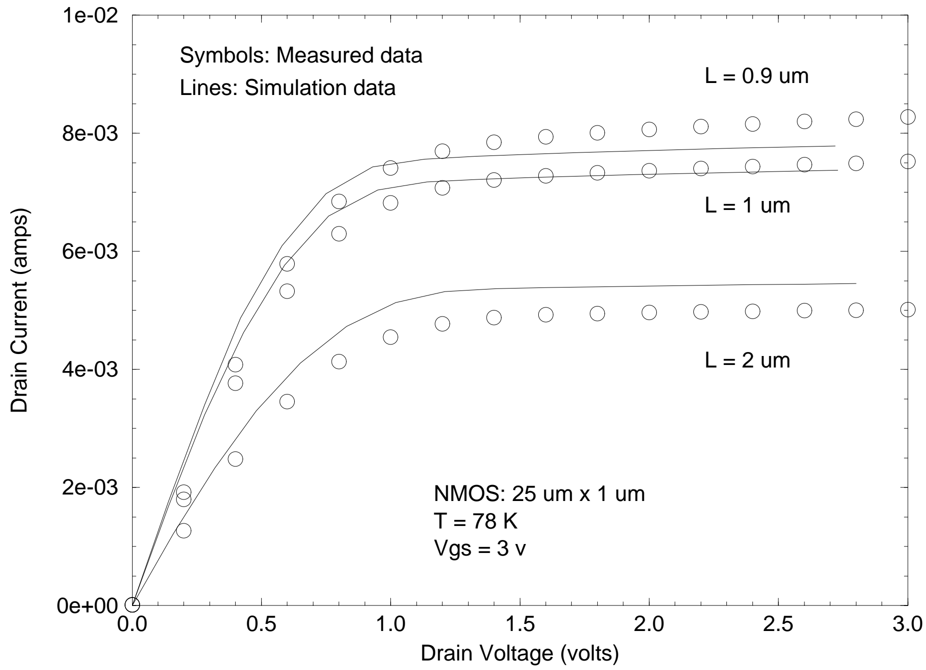

Figure 3.8: Simulated and measured I_{ds}-V_{ds} characteristics of NMOS transistors at 78 K. The gate bias for all three curves if 3 v. Since the devices are based on a $1\mu m$ CMOS process, the curves are fitted such that the $1\mu m$ device has the best fit.

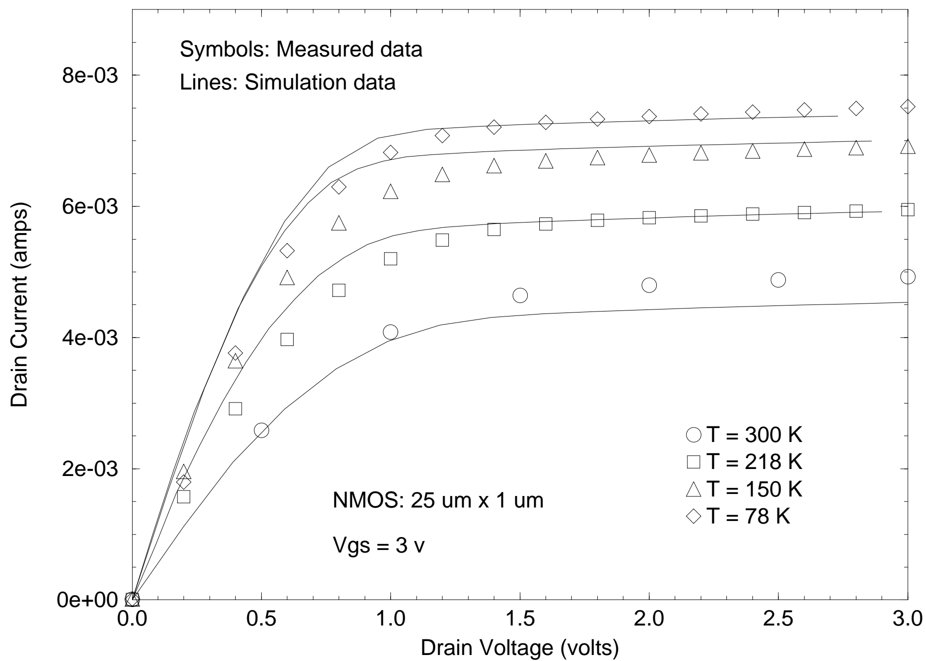

Figure 3.9: Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm NMOS transistor at various temperatures. The gate bias for all curves is 3 v. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

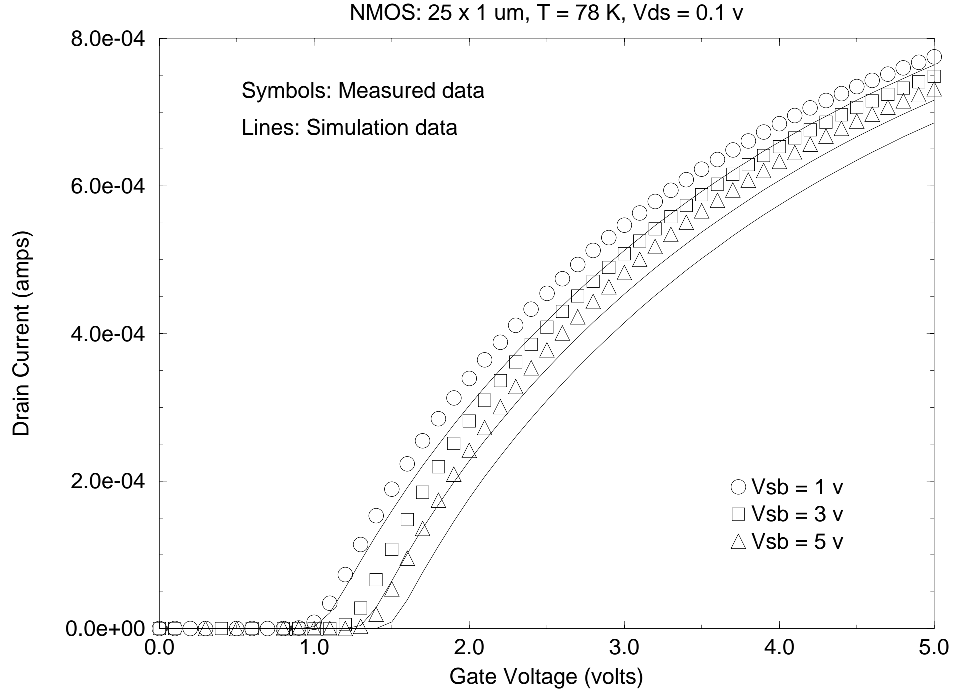

Figure 3.10: I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 300 K. The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

of voltages than the linear characteristics, it is essential to have good agreement here between simulation and measurement results.

3.3.3 I_{ds} vs. V_{gs} With V_{bs} Varied

I_{ds}-V_{gs} curves at low drain bias with varying substrate bias and temperature are shown in Figures 3.10-3.13. Adjusting these curves requires modifying the doping profile, which is not easy because doing so will alter the already fitted linear and saturation characteristics. Therefore, minimal work was done to improve the fit for these curves, since the simulation results were already close enough to the measurements.

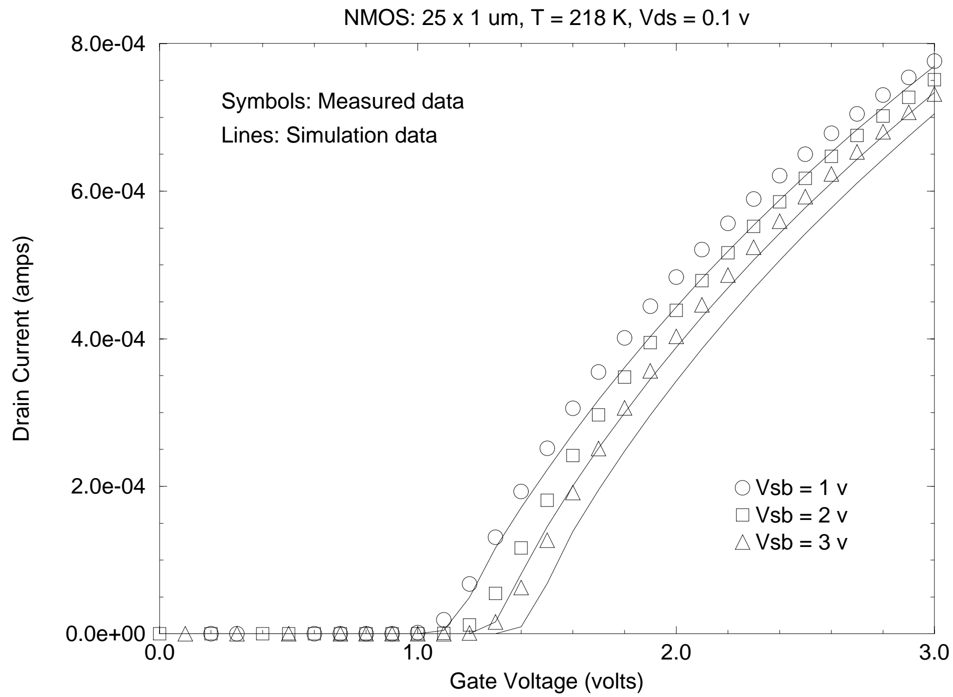

Figure 3.11: I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 218 K. The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

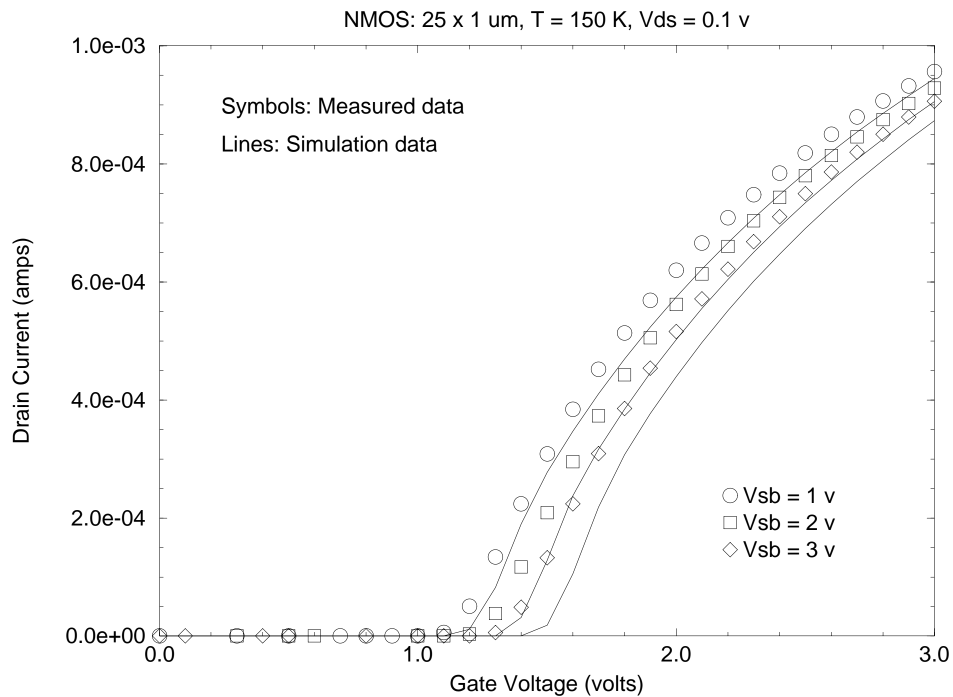

Figure 3.12: I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 150 K. The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

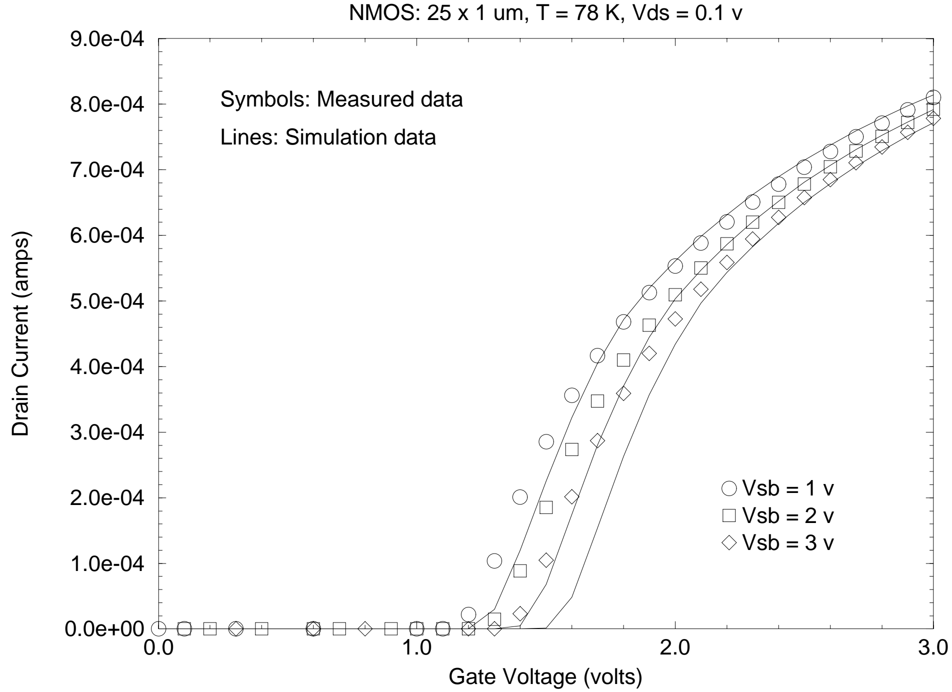

Figure 3.13: I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm NMOS device at 78 K. The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

3.4 PMOS Fit

Due to freezeout effects in the channel region, fitting the PMOS devices at low temperature is more challenging. This is further complicated by the fact that most of the PMOS current flows well below the Si-SiO_2 interface due to the gate material (n^+-poly) and threshold adjustment implant (boron). Nevertheless, adequate agreement between simulation and measurements is still obtained for these devices after adjusting some process parameters.

3.4.1 I_{ds} vs. V_{ds} With V_{gs} Varied

Figure 3.14 shows the simulated and measured I_{ds}-V_{gs} characteristics at 300 K for different channel lengths. In order to get a good match in the threshold voltage,

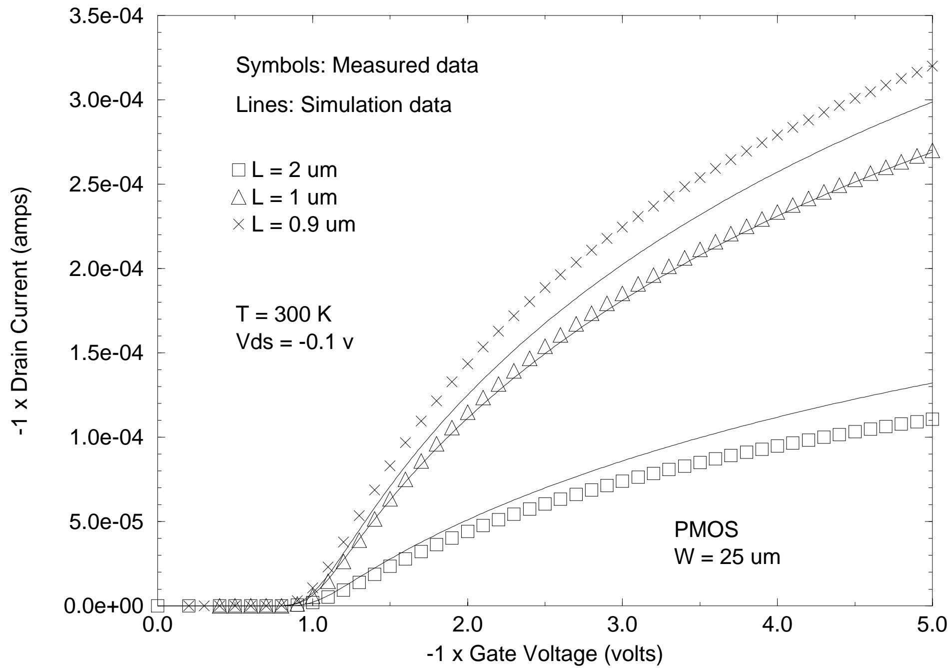

Figure 3.14: Simulated and measured I_{ds}-V_{gs} characteristics of PMOS devices at 300 K (linear scale). Points indicate measurement data. Solid lines indicate MEDICI simulation results. The punchthrough implant was made shallower to get a better V_{th} fit.

the punchthrough implant energy was decreased by 20%. The species used for this particular implant was doubly ionized phosphorous (P_2^{++}), which is cannot be directly simulated by SUPREM-III. Therefore, a slight adjustment to the equivalent singly-ionized phosphorous (P^+) implant can be justified.

At low temperatures, additional adjustments to the gate profile were necessary to obtain a good fit in the threshold voltage. The boron doping underneath the channel was decreased by 10%, 25%, and 40% to fit the threshold voltages at 218 K, 150 K, and 78 K, respectively. This reduction can be associated with freezeout effects (see Section 2.3.2), which reduces the ionized boron concentration. After this modification, a good fit across temperature is achieved (Figure 3.15).

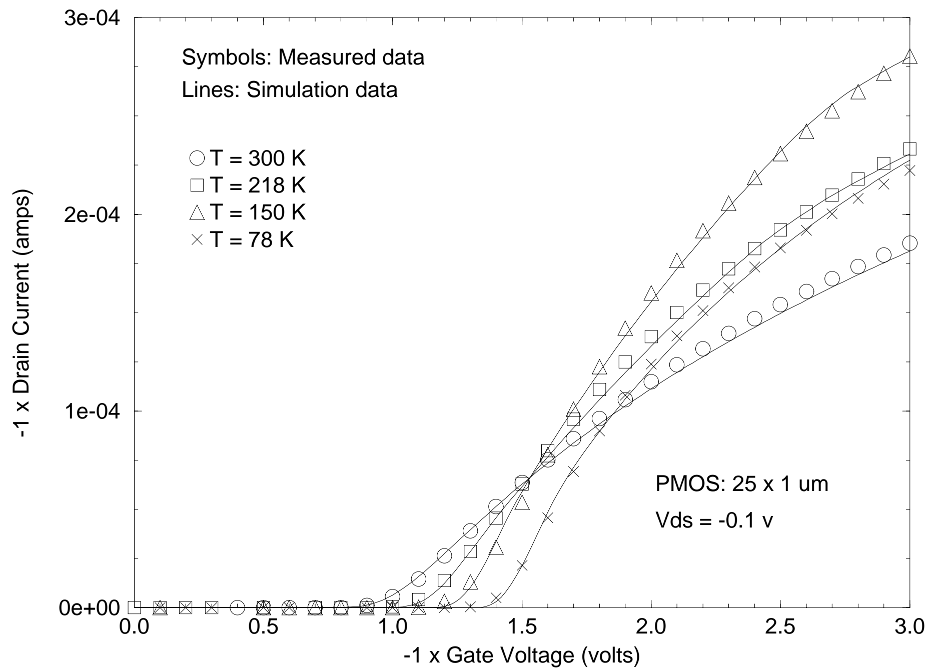

Figure 3.15: Simulated and measured I_{ds}-V_{gs} characteristics of a 25 μm x 1 μm PMOS transistor at different temperatures. Points indicate measurement data. Solid lines indicate MEDICI simulation results. At lower temperatures, the boron doping concentration is reduced by a greater percentage to account for carrier freeze-out.

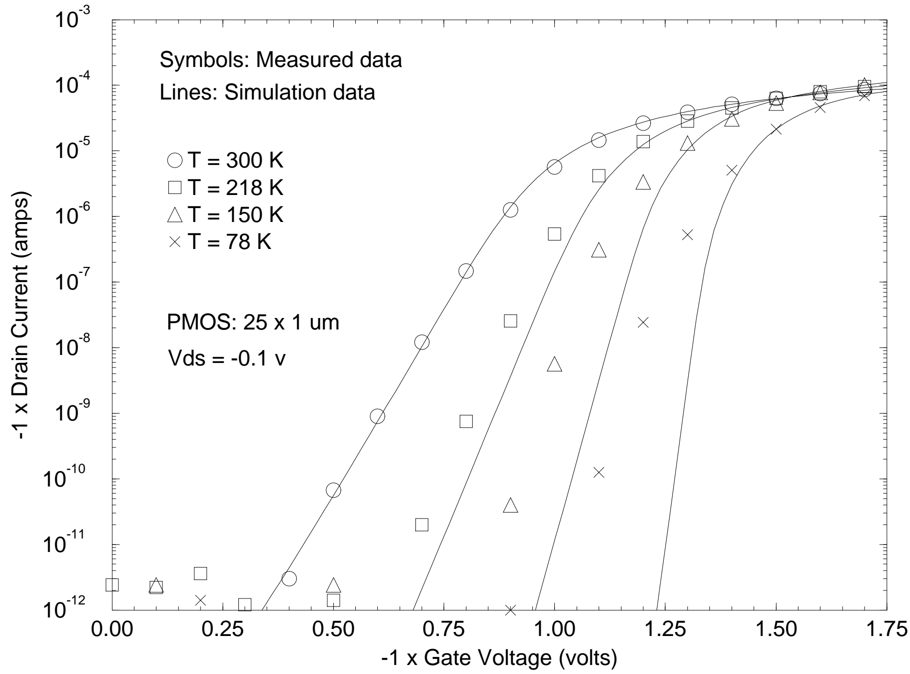

Figure 3.16: Simulated and measured I_{ds}-V_{gs} characteristics of a PMOS transistor at different temperatures (log scale). Points indicate measurement data. Solid lines indicate MEDICI simulation results. The poor fit at lower temperatures results from inadequate freezeout modeling in MEDICI.

Freezeout effects also degrade the subthreshold behavior. Figure 3.16 shows the subthreshold behavior of a 25 μm x 1 μm PMOS devices at different temperatures. A poorer fit is expected at lower temperatures since the carrier freezeout option is not used in the simulations. Instead, additional source/drain resistance and a fixed boron reduction are used, both of which cannot account for the variation in effective channel doping as a function of band bending, which leads to the poor subthreshold swing. Using the freezeout model should result in a better subthreshold fit, but the linear and saturation characteristics end up fitting much worse, as discussed in Section 3.3.1.

3.4.2 I_{ds} vs. V_{ds} With V_{gs} Varied

The results of fitting the simulated PMOS saturation characteristics at 300 K are shown in Figure 3.17. In order to keep the room temperature saturation velocity reasonably low, BETAP was increased from 1 to 3.5. The large shift is necessary because MEDICI incorrectly uses the Si-SiO_2 interface to calculate mobility degradation due to the vertical electric fields. The PMOS transistors in this case act as buried channel devices and should show less mobility degradation due to surface scattering effects. Instead of increasing BETAP to counteract the overdegradation of mobility, the parameters in the second term of 3.1 could have been reduced. By increasing the saturation velocity at lower temperatures, a good fit (Figure 3.18) can be achieved across a variety of channel lengths (Figure 3.19) all the way down to 78 K. BETAP is set to 1 for low temperature simulations due to difficulties in convergence. The final values of VSATP remain reasonable under this case.

3.4.3 I_{ds} vs. V_{gs} With V_{bs} Varied

Figures 3.20 and 3.21 show the linear I_{ds}-V_{gs} characteristics with varying substrate bias at 300 K and 78 K, respectively. In order to obtain a better fit, the punchthrough dose was decreased from $5x10^{12}cm^2$ to $3x10^{12}cm^2$. This reduced the horizontal shift in the I-V curves under the same substrate bias and matched the measurement results better. The uncertainty in modeling the P_2^{++} punchthrough implant in SUPREM-III can be used to justify this correction. With this process change, all previous curves need to be resimulated to confirm that there is still agreement between simulation and measurement.

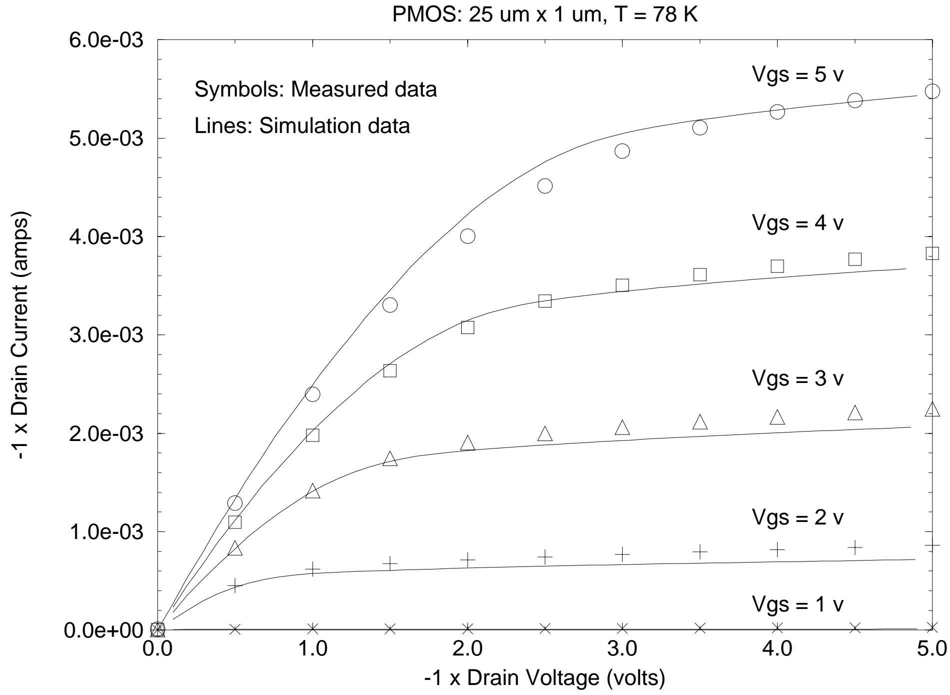

Figure 3.17: Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm PMOS transistor at 300 K. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

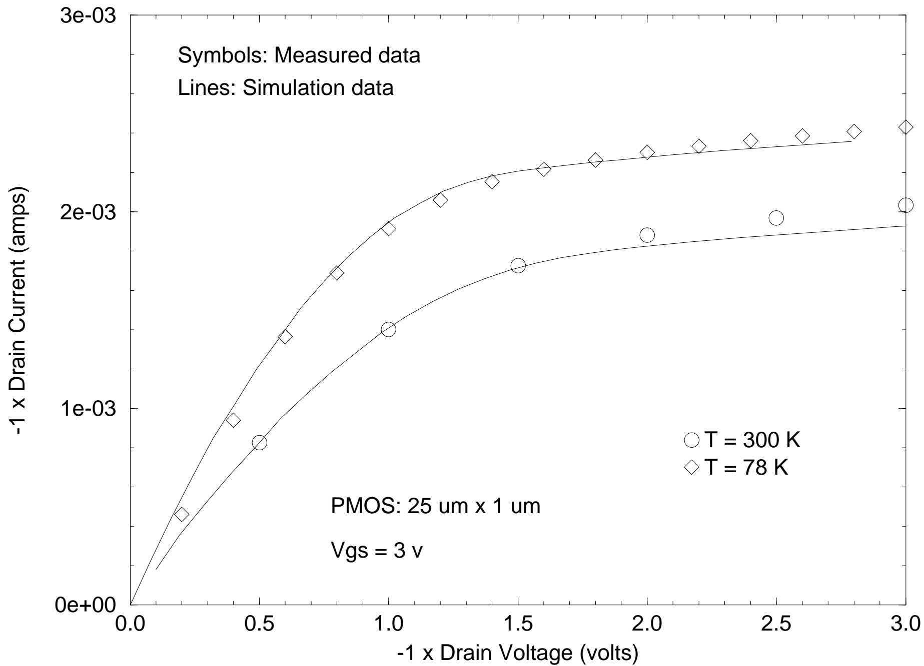

Figure 3.18: Simulated and measured I_{ds}-V_{ds} characteristics of a 25 μm x 1 μm PMOS transistor at 300 K and 78 K. The gate bias for all curves is 3 v. Points indicate measurement data. Solid lines indicate MEDICI simulation results. The saturation velocity is increased from $7.5x10^6$ cm/s to $1.2x10^7$ cm/s at 78 K.

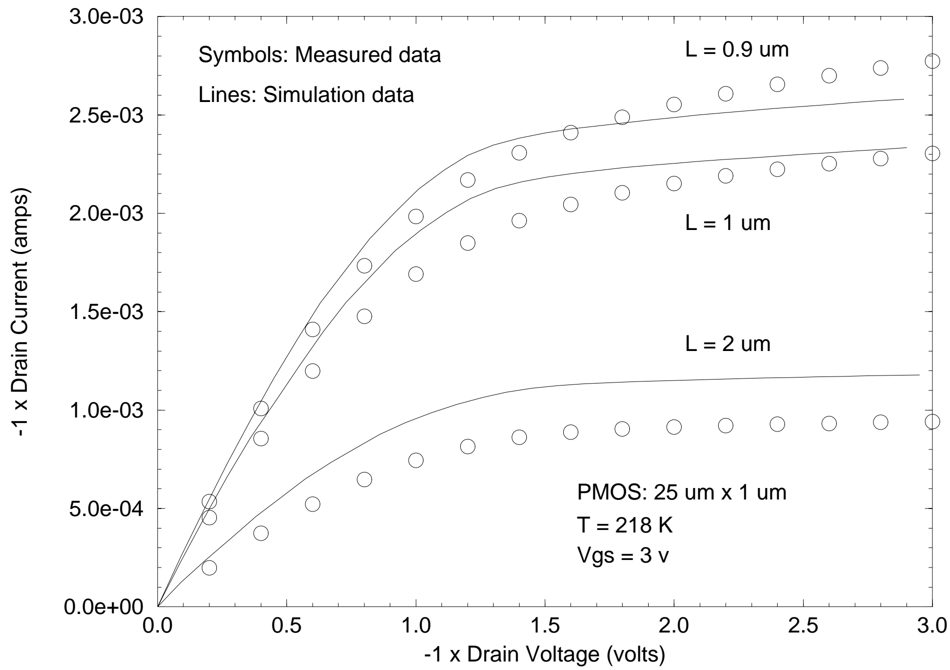

Figure 3.19: Simulated and measured I_{ds}-V_{ds} characteristics of PMOS transistors at 218 K. The gate bias for all curves is 3 v. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

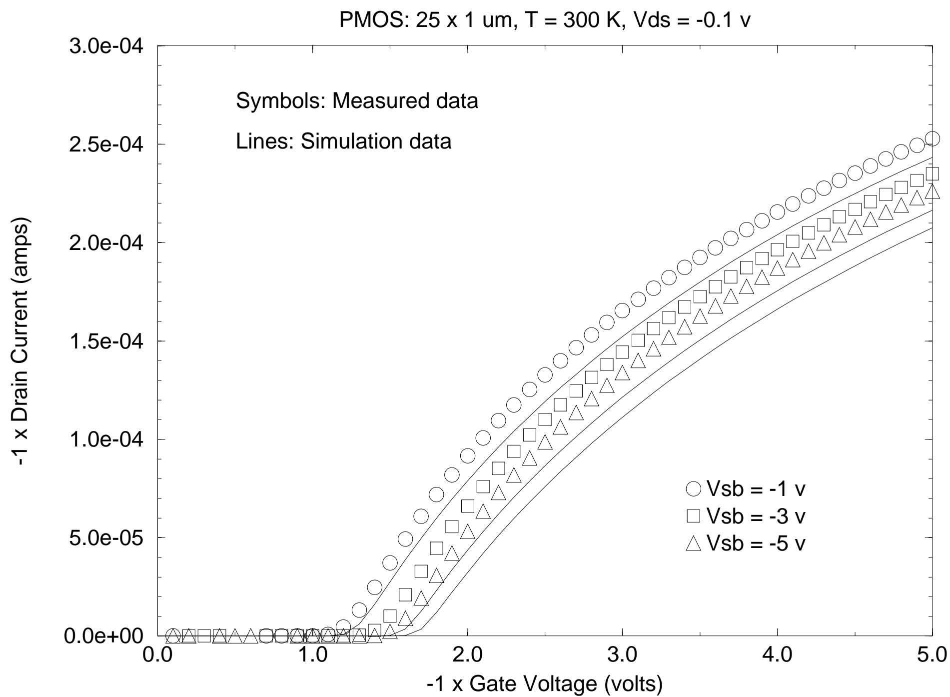

Figure 3.20: I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm PMOS device at 300 K. The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

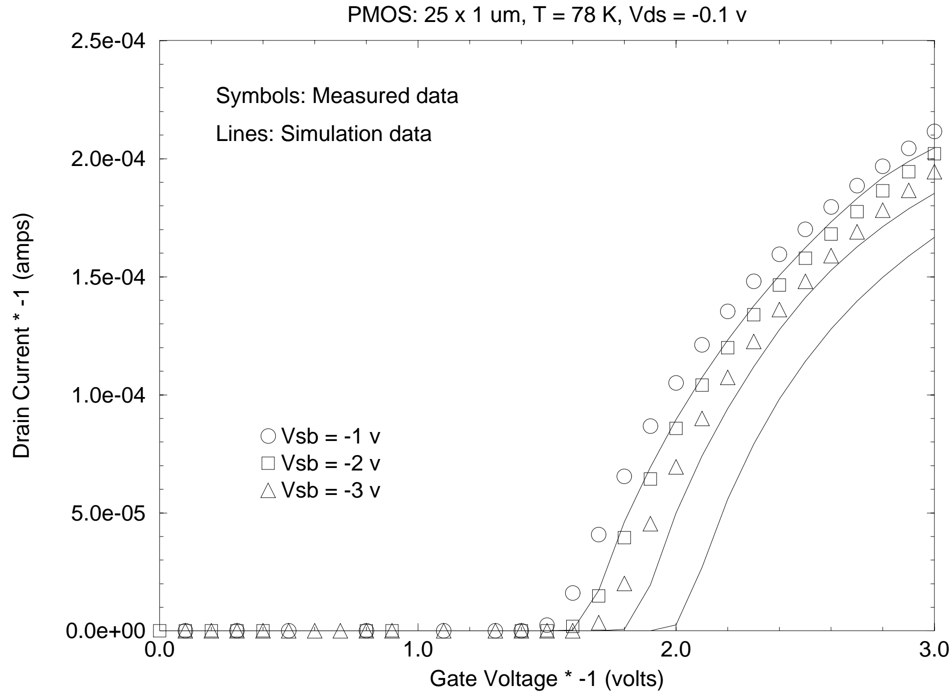

Figure 3.21: I_{ds}-V_{gs} curve with variable V_{sb} for a 25 x 1 μm PMOS device at 78 K. The channel width of all devices is 25 μm. Points indicate measurement data. Solid lines indicate MEDICI simulation results.

3.5 Conclusion

TCAD process, device, and circuit simulation tools can be used to design high performance CMOS for low temperature operation. Running simulations costs far less and yields faster results than physically fabricating the devices. Fewer process splits are needed if the design window can be narrowed through simulation. But in order to accurately predict CMOS behavior at low temperature, the tools must be initially calibrated to measurement results. By taking into account improvements in mobility and degradation due to freezeout, a good initial fit can be obtained across temperature, channel length, and bias condition. Additional tinkering of the process parameters (within reason) may be necessary to better fit the results.

Card	Option	Simulation Temperature			
		300 K	218 K	150 K	78 K
Models	Temperature	300	218	150	78
	Boltzmann	True	False	False	False
	FermiDir	False	True	True	True
Mobility	MuN0 ($\frac{cm^2}{V\ s}$)	default value	900	1500	2000
	EtaN	default value	0.59	0.67	0.75
	SrfMob2 parameters	default values	linearly extrapolated	linearly extrapolated	[16]
	VSatN (cm/s)	$7.5x10^6$	$9.1x10^6$	$1.05x10^7$	$1.22x10^7$
Contact	S/D Resistance (ohm*um)	75	400	500	900

Table 3.1: Parameters used in NMOS MEDICI simulations.

Tables 3.1 and 3.2 summarize the parameters which were modified in order to obtain good agreement between the simulated and measured I-V characteristics for a given temperature. As expected from experimental observations, the modeling parameters indicate an improvement in mobility and an increase in LDD freezeout at lower temperatures.

To obtain a better fit, certain process parameters were also modified. The NMOS and PMOS channel lengths were increased to achieve better agreement with the measurement results. In addition, the PMOS channel profile was adjusted to account for freezeout effects, which cannot be accurately simulated by MEDICI. The procedure and methodology for fitting the I-V characteristics are generalized enough where they can be applied to any operating temperature and extrapolated to any TCAD device simulation tool.

Card	Option	Simulation Temperature			
		300 K	218 K	150 K	78 K
Models	Temperature	300	218	150	78
	Boltzmann	True	False	False	False
	FermiDir	False	True	True	True
Mobility	MuP0 ($\frac{cm^2}{V\,s}$)	225	280	400	455
	EtaP	default value	0.37	0.41	0.45
	SrfMob2 parameters	default values	linearly extrapolated	linearly extrapolated	[16]
	VSatP (cm/s)	$7x10^6$	$6.5x10^6$	$6x10^6$	$1.22x10^7$
	BetaP	3.5	3.5	3.5	3.5
Contact	S/D Resistance (ohm*um)	452.5	1050	1150	2250

Table 3.2: Parameters used in PMOS MEDICI simulations.

Chapter 4

Redesigning CMOS for Improved Low Temperature Performance

4.1 Introduction

Significant improvements are expected in CMOS as the operating temperature is reduced. In order to take full advantage of these enhancements, devices need to be redesigned so that undesirable low temperature effects, such as freezeout, are minimized. With TCAD design tools properly calibrated at the operating temperature of interest, proposed design changes can be thoroughly studied and evaluated before being implemented. In this chapter, a conventional, room-temperature process flow will be modified so that better low temperature performance is achieved. First, the threshold voltage will be reduced so that higher current drive is obtained without an increase in off-state leakage. Next, the channel profile will be altered so that better mobility improvement will occur. Re-engineering of the source/drain regions will be

necessary to minimize freezeout effects. And finally, the advantages of using dual-polarity gates (i.e. p^+-poly PMOS) for low temperature operation will be addressed. Simulations will be used to support some of the proposed design changes.

4.2 Design Methodology

Figure 4.1 illustrates the general approach which will be used to design CMOS for low temperature operation. A conventional, high performance, room temperature process flow will be used both as a starting point and for comparison. After calibrating the tools so that the simulation results match the measurements (see Chapter 3), proposed changes in the process flow are first run through SUPREM to obtain new doping profiles. MEDICI then imports the profiles and produces I-V characteristics. Improvement on the device level can be obtained by comparing these new I-V curves with results based on the baseline, room temperature process flow. To determine circuit level improvement, extraction of parameters for HSPICE [17] simulations should be performed. By just slightly modifying an already existing process flow, one can better understand the impact of certain process steps on low temperature CMOS performance. In addition, the cost of implementing these changes is far less than designing and testing a completely new process flow.

4.3 Threshold Voltage Adjustment

4.3.1 To Room Temperature Value

Using the aforementioned technique, the threshold voltage implants were modified to decrease the high threshold voltage at low temperature. The NMOS V_{th} implant

Figure 4.1: Illustration of how TCAD process (SUPREM), device (MEDICI), and circuit (HSPICE) simulation tools are used together to improve the performance of low temperature CMOS. Input decks similar to those used to fit the original, measured results are used as the starting point.

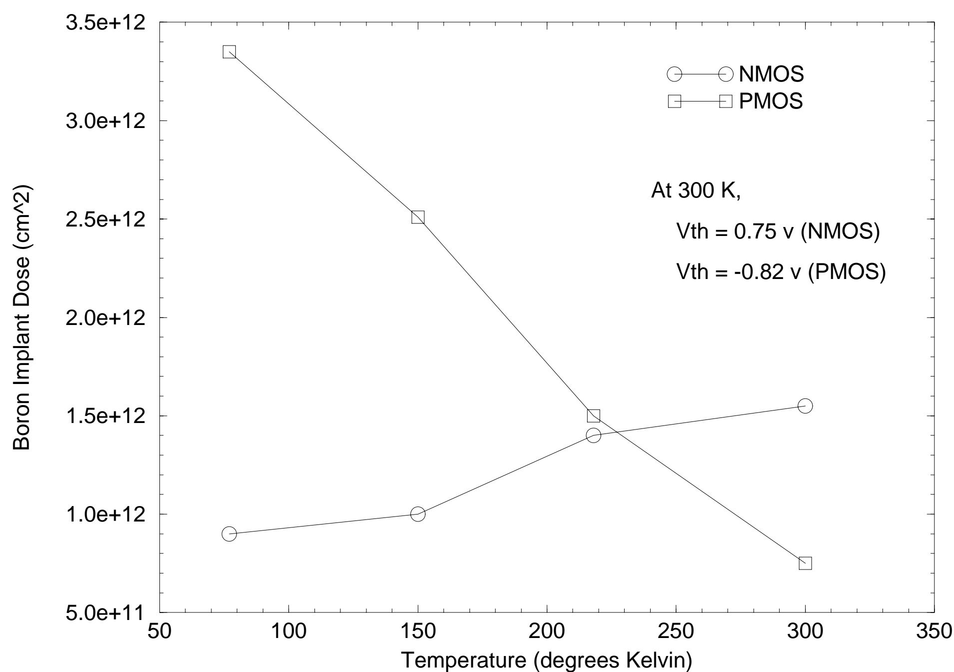

Figure 4.2: Threshold adjustment implant doses (boron) necessary to maintain a 300 K V_{th} for a given operating temperature.

dose was decreased, and the PMOS V_{th} implant dose was increased. By not adjusting the threshold voltage, too large of an "off-zone" will exist at low temperatures, which will limit current drive and switching performance. Figure 4.2 shows the implant doses needed to obtain the same room temperature threshold voltage. As expected, a larger change is needed for the PMOS devices in order to counter channel dopant freezeout. However, the changes are not significant enough to worsen short channel effects.

The improvement in NMOS linear drain current at 78 K after readjustment is illustrated in Figure 4.3. A better match with the 300 K I-V curve is obtained when the threshold voltage is defined by extrapolating back from the peak transconductance (as opposed to using a constant current level). Although the linear current improves only slightly, the saturation current improves by a larger amount (Figure 4.4). Nevertheless, the linear drain current improvement is an important factor to consider in circuit switching because the output node must swing all the way between V_{dd} and zero. Between zero volts and the drain saturation voltage, V_{dsat}, the transistor will operate in the linear regime. Although V_{dsat} decreases at lower temperatures, it can still take up a significant portion of the output swing, especially for low V_{dd}.

The impact of the threshold adjustments on the circuit level can be determined by extracting the new I-V characteristics to HSPICE. The resulting parameters were similar to those obtained for the original, measured I-V characteristics. Figure 4.5 shows the simulated propagation delays of a 41 stage ring oscillator at 78 K for the original and threshold adjusted cases. Delay improvement at 218 K, 150 K, and 78 K is illustrated in Figure 4.6. Low supply voltages now show significant

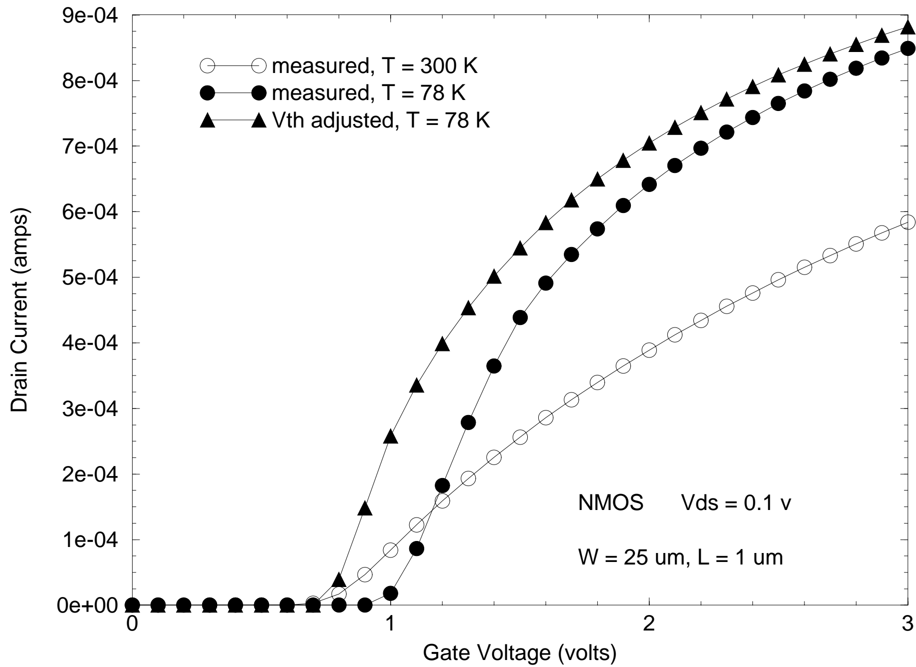

Figure 4.3: Improvement in linear drain current of a 25 μm x 1 μm NMOS device at 78 K after threshold adjustment. The original 78 K I-V curve is shown for comparison. A better match with the 300 K threshold voltage is evident. Freezeout in the LDD region limits the improvement in current and high V_{gs}.

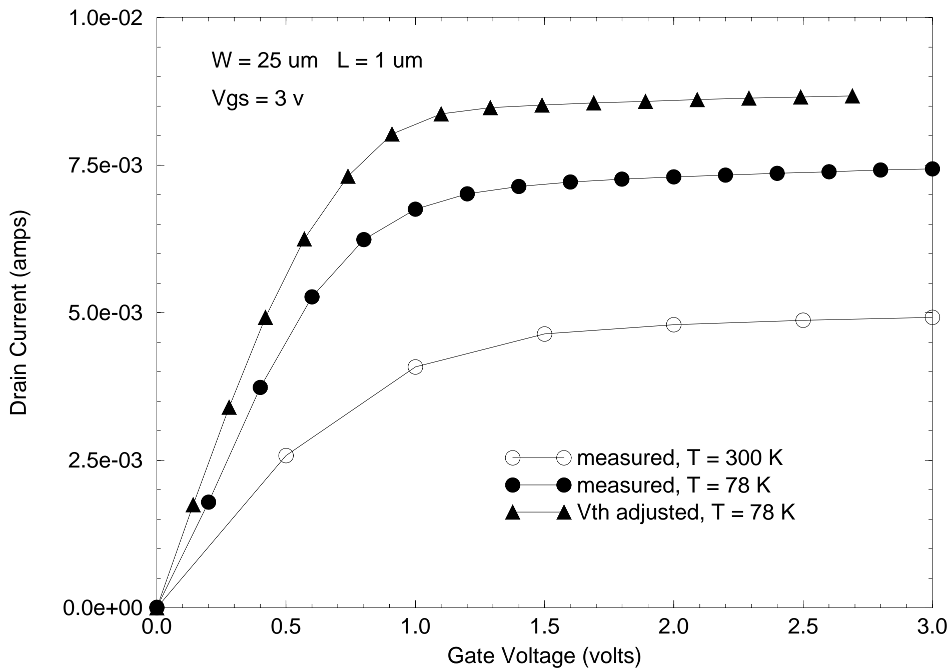

Figure 4.4: Improvement in saturation drain current of a 25 μm x 1 μm NMOS device at 78 K after threshold adjustment. The original 78 K I-V curve is shown for comparison.

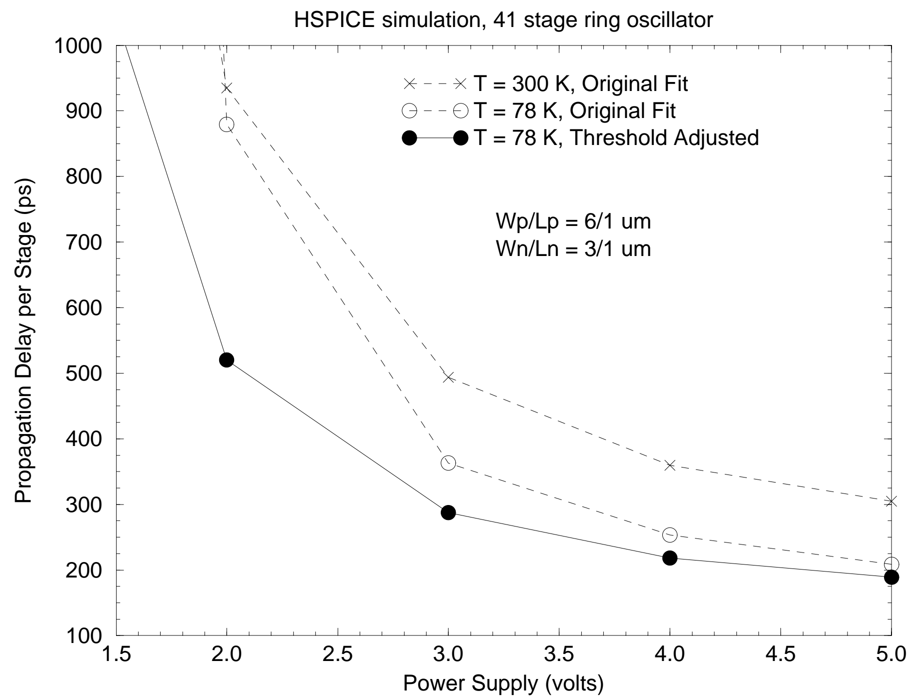

Figure 4.5: Propagation delay versus power supply voltage of a 41 stage ring oscillator at 78 K after threshold implant adjustment. Original results (Figure 2.6) are shown for comparison. A much larger decrease in propagation delay is evident.

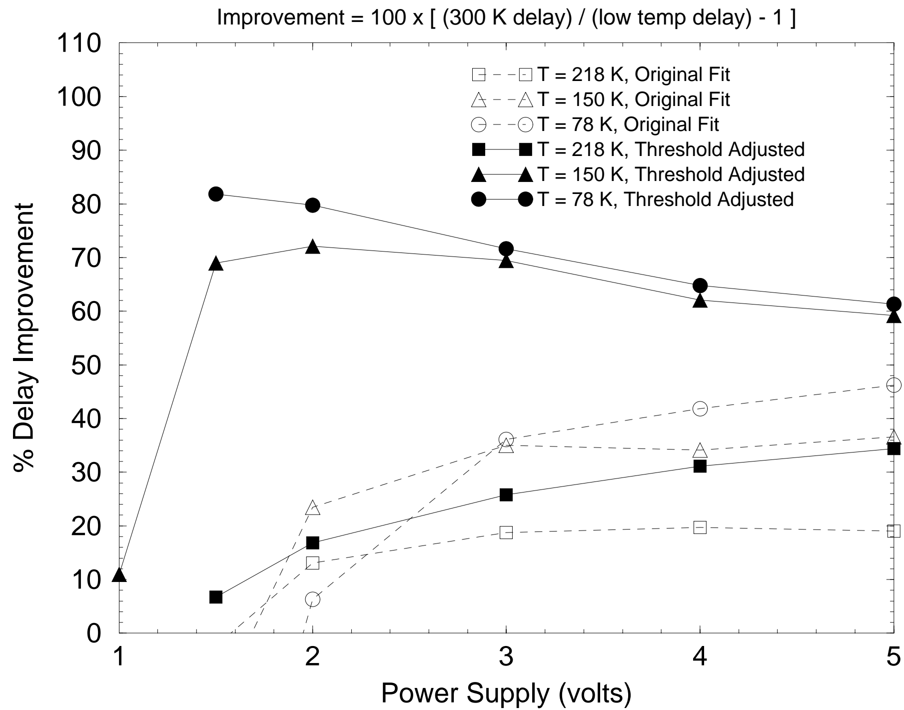

Figure 4.6: Improvement in ring oscillator delay versus power supply voltage as a result of threshold implant adjustment. Delays at 218 K, 150 K, and 78 K were simulated using HSPICE (41 stage ring oscillator: $W_p/L_p = 6/1\mu m$ and $W_n/L_n = 3/1\mu m$) and normalized to 300 K. Original results (Figure 2.7) are shown for comparison.

improvement. The change in improvement is greater at lower temperatures due to the larger shift in threshold voltage. Freezeout of the LDD regions still limits the improvement between 150 K and 78 K. Since only the channel profile was changed, minimal change from 150 K to 78 K is still expected. By increasing the dose of the LDD implant to reduce the high parasitic resistance (see section 4.5.2), enhanced 78 K circuit performance is possible.

4.3.2 Below Room Temperature Value

Due to the improvement in subthreshold slope (see Figure 2.1), a lower than room temperature threshold voltage can be used at low temperatures without increasing the off-state leakage. Figure 4.7 shows the amount of performance enhancement possible at 78 K when the threshold voltage is 25% lower than the value at room temperature. To run this set of HSPICE simulations, only VTO was changed. At high power supply voltages, freezeout still limits the improvement. However, a 2x (100%) improvement in delay is easily achieved at low power supply voltages, which will be the desired operating voltage, as suggested by the lifetime data (see Figure 2.15). At 218 K, an additional 20% improvement is obtained at 3 v (Figure 4.8).

Reducing the threshold voltage to 25% below the room temperature value is somewhat arbitrary. As long as the off-state current stays below the value at room temperature, the threshold voltage can continue to be reduced. At room temperature, a 25 μm x 1 μm NMOS device has a threshold voltage of 0.57 v, a subthreshold slope of 86 mV/decade, and an off-state current of 4 pA. If the desired operating temperature is 78 K, the threshold voltage can be lowered all the way down to 0.13 v (Figure 4.9) without increasing the off-state current since the subthreshold slope

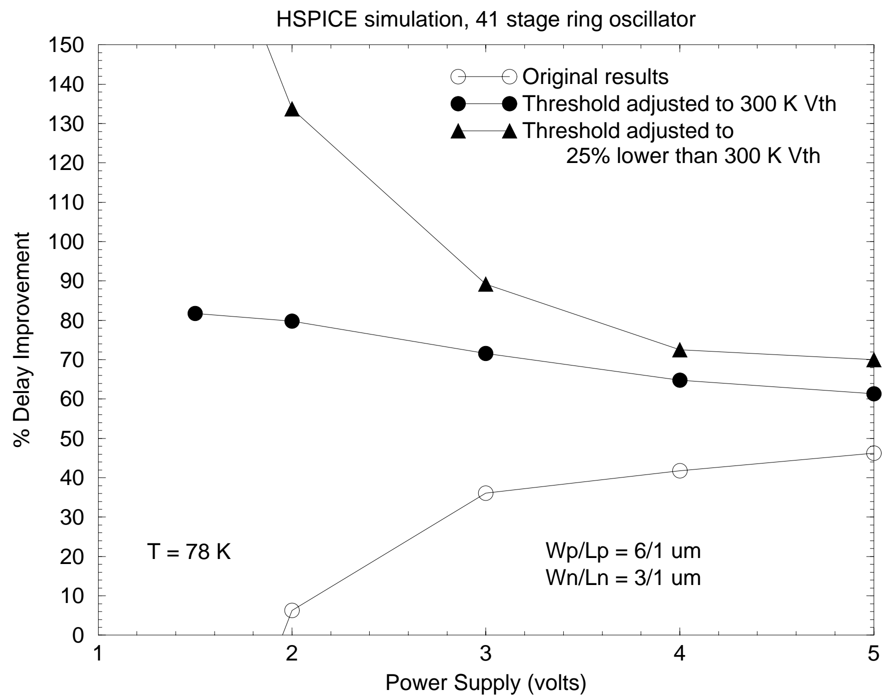

Figure 4.7: Improvement in delay at 78 K versus power supply voltage for a lower than room temperature threshold voltage. Delay improvements for the original (Figure 2.7) and the threshold readjustment to room temperature case (Figure 4.6) are shown for reference. A 2x improvement in delay at low power supply (2.7 v) is possible with the 25% V_{th} reduction.

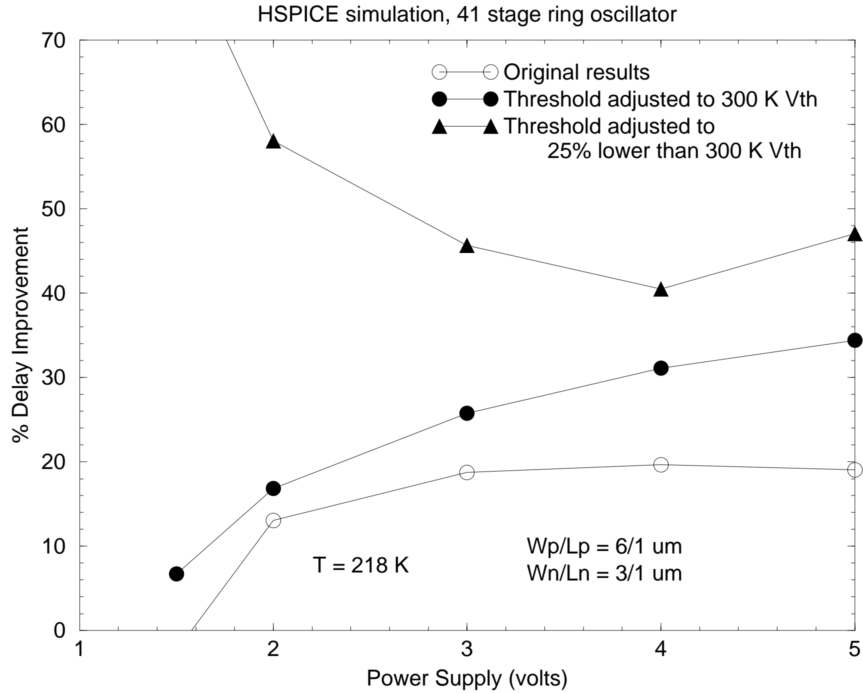

Figure 4.8: Improvement in delay at 218 K versus power supply voltage for a lower than room temperature threshold voltage. Delay improvements for the original (Figure 2.7) and the threshold readjustment to room temperature case (Figure 4.6) are shown for reference.

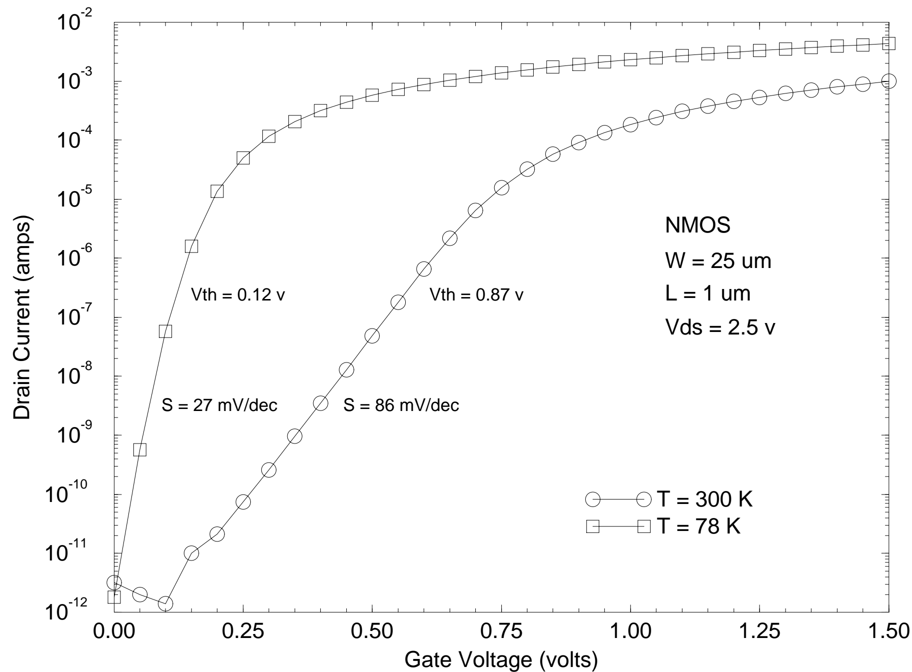

Figure 4.9: Optimized threshold voltage for NMOS transistors at 78 K. V_{th} is chosen such that the off-state current matches the results at 300 K. The 78 K curve is measured data that has been shifted to the left by 0.75 v.

is only 27 mV/decade. This is over 75% below the room temperature value. Therefore, an even larger improvement in 78 K device and circuit performance than what is shown in Figure 4.7 can be expected.

Having a very low threshold voltage at 78 K introduces several new issues which need to be considered. Since the threshold voltage increases by roughly 0.3 v from 300 K to 78 K, a 0.13 v NMOS threshold voltage at 78 K translates to a -0.17 v threshold voltage at room temperature. As a result, testing and evaluation of circuits can only be done by first cooling down the circuitry. In addition, if the cooling system fails, the circuit will no longer function correctly. Low threshold voltage devices are also more susceptible to process variations. A variation in threshold voltage due to poor gate oxide quality or interface states will dramatically affect the leakage current. Less aggressive scaling of the threshold voltage can minimize this problem, with the trade-off being a loss of additional performance improvement. Good device isolation also becomes critical in designing a low threshold, low temperature device. A poor field implant can lead to larger than expected leakage currents.

4.4 Retrograde Channel Profiling

Reducing the threshold voltage by modifying the channel implant dose leads to significant improvement in low temperature performance. The threshold voltage can alternatively be lowered by increasing the implant energy. This creates a retrograded doping profile where the concentration near the oxide-silicon interface is lower than what it is further into the substrate. Retrograde profiling has been studied at room temperature as a means of improving short channel behavior without significantly

increasing the device capacitance. However, its performance has received mixed reviews [19]-[20].

At low temperatures, improved mobility is expected if a retrograded profile is used. This results from the fact that mobility is limited by impurity scattering at low temperatures [21]. Figure 4.10 illustrates the effective electron mobility for different doping concentrations at 300 K and 78 K. Having a lightly doped channel is extremely desirable at low temperature, especially at low voltages. However, in order to maintain good short channel behavior, the doping concentration needs to be high. Both requirements can be met by using a retrograded profile. The region immediately underneath the Si-SiO_2 interface (where most of the carriers flow) is lightly doped, to take advantage of the improved mobility, while the area further underneath is more heavily doped, to prevent subsurface punchthrough and poor short channel behavior.

The fast diffusivity of boron in silicon makes it difficult to form a retrograded profile using this species for the threshold implant. As an alternative, a heavier implant species, such as indium, can be used. Figure 4.11 illustrates the doping profiles obtained using SUPREM-III for different implant species and anneals. Since indium was not available in this simulation tool as an implant species, a similarly weighted ion (antimony) was used, and the signs in the resulting doping profile were reversed. By reducing the thermal cycles after the threshold implant, a more retrograded profile can be obtained with boron. The gate oxide anneal can be removed without a significant change in reliability, and rapid thermal annealing can be used instead of a furnace to activate all of the dopants. Nevertheless, using indium results in a far better retrograded profile.

Figure 4.10: Effective electron mobility versus field at 300 K and 78 K (from [16]). At room temperature, the mobility is independent of doping (phonon scattering limited). At low temperature, the mobility is depends on the carrier concentration. A similar trend occurs for holes.

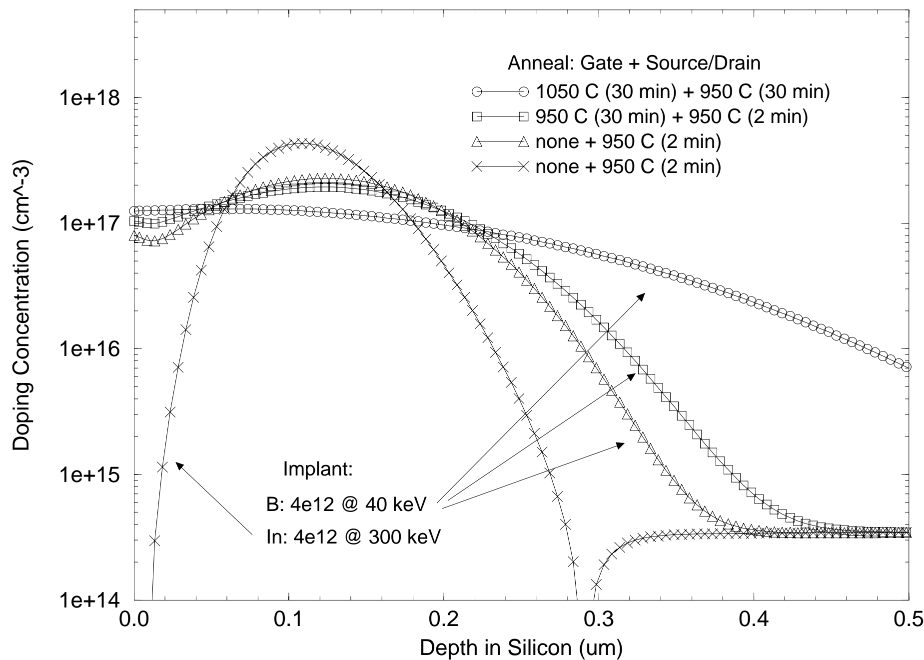

Figure 4.11: SUPREM-III doping profiles underneath the gate for different implant species and anneals.

4.5 Source/Drain Engineering

As channel lengths are scaled down to deep sub-micron dimensions, the source/drain regions take up a larger portion of the overall device area. Therefore, proper design of these areas is essential, especially at low temperature, where freezeout can introduce significant parasitic resistance.

4.5.1 Halo Implant

Similar to a retrograde implant, a halo implant can be used to enhance device performance. Instead of performing a retrograded threshold adjustment implant, a large angle implant is done prior LDD spacer formation (Figure 4.12). This results in a small, highly doped pocket near the source and drain, which minimizes

Figure 4.12: Illustration of a halo implant for an NMOS device. The majority of the channel remains lightly doped, allowing for better mobility improvement at low temperature, while maintaining good short-channel behavior due to the high doping near the source and drain.

punchthrough. Most of the region underneath the channel remains lightly doped and will exhibit higher mobility at low temperatures. Unlike retrograded profiles, the improvement should occur regardless of whether or not the device is buried or surface channel. Halo implants also offer a wider design window than retrograde implants since the tilt angle of the implant can be adjusted to optimize performance improvement. Like retrograde designs, much of the existing work on halo implants has focused on minimizing short channel effects [22]-[23], but with reduced operating temperature, halo implants should also further enhance mobility and current drive.

4.5.2 LDD Implant

In Section 2.3.3, it was shown that conventional LDD designs result in carrier freezeout at low temperature because the doping concentration falls below Mott's transition ($4x10^{18}cm^{-3}$). In fact, the reliability can be even worse at low temperature than non-LDD devices [24]. To prevent carrier freezeout, the LDD dose should be increased so that the doping concentration is above Mott's transition. In addition, the thermal cycle can be reduced so that less diffusion of dopants in the LDD region occurs. In this case, the LDD region should actually be thought of as a shallow, source/drain extension region, due to the higher doping level.

MEDICI simulations can be performed to examine the effect of replacing the LDD with a source/drain extension. For maximum reliability, the peak electric fields should be kept low and away from the spacer region for maximum reliability. With the trend towards using reduced power supplies, hot carrier effects and proper LDD design are becoming a less critical issue in terms of device reliability. In order for charge injection to occur into the spacer and thus lead to device degradation, a fixed energy barrier of roughly 3.5 eV must be surmounted. Under low power operation, there is less probability that a carrier can generate this energy to inject itself into the spacer.

Increasing the LDD implant dose and reducing the thermal cycle has been shown to result in better low temperature CMOS performance and reliability for a 1 μm CMOS technology [25]. Freezeout effects can be removed from the 1 μm CMOS technology presented in Chapter 2 by extrapolating the HSPICE source/drain resistance from the results at 343 K, 300 K, 218 K, and 150 K (see Figure 2.14). This results in an additional 40% improvement in CMOS performance at 78 K (Figure 4.13).

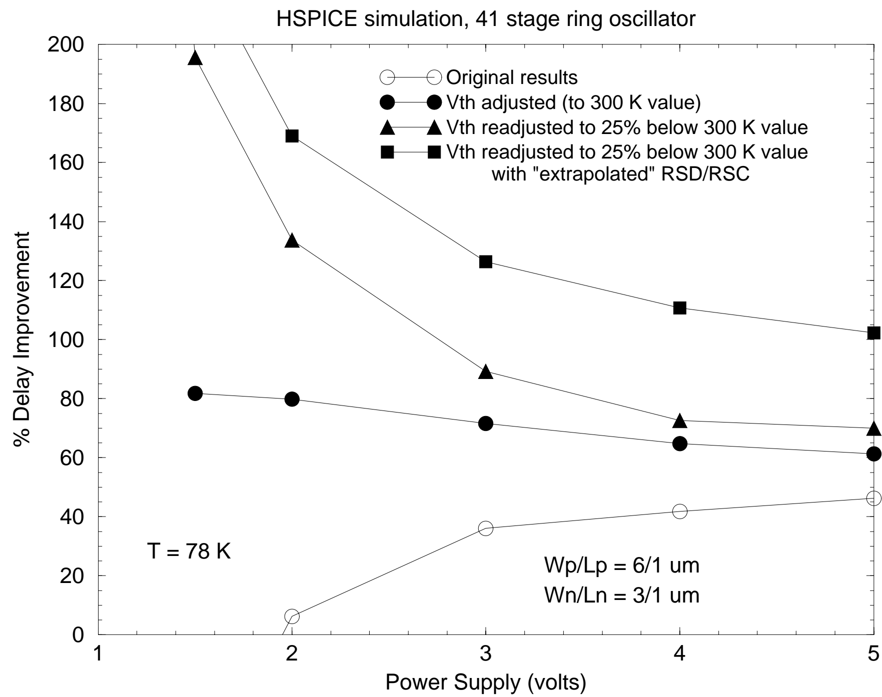

Figure 4.13: Improvement in CMOS ring oscillator performance at 78 K if no freeze-out occurs. Original results, threshold adjusted, and threshold reduced cases are shown for reference.

4.6 Gate Engineering

Freezeout of the PMOS channel dopants, which leads to a larger shift in threshold voltage and poorer improvement in subthreshold slope, can be eliminated by switching the polarity of the polysilicon gate. By using a p^+-poly gate instead of an n^+-poly gate, phosphorous is used (instead of boron) to adjust the PMOS threshold voltage. As a result of the band bending (Figure 2.9), the donor level of phosphorous (located just below the conduction band) remains far away from the Fermi level near the threshold voltage and will not freezeout.

With a p^+-poly gate, the PMOS transistor now operates as a surface channel device instead of buried channel. Recent work [26] has shown that surface channel PMOS devices exhibit better performance than buried channel. Using a PMOS retrograded profile can now lead to additional mobility improvement at low temperatures. Since phosphorous is much heavier than boron, retrograded profiles can also be more easily created.

Now that dual-polarity gates are employed (n^+-poly for NMOS and p^+-poly for PMOS), POCL doping of the polysilicon immediately after deposition should be not be performed. Instead, the polysilicon can be doped using the NMOS and PMOS source/drain implant. In order to create shallow source/drain junctions and also adequately diffuse the dopants all the way down to the poly-oxide interface, the polysilicon thickness must be kept reasonably thin. If thick polysilicon is required, two additional masking steps and deeper n and p implants can be performed immediately after the polysilicon deposition. However, care must be taken not to overdope the poly (which can cause dopant segregation and high resistance or depletion effects), since an additional source/drain implant will occur later in the process flow.

Furthermore, etching of the polysilicon to form the gates may be problematic in this case since the etch rates of n^+-poly and p^+-poly are typically different.

The main disadvantage of switching to a p^+-poly PMOS gate is boron penetration across the gate oxide [27]. Boron diffuses extremely fast in polysilicon. For each successive generation of CMOS, the gate oxide thickness must be scaled down in order to maintain adequate short channel performance. As a result, it is very easy for boron to diffuse across the oxide and into the channel. This results in a threshold voltage shift, poorer gate oxide quality, and higher leakage currents.

To minimize boron penetration, the thermal cycles after the poly implants need to be reduced. This can be accomplished using rapid thermal annealing (RTA). The ramp up, annealing, and cool down phases for RTA are much shorter than conventional furnaces. As a result, less diffusion of dopants will occur while still providing sufficient energy for activation.

In addition to lowering the thermal budget, alternative gate dielectrics can be used to minimize boron penetration. High κ materials, such as nitrous oxide, can prevent boron penetration better than standard SiO_2. By simply changing the ambient gas to N_2O, conventional furnaces can be used to grow the nitrous oxide gate. The oxidation rate is low and self-limiting, which can result in thin and uniform gate oxides. However, reviews of the performance and reliability of nitrous oxide are mixed [28]-[31], and very little work has been done in studying its performance at low temperature.

As an alternative to changing the gate material, nitrogen can be implanted into the polysilicon gate to reduce boron penetration. The presence of nitrogen in polysilicon reduces the diffusivity of boron [32] without affecting the gate oxide quality [33].

Again, little is known about the performance or reliability of such a process change at low temperature.

Because of the limitations in simulating both boron diffusion and the growth rate of alternative gate materials, much of the work to minimize boron penetration has to be done experimentally. A first order estimation on the thermal cycle needed to fully activate dopants while minimizing boron penetration can be obtained via SUPREM-III, but needs to be confirmed by fabricating test capacitors. Similarly, the steps needed to grow a high quality nitrous oxide gate can be determined by process development and test capacitor fabrication.

4.7 Conclusion

Several changes to a conventional CMOS process flow have been proposed to achieve higher performance at low operating temperatures. The modifications are minor and can be easily implemented without introducing new processing or design complications. By accounting for the increase in threshold voltage, an additional 40% improvement in delay can be achieved at 78 K for a 3 volt power supply. Further reduction of the threshold voltage can be done without an increase in off-state power dissipation, resulting in a 2-3x improvement in delay from 300 K to 78 K.

Because the mobility at low temperature is limited by ionized impurity scattering, retrograded or halo implants can be added to the process flow to further enhance low temperature performance without degrading short channel effects. For operating temperatures below 150 K, the LDD should be replaced by a more heavily doped source/drain extension in order to continue to have significant improvement with

decreased operating temperature. Switching to dual-polarity gates will eliminate PMOS channel freezeout and allow for better use of its retrograded profile.

Chapter 5

Performance of Optimized Low Temperature CMOS

5.1 Introduction

Up to this point, the study and redesign of CMOS for low temperature applications has focused on using TCAD design tools. After identifying the challenges of operating CMOS at low temperature, several minor process changes have been shown (through simulations) to further improve performance. But in order to truly demonstrate the merits of low temperature CMOS, actual devices and circuits need to be fabricated and shown to exhibit far better performance at low temperature than at room temperature. In this chapter, the measurement results of NMOS and PMOS transistors and CMOS ring oscillators will be presented. The development of the test chip and process flow used to manufacture these components will be described in detail. Process splits in the threshold implant, gate oxide formation, extension implant, halo implant, and dopant activation will be summarized. The

impact of the process modifications listed in Chapter 4 will be reviewed. Finally, a breakdown of the expected improvements for deep submicron CMOS will be given for an operating temperature of 78 K.

5.2 Test Chip Design

In order to evaluate a proposed process flow for fabricating CMOS devices and circuits, a test chip has been laid out using Cadence Opus [34]. One quarter of the chip is composed of process test structures. Another 50% is made up of discrete transistors with varying channel widths and lengths. The remaining 25% of the test chip contains simple ring oscillators and circuits. The test chip is robust enough to be used for a bulk CMOS or a fully-depleted SOI-CMOS process (no body contact). Figure 5.1 shows the test chip in its entirety. This pattern is repeated over an entire 4" silicon wafer using an Ultratech stepper. Table 5.1 lists the masks levels (11) that were generated for this chip. Although the chip is designed to handle dual-polarity gates, single-polarity gates can still be realized by POCL or BBR doping the gates after poly deposition, and slightly reducing the source/drain implant dose.

5.2.1 Process Test Structures

Figure 5.2 shows the section of the chip with the test structures. The upper portion consists of large area transistors, which can be used for split-CV measurements or mobility extraction, and two terminal capacitors, which can be used to evaluate gate oxide quality (CV, IV, and TDDB). The field transistors on the right are used to confirm adequate device isolation. The SPICE capacitors on the lower right are used to obtain the sidewall and junction capacitances for HSPICE simulation. The

Figure 5.1: Test chip used to fabricate devices and circuits for low temperature operation. Test structures (lower left), discrete devices (top), and circuits (lower right) are available.

Layer Name	Type	Aligns To
N-Well	Dark Field	Wafer (Mechanical)
Active	Clear Field	N-Well
Field	Clear Field	Active
E-Beam	Dark Field	Active
PMOS V_{th}	Dark Field	Active
NMOS V_{th}	Dark Field	Active
Poly	Clear Field	Active
N-Diffusion	Dark Field	Poly
P-Diffusion	Dark Field	Poly
Contact	Dark Field	Poly
Metal	Dark Field	Contact

Table 5.1: Masks layers used in the low temperature CMOS test chip.

Figure 5.2: Process test structure portion of the low temperature CMOS test chip. A wide variety of different structures are available to fully characterize a proposed process flow.

latchup structures along the bottom consist of a p^+ contact in p-sub, an n^+ contact in p-sub. a p^+ contact in n-well, and an n^+ contact in n-well, from left to right. They are used to determine or verify the design rules for latchup prevention.

The process capacitors in the center are used to evaluate the field and/or CVD oxide quality. Contact hole quality between the metal and any other conducting layer underneath can be tested electrically using the contact resistors on the lower left. The edgeless transistors on the left can be used to study device behavior if poor field isolation occurs. Finally, the cross bridges in the center consist of a Van der Pauw structure (used to measure sheet resistance) and a Kelvin resistor (used to measure effective linewidth). The Van der Pauw structure is especially important in determining the effectiveness of silicidation.

Figure 5.3: n^+-poly NMOS transistor array and SEM bar. The primary transistor width is 10 μm. Each transistor has its own drain, gate, source, and substrate pads (starting from the upper left and counting clockwise).

5.2.2 Transistors

The upper half of the test chip is made up of discrete transistors. The devices are organized into two 5 x 5 arrays: one n^+-poly NMOS and one p^+-poly PMOS. Vertical SEM bars with varying gate lengths are placed next to the arrays on opposite ends of the die. Figure 5.3 shows one of the transistor arrays in more detail. Channel widths vary from 3 μm to 20 μm, and channel lengths vary from 10 μm down to 0.25 μm. Each transistor has its own gate, source, drain, and substrate pad, so that reliability testing can be performed without affecting the other devices. Metal is routed from the pads of 17 out of the 25 devices to the outside edges of the array so that wire bonding on a 68 pin (17 x 4) package can be easily performed.

If an NMOS-only or PMOS-only transistor run is desired, both transistor arrays can still be utilized. Blanket implants can be performed for the threshold, extension, and source/drain implants. This simplifies the processing because masking is no longer required before any of these implants. However, the N and P masks can still be used so that process splits of any of these three implants can be performed on the same wafer. Note that if both transistor arrays are to made into NMOS-only or PMOS-only, the backside of the wafer must be used as the substrate contact.

5.2.3 Circuits

The circuit blocks in the test chip are shown in Figure 5.4. A W_p to W_n ratio of 10 μm / 5 μm is used throughout. The majority of the circuit elements are ring oscillators, with channel lengths ranging from 2 μm down to 0.25 μm. The basic element of each ring oscillator is either an inverter, a 2-input NOR gate, or a 3-input NAND gate. For the NOR and NAND gates, the output of each stage is fed into every input of the next stage. Individual circuit elements are located along the bottom portion of the circuit block. To drive the output pad, the ring oscillator signal is passed through a four-stage buffer. At each stage of the buffer, the gate widths are either tripled or quadrupled so that a 30 pF output load can be driven. The buffers are replicated in the lower right of the circuit block for debugging purposes. Separate V_{dd} lines are used for the buffer and ring oscillator so that the actual ring oscillator power dissipation can be obtained.

The remaining circuit blocks are all drawn with 1 μm channel lengths, using "standard" design rules, and a single level of metal. Along the top of the circuit

Figure 5.4: Circuit blocks designed in the low temperature CMOS test chip.

module is a D-register, a one-bit adder, and a dynamic NOR delay line. Alternating dies also have a 7-bit counter.

5.3 Process Flow

Table 5.2 summarizes the fabrication steps used to make the CMOS devices and circuits. The process flow was based on Stanford's BiCMOS process and several recent SOI runs fabricated at Stanford University's Center for Integrated Systems. The starting material for each process run was 100 nm (4") silicon wafers, with a resistivity of roughly 10 Ω-cm. For CMOS runs, an n-well was first defined by implant and drive-in. In order to align subsequent levels to the n-well, an initial oxidation and patterned etch were necessary before the implant, followed by a reoxidation before the drive-in.

Step	Description	Details
1	Starting material	CMOS/NMOS: p-type, $< 100 >$, 10-20 Ω-cm PMOS: n-type, $< 100 >$, 5-10 Ω-cm
2	N-well formation (CMOS only)	Phosphorous implant: $2.5x10^{12}cm^2$ @ 100 keV Drive-in: 16 hours at 1150°C
3	Pad oxidation	Dry O_2: 950°C, 25 nm
4	Nitride deposition	LPCVD: NH_3 + DCS, 785°C, 100 nm
5	Active area definition	Lithography and RIE: $CF_3Br + SF_6$
6	Field implant	Boron: $1x10^{13}cm^2$ @ 100 keV
7	E-beam alignment mark formation	Lithography and silicon RIE etch $C_2C\ F_5 + SF_6$: 2 μm
8	LOCOS isolation	Wet O_2: 1000^oC, 750 nm
9	Nitride strip	Wet etch: H_3PO_4
10	NMOS and PMOS threshold implant	Lithography and ion implantation (variable doses and energies)
11	Gate oxidation	Standard: Dry O_2, 800^oC + N_2 anneal Nitrous oxide: diluted N_2O, 850^oC + N_2, 1050^oC, 15 min.
12	Polysilicon deposition	LPCVD: $SiH_4 + H_2$, 620^oC, 175 nm
13	Poly gate processing	N_2 implant or $POCL_3$ doping (both optional)
14	Poly lithography	E-beam lithography
15	Poly etch	RIE: HBr + $C\ _2$ OR NF_3
16	Reoxidation	Dry O_2: 850^oC, 40 min.
17	NMOS and PMOS extension/halo implant	Lithography and ion implantation (variable doses and energies)
18	Spacer deposition	LPCVD: $SiH_4 + O_2$, 125 nm
19	Spacer etch	RIE: $CHF_3 + O_2$
20	Source/drain/(gate) implant	NMOS: Arsenic $1x10^{16}cm^2$ @ 40 keV PMOS: Boron $2x10^{15}cm^2$ @ 12 keV
21	Source/drain/(gate) anneal	Furnace or RTA (variable temperature and time)
22	Silicidation	Ti deposition (25 nm) + RTA (2-step)
23	Contact hole deposition	LPCVD: $SiH_4 + O_2$, 500 nm
24	Contact hole etch	Lithography and RIE: $CHF_3 + O_2$
25	Metalization	Al/1% Si deposition (800 nm)
26	Metal Etch	Lithography and RIE: $CHCL_3 + N_2 + SiC\ _4$
27	Backside oxide/poly removal	BOE etch and RIE: $C_2C\ F_5 + SF_6$
28	Forming gas anneal	H_2: 400^oC, 60 min.

Table 5.2: Process flow for low temperature CMOS devices and circuits.

Standard local oxidation of silicon (LOCOS) and a field implant were performed to isolate the devices. In order to use e-beam lithography for the poly level, 2 μm trenches in the shape of a cross were defined in the corners of each die using standard lithography and deep RIE silicon etching. After implanting the channels to adjust the threshold voltage or create a retrograded profile, a 7 nm or 5 nm gate oxide was grown followed by LPCVD polysilicon deposition. The nitrogen annealing conditions after gate oxidation were varied to examine tradeoffs in reliability and retrograded channel profiling. Wafers needing nitrous oxide gates were grown in diluted nitrous oxide (instead of dry O_2) and annealed in N_2 at 1000^oC for 15 minutes.

A Hitachi HL-700F electron beam lithography machine was used to define the polysilicon gates. Since the system accepted masks electronically, modification of the channel lengths could be easily made without having to remanufacture a new photomask. Features down to 0.15 μm were generated using Shipley SAL-601 e-beam resist. An Ultratech Model 1000 Stepper was used for masking all other levels because of faster throughput and the fact that only the poly level had feature sizes smaller than 1 μm (the minimum resolution of the Ultratech). Using the trench crosses formed earlier, good alignment to the previous layers was achieved. To align subsequent mask layers to the poly level, Ultratech alignment crosses were drawn in the scribe area of the wafer. In addition, verniers were drawn in the field regions of the test chip for alignment verification.

Poly etching was performed using either NF_3 or HBr + $C\ _2$. The latter etch chemistry has higher selectivity between polysilicon and oxide. After gate definition, a short reoxidation was performed to clean up any damage from the etch. The extension and (optional) halo regions were then implanted, followed by an LTO

spacer deposition and etch. Next, implantation the source and drain regions was done. These implants also doped the poly gate (if not POCL doped already) and helped to form substrate/n-well contacts. A furnace or RTA anneal to activate the dopants completed the front-end processing.

To lower the sheet resistance of the polysilicon gates and the source/drain, a standard, titanium silicidation process was used. It is assumed that nearly all of the sputtered titanium will react to form about 50 nm of $TiSi_2$. A layer of undensified LTO oxide was then deposited and patterned to form contact holes. After metalization using 800 nm of Al/Si, backside removal of the residual oxide and polysilicon was performed, followed by a forming gas anneal.

Figure 5.5 shows the cross section of the resulting NMOS and PMOS devices with expected dimensions. In order to perform low temperature measurements, the finished wafers were diced, attached to ceramic packages using silver epoxy, and wire bonded using gold or aluminum wire. Measurements were then performed using a computer controlled HP 4142B connected to a temperature controlled Lakeshore Cryotronics dewar.

5.4 Process Splits

The process steps which received different implants or thermal cycles are listed in Tables 5.3 and 5.4. A wide range of threshold voltages are created by varying the implant dose and energy. Performance tradeoffs (i.e. low threshold voltage versus high off-state currents) can thus be studied at a variety of different temperatures. Wafers which have "negative" threshold voltages at room temperature will have low

Figure 5.5: Cross section of fabricated NMOS and PMOS devices. Figure is not drawn to scale.

threshold voltages at low temperature and can be compared to low threshold devices at room temperatures.

The NMOS threshold implant splits include heavy ion implants (indium) which are used to create retrograded channel profiles. The thermal cycles (gate oxide anneal and source/drain anneal) are also varied to study the tradeoffs between retrograde channel formation and device reliability/dopant activation. Halo implants are employed to improve short channel behavior, especially for retrograded profiles since the channel doping near the Si-SiO_2 interface is low. A halo implant can also be used to keep the overall channel doping light, resulting in improved low temperature mobility.

Step	Details
NMOS Threshold Implant	Boron: $4x10^{12}cm^2$ @ 15 keV
	Boron: $4x10^{12}cm^2$ @ 15 keV + Boron: $5x10^{12}cm^2$ @ 100 keV
	Boron: $2x10^{12}cm^2$ @ 40 keV
	Boron: $4x10^{12}cm^2$ @ 40 keV
	Boron: $8x10^{12}cm^2$ @ 40 keV
	Boron: $4x10^{12}cm^2$ @ 40 keV + Indium: $5x10^{12}cm^2$ @ 100 keV
	Boron: $2x10^{13}cm^2$ @ 40 keV
	Boron: $4x10^{12}cm^2$ @ 80 keV
	Boron: $2x10^{13}cm^2$ @ 80 keV
	Indium: $1x10^{12}cm^2$ @ 150 keV
	Indium: $4x10^{12}cm^2$ @ 300 keV
	Indium: $6x10^{12}cm^2$ @ 300 keV
	Indium: $8x10^{12}cm^2$ @ 300 keV
PMOS Threshold Implant (p^+-poly gate)	Arsenic: $3x10^{12}cm^2$ @ 100 keV
	Arsenic: $3x10^{12}cm^2$ @ 140 keV
	Arsenic: $6x10^{12}cm^2$ @ 140 keV
	Arsenic: $3x10^{12}cm^2$ @ 200 keV
PMOS Threshold Implant (n^+-poly gate)	Boron: $3x10^{12}cm^2$ @ 15 keV + Phosphorous: $5x10^{12}cm^2$ @ 150 keV
	Boron: $3x10^{12}cm^2$ @ 15 keV + Phosphorous: $5x10^{12}cm^2$ @ 200 keV
Gate Oxide and Anneal	SiO_2: 7 nm; 1050^oC, 30 min.
	SiO_2: 7 nm; 950^oC, 30 min.
	SiO_2: 7 nm; no N_2 anneal
	SiO_2: 5 nm; 950^oC, 30 min.
	N_2O: 7 nm; 1000^oC, 30 min.
	N_2O: 5 nm; 1000^oC, 30 min.
NMOS Halo Implant	none
	BF_2: $5x10^{12}cm^2$ @ 120 keV (30° tilt)
	BF_2: $1x10^{13}cm^2$ @ 120 keV (30° tilt)
PMOS Halo Implant	none
	Phosphorous: $1x10^{13}cm^2$ @ 120 keV (30° tilt)

Table 5.3: Process splits for low temperature CMOS devices and circuits.

Step	Details
NMOS Extension Implant	Arsenic: $7x10^{13}cm^2$ @ 20 keV
	Arsenic: $2x10^{14}cm^2$ @ 20 keV
	Arsenic: $5x10^{14}cm^2$ @ 20 keV
PMOS Extension Implant	Boron: $5x10^{13}cm^2$ @ 10 keV
	Boron: $7x10^{13}cm^2$ @ 10 keV
	Boron: $2x10^{14}cm^2$ @ 10 keV
	Boron: $4x10^{14}cm^2$ @ 10 keV
Source/Drain Anneal	Furnace: 850^oC, 15 min.
	Furnace: 950^oC, 30 min.
	RTA: 950^oC, 2 min.
	RTA: 950^oC, 30 sec.

Table 5.4: Process splits for low temperature CMOS devices and circuits (continued).

To better understand LDD freezeout, a variety of different doses for the extension implant are used. The tradeoff between parasitic source/drain resistance and device reliability can be easily studied this way.

5.5 Low Temperature NMOS Performance

Linear I_{ds}-V_{gs} characteristics of a 10 μm x 0.25 μm NMOS transistor at different temperatures are shown in Figures 5.6 and 5.7. Unlike previous results (see Figure 3.6), freezeout does not limit the linear drain current at 78 K. This is because a higher implant dose was used in the extension region. Good short channel behavior was obtained all the way down to channel lengths of 0.25 μm (Figure 5.8).

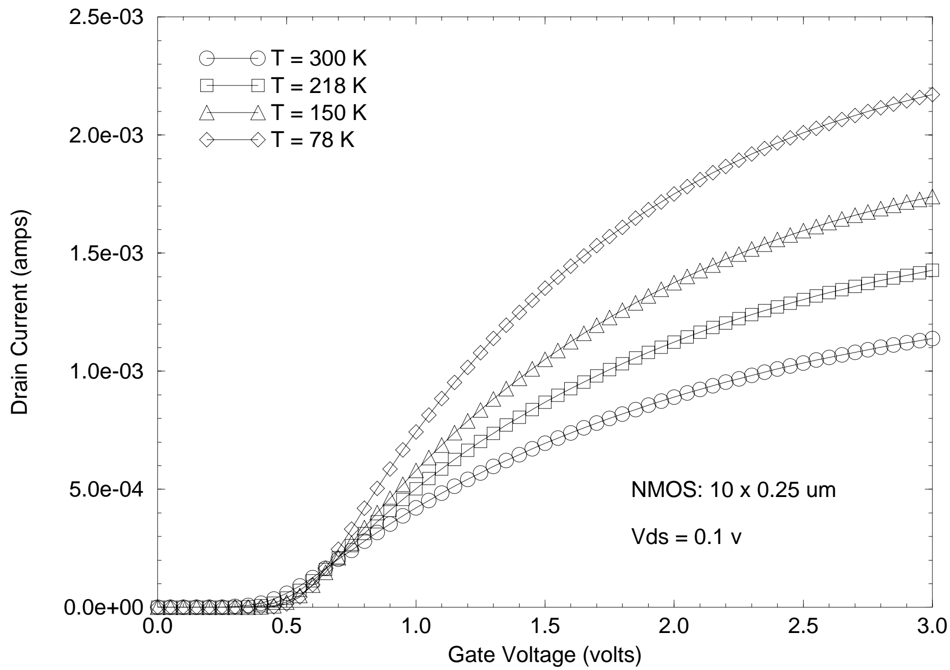

Figure 5.6: Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.25 μm NMOS transistor at different temperatures (linear scale). Because a higher implant dose is used in the LDD/extension region, freezeout does not limit the current at 78 K.

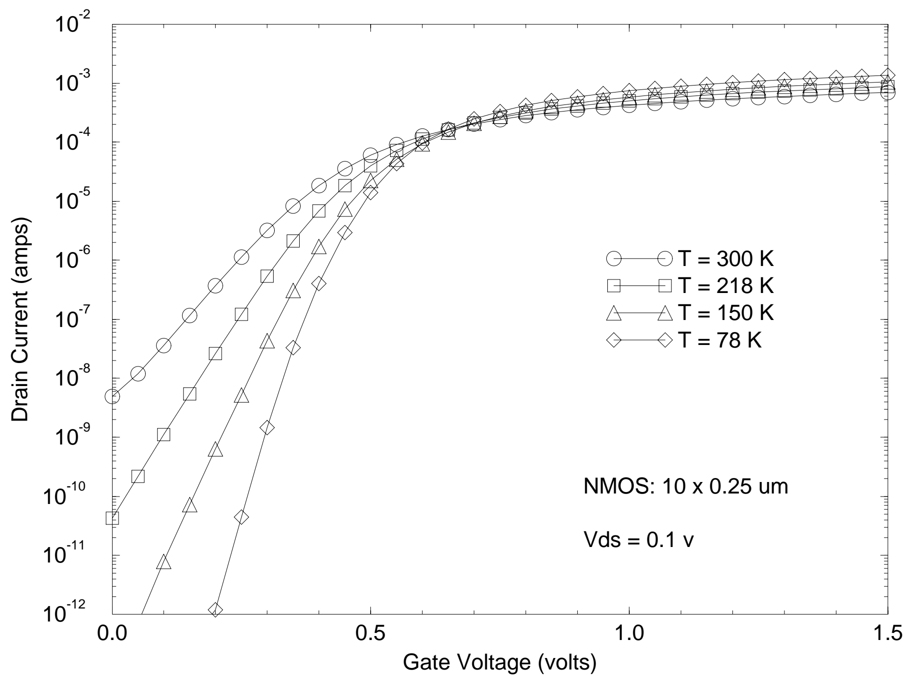

Figure 5.7: Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.25 μm NMOS transistor at different temperatures (log scale).

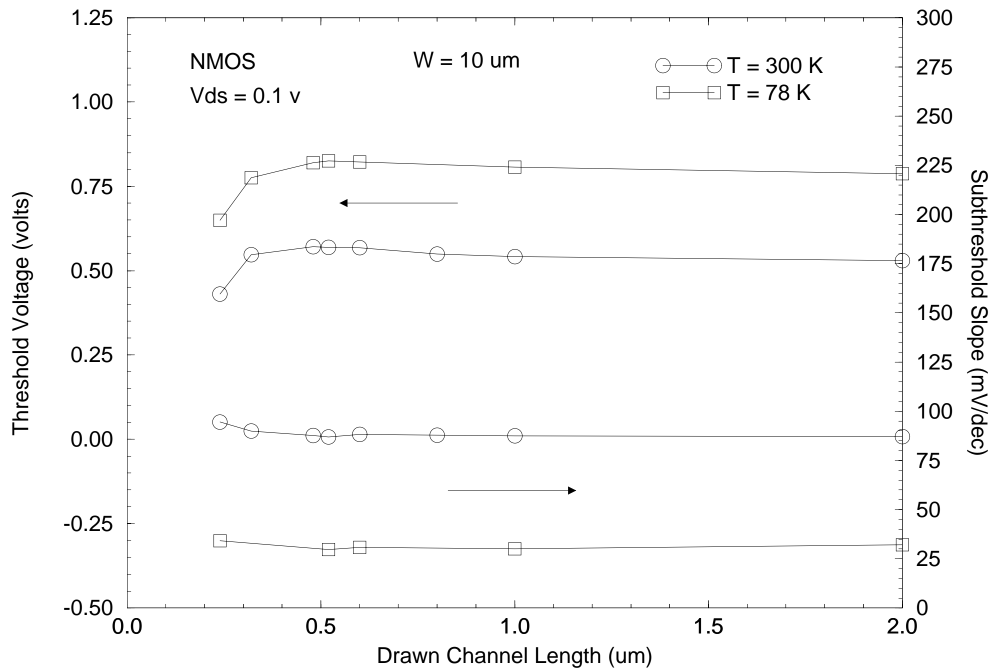

Figure 5.8: Measured threshold voltage and subthreshold slope of NMOS transistors at 300 K and 78 K. Threshold voltage is defined as the gate bias which corresponds to 10 nA x W/L of drain current. All measurements were done with $V_{ds} = 0.1$ v.

Figure 5.9: Saturation current for NMOS transistors versus drawn channel length at 300 K and 78 K. Measurements were made at a gate and drain bias of 3 v.

Figure 5.10: I_{ds}-V_{ds} characteristics of a 10 μm x 0.25 μm NMOS transistor at 300 K and 78 K.

Figure 5.9 illustrates the saturation current versus channel length at room and liquid nitrogen temperatures. Even at deep submicron dimensions, significant improvement can be seen (Figure 5.10). To first order, the gate delay of a single inverter stage is:

$$t_{delay} = C_{out} \frac{V_{dd}}{I_{dsat}} \tag{5.1}$$

where C_{out} is the output capacitance and V_{dd} is the power supply voltage. In this dissertation, the saturation current will be used as the primary figure of merit for measuring CMOS performance.

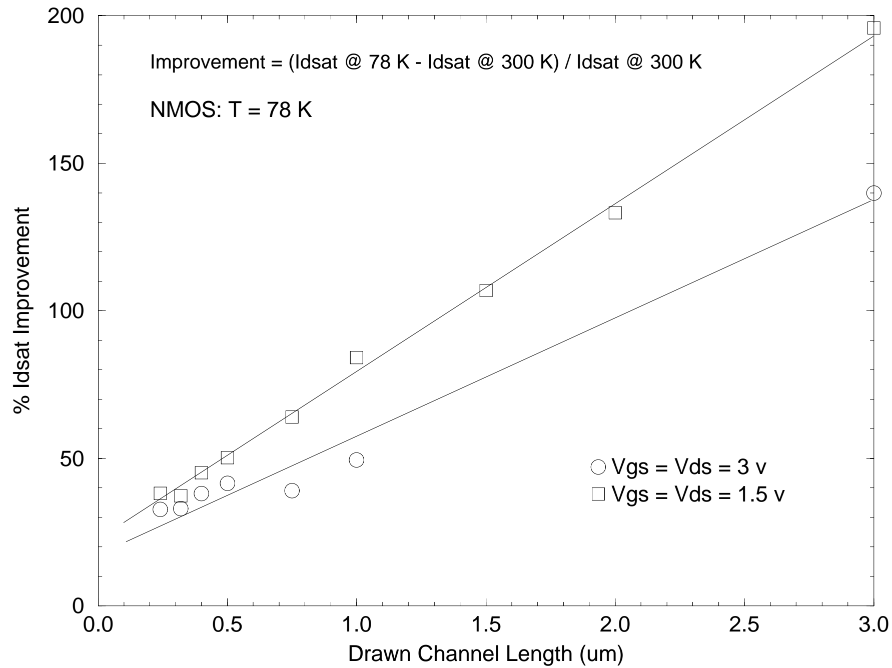

Figure 5.11: Improvement in saturation current for NMOS transistors (normalized to 300 K). Better improvement occurs if lower power supplies are used. However, higher power supply voltages result in higher current drive.

5.5.1 Low Power Performance

The amount of improvement in saturation current at 78 K, normalized to 300 K, is shown in Figure 5.11. The increase is linear and continuous across all channel lengths from high to low temperatures (Figure 5.12). At short channel lengths, a 40% improvement in current drive can be obtained when the temperature is lowered from 300 K to 78 K. This amount is similar to the results obtained from cooling down the 1 μm CMOS process in Chapter 2.

Lowering the power supply voltage (V_{ds}) from 3 volts to 1.5 v results in better improvement in saturation current because of reduced mobility degradation from carrier scattering. Since a greater increase in saturation current can be achieved at lower power supply voltages, better low temperature performance can be expected

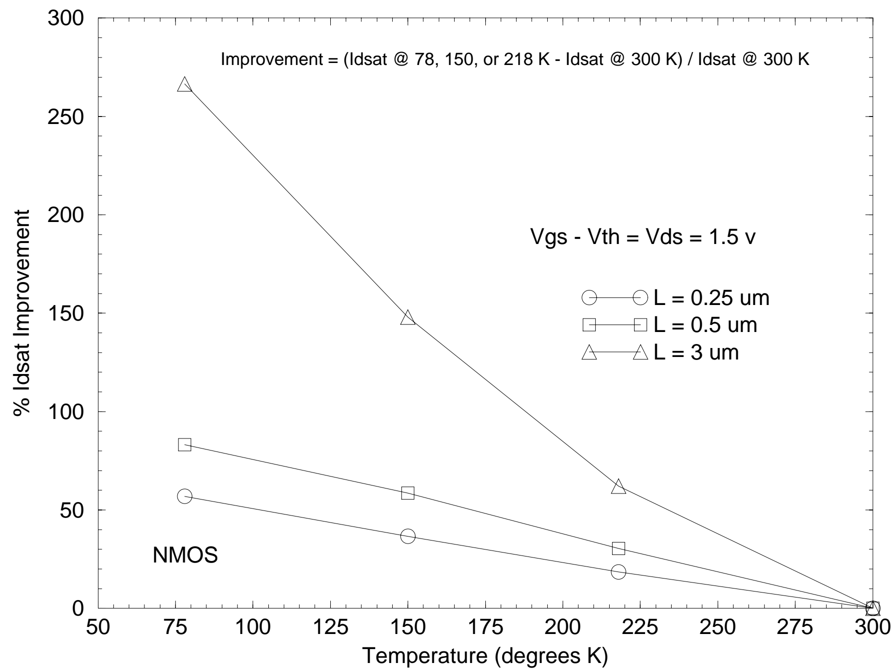

Figure 5.12: Improvement in saturation current at different temperatures for NMOS transistors of varying channel lengths. Results are normalized to 300 K at a drain and gate above threshold bias of 1.5 v. A 50% improvement in performance can be realized even at 0.25 μm.

in low power systems. In addition, improved reliability can also be expected due to the lower electric fields. However, the overall system power consumption will rise dramatically due to the need for a cooling system. System portability and battery operation are lost in low power systems if low temperature operation is used.

5.5.2 Retrograde Profile Performance

By reducing the thermal cycles and using indium (instead of boron) to implant the channel region, better low temperature performance is expected because the profile will be more retrograded. Figures 5.13 and 5.14 show the I_{ds}-V_{gs} and I_{ds}-V_{ds} characteristics of NMOS transistors with indium implanted channels at different temperatures. The IV curves and short channel behavior (Figure 5.15) are comparable to the results presented earlier, which used boron implants to shift the threshold voltage.

Figure 5.16 illustrates the improvement in saturation current at low temperature between standard (boron) and retrograded (indium) profiles. To eliminate the difference in threshold voltage on the results, the saturation current is measured using a gate bias that is three volts above V_{th}. Using a retrograded profile, an additional 35% improvement in performance can be achieved.

The improvement in saturation current for different indium profiles is shown in Figure 5.17. As the dose is increased, the channel doping near the Si-SiO_2 interface also increases. Therefore, mobility enhancement at low temperature decreases due to increased ionized impurity scattering. Reducing the implant dose can further lighten the channel doping, but can also worsen short channel behavior (Figure 5.18).

Figure 5.13: Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.52 μm NMOS transistor with indium implanted channel at different temperatures (log scale).

SUPREM-III simulation results (Figure 4.11) confirm that the $4x10^{12}cm^2$ @ 300 keV indium implant results in the best retrograded profile.

5.5.3 Halo Implant Performance

To achieve better short channel behavior for a more lightly doped retrograded profile, a halo implant can be used. Figure 5.19 shows the saturation current improvement when a halo implant is added. Although the improvement in current is reduced by 25% at 3 v, at lower power supply voltages, the difference is negligible (Figure 5.20). Over 2x improvement in current drive at 0.5 μm channel lengths is achievable if the saturation currents are normalized to measurements at 70^oC (343 K) (Figure 5.21). Most electronic components designed for room temperature operation are specified

Figure 5.14: Measured I_{ds}-V_{ds} characteristics of a 10 μm x 0.24 μm NMOS transistor with indium implanted channel at 300 K and 78 K. At low gate, high drain biases, the short-channel output characteristics exhibited a "kink" behavior which was attributed to freezeout in the substrate. Using a more highly doped substrate (or a deep substrate implant) should remove this anomalous behavior.

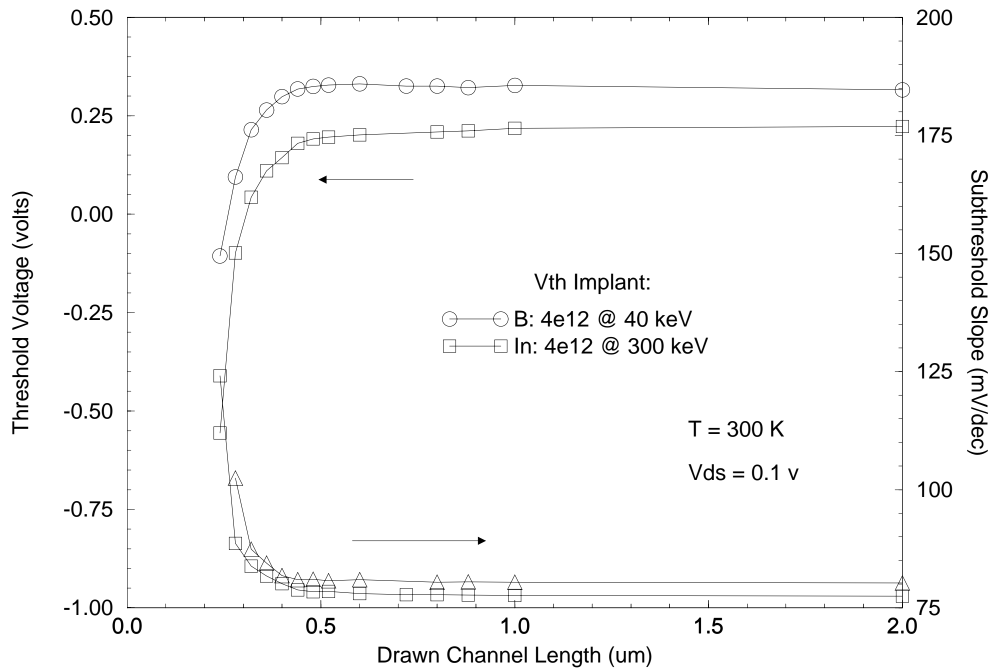

Figure 5.15: Measured threshold voltage and subthreshold slope of NMOS transistors with different threshold implants at 300 K. The indium implanted wafers exhibit nearly identical short channel behavior, after correcting for the difference in threshold voltage.

Figure 5.16: Improvement in saturation current for NMOS transistors with different threshold implants at 78 K. Results are normalized to 300 K at a drain and gate bias above threshold of 3 v.

Figure 5.17: Improvement in saturation current for NMOS transistors with different retrograded profiles at 78 K. Results are normalized to 300 K at a drain and gate bias above threshold of 3 v. The results for a standard (boron) profile are shown for reference.

Figure 5.18: Threshold voltage rolloff for NMOS transistors with different threshold implants at 78 K. The light dose retrograded implant exhibits the worst short channel behavior.

Figure 5.19: Improvement in saturation current for NMOS transistors with different halo implants at 78 K. Results are normalized to 300 K at a drain and gate bias above threshold of 3 v.

to run over a large temperature range (0^oC and 70^oC) and typically tend to run hot, even when cooling fans are used.

As expected, the threshold voltage rolloff is less with a halo implant (Figure 5.22, although additional refinement of the halo implant is necessary to reduce the large reverse short channel effect. However, using this particular halo implant with the $1x10^{12}cm^2@150keV$ indium implant may result in less reverse short channel behavior as well as better improvement in current drive.

5.5.4 Threshold Voltage Considerations

Figure 5.23 illustrates the improvement in saturation current at a fixed gate bias above the threshold voltage. At short channel lengths, these results are 15% higher

Figure 5.20: Improvement in saturation current for NMOS transistors with different halo implants at 78 K. Results are normalized to 300 K at a drain and gate bias above threshold of 1.5 v. The difference between halo and non-halo implanted wafers are minimal at this lower voltage.

Figure 5.21: Improvement in saturation current for NMOS transistors with different halo implants at 78 K. Results are normalized to 343 K at drain and gate biases above threshold of 3 v and 1.5 v. Comparing the low temperature data to 343 K is more accurate since typical room temperature electronics run hot.

Figure 5.22: Threshold voltage and subthreshold slope versus channel length for NMOS transistors with different halo implants at 300 K. The threshold voltage for the halo implanted wafer increases at short channel lengths because the effective channel doping underneath the gate is higher due to the halo pocket.

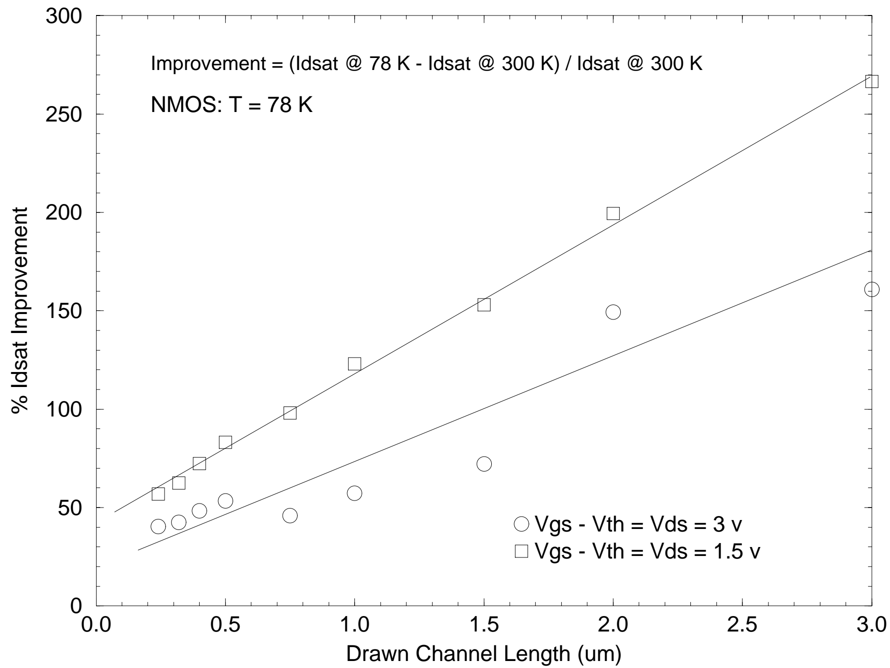

Figure 5.23: Improvement in saturation current for NMOS transistors at gate biases above the threshold voltage. The data has been normalized to measurement results at 300 K. An additional 15% improvement in performance can be seen after comparing to Figure 5.11

than earlier results (Figure 5.11) because the shift in threshold voltage from room to low temperature has been eliminated. In order to determine the amount of additional improvement when the threshold voltage is properly chosen, the saturation current from a low threshold low temperature device should be compared to the saturation current from a low threshold room temperature device. Figure 5.24 shows the I_{ds}-V_{gs} curves from two NMOS transistors. Only the implant dose and energy in the channel are different. Both devices are optimized for operation at their respective temperatures because of their low off-state current (less than 1 nA) and minimal off-zones. At 78 K, a lower threshold voltage should be used because of the improved subthreshold slope. This allows for higher current drive at a given drain and gate bias.

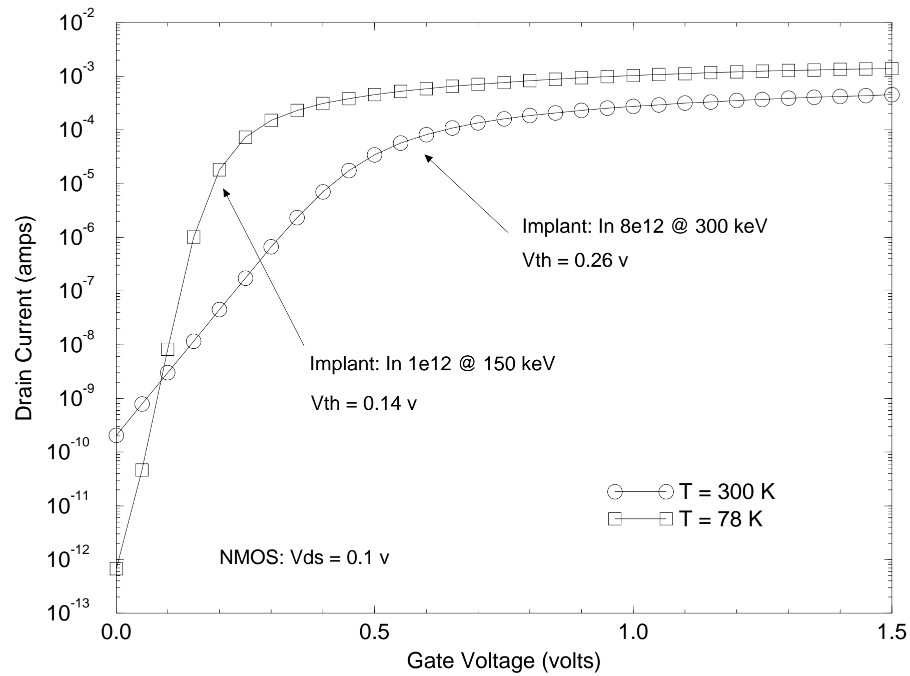

Figure 5.24: I_{ds}-V_{gs} characteristics (log scale) of two 10 μm x 0.44 μm NMOS transistors at 300 K and 78 K. Each device is optimized for operation at its respective temperature.

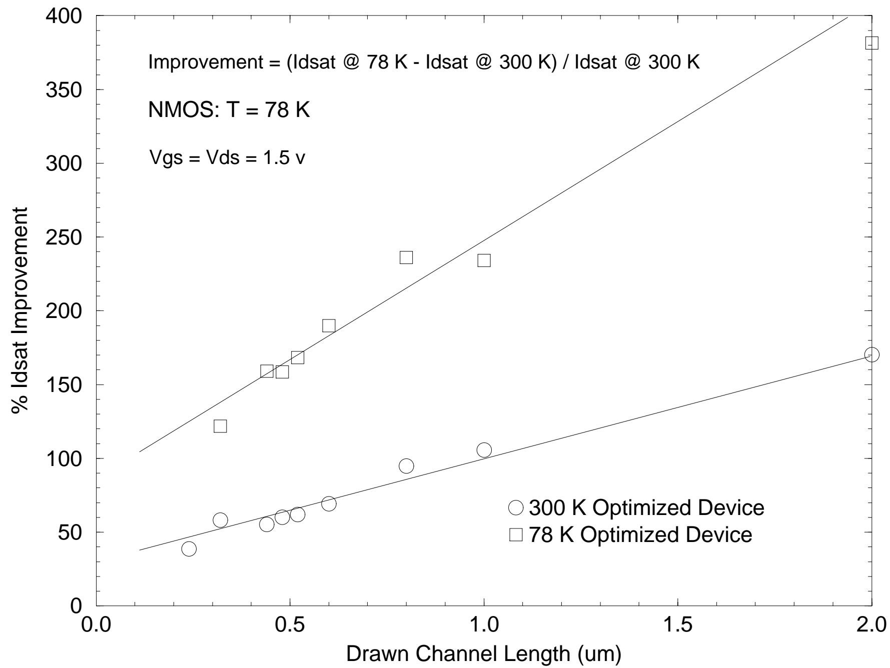

Figure 5.25: Improvement in saturation current at 78 K for NMOS transistors optimized for 300 K and 78 K operation. All data here has been normalized to the 300 K optimized devices. With proper adjustment of the threshold voltage, an additional 60% improvement in performance can be achieved.

Figure 5.25 shows the improvement in saturation current at 78 K for the optimized 300 K and 78 K NMOS devices. In *both* cases, the current is normalized to the optimized 300 K device. As expected, the optimized 300 K device only improves by only about 50% at 78 K. This is because the threshold voltage increases to 0.50 v at 78 K (Figure 5.26), creating too large of an "off-zone", which limits the amount of current that can be generated for a drain and gate bias of 1.5 v. With proper shifting of the threshold voltage, a 2x improvement in saturation current can be easily achieved at short channel lengths.

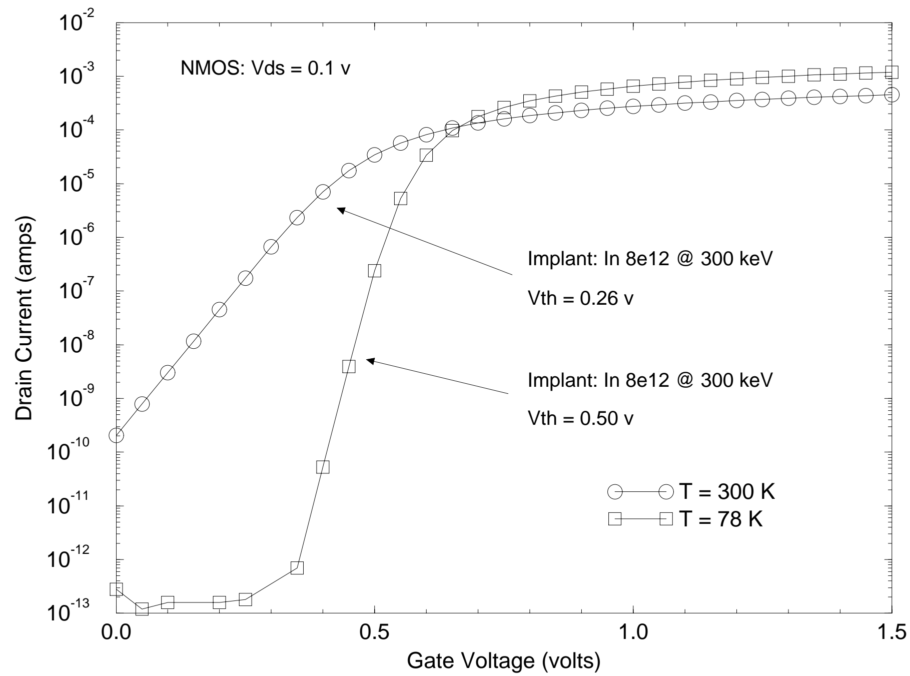

Figure 5.26: I_{ds}-V_{gs} characteristics (log scale) of a 10 μm x 0.44 μm NMOS transistor (optimized for 300 K operation) at 300 K and 78 K. At 78 K, the threshold voltage shifts too high, limiting the improvement in current drive.

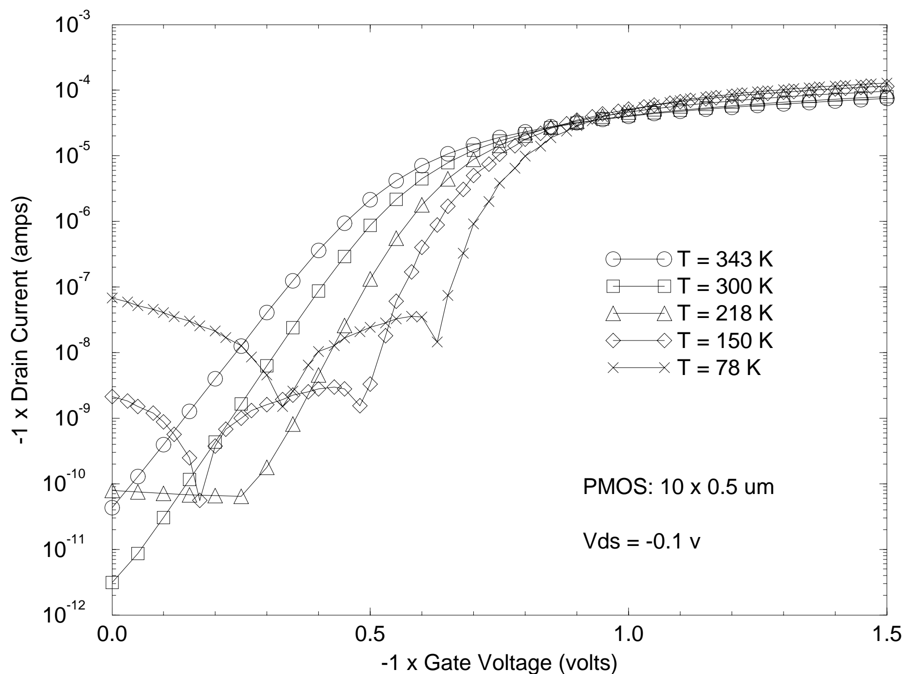

Figure 5.27: I_{ds}-V_{gs} characteristics (log scale) of a 10 μm x 0.5 μm PMOS transistor at different temperatures. The higher leakage currents at low temperature may be the result of poor gate oxide quality due to boron penetration.

5.6 Low Temperature PMOS Performance

PMOS device performance and current improvements at low temperature are comparable to the NMOS results. Figures 5.27 and 5.28 show the I_{ds}-V_{gs} and I_{ds}-V_{ds} characteristics of a 10 μm x 0.5 μm PMOS transistor at different temperatures. The higher leakage currents at low temperature may limit the effectiveness of operating CMOS at low temperatures. In addition, the amount of reduction that can be made in the threshold voltage for an optimal low temperature PMOS device will also be limited.

Due to the higher diffusivity of boron, good PMOS I-V characteristics were obtained only down to channel lengths of 0.4 μm. Devices below this size showed punchthrough between the source and drain. Further scaling is necessary to achieve

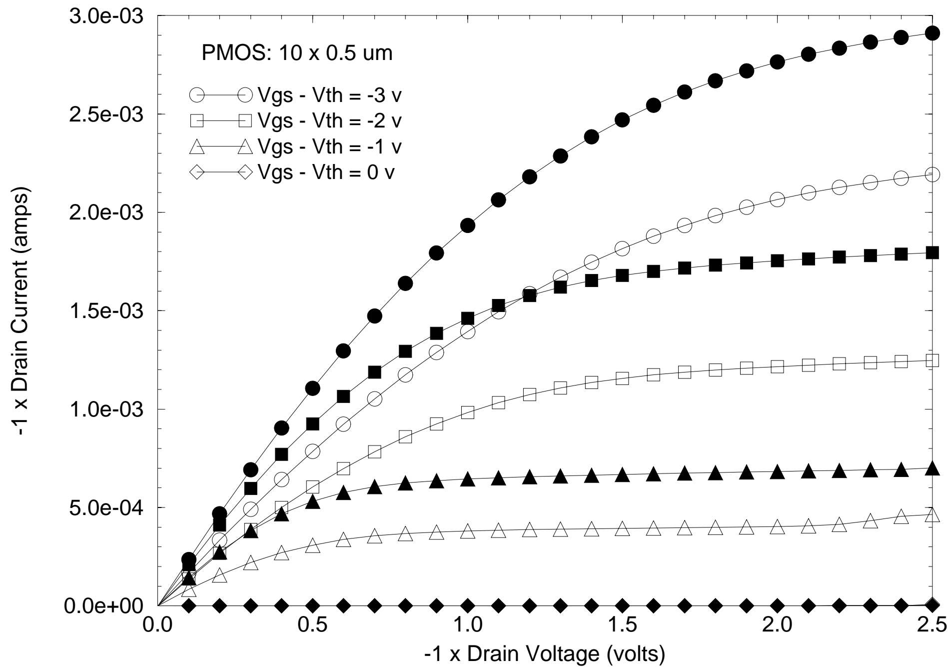

Figure 5.28: I_{ds}-V_{ds} characteristics of a 10 μm x 0.5 μm PMOS transistor at different temperatures.

good PMOS behavior at 0.25 μm. This can be achieved through the following process modifications:

- Reduction of the source/drain anneal. The limiting factor here will be sufficient activation of the NMOS source/drain dopants.
- Use of a lower energy implant for the boron extension and source/drain implants. High dose, low energy boron implants typically cost more and reduce wafer throughput.
- Growth of a thinner gate oxide. Gate oxide scaling is necessary to maintain good short channel behavior at shorter channel lengths. However, boron penetration worsens for thinner oxides.

Figure 5.29: Improvement in saturation current versus channel length for PMOS transistors at different temperatures. Results are normalized to 300 K at a drain and gate above threshold bias of 1.5 v.

- Counterdoping through a halo implant. Implanting phosphorous prior to the extension implant should minimize the lateral diffusion of boron, resulting in better short channel performance.

- Increasing the dose of the retrograde channel implant. A higher doping concentration in and below the channel should reduce punchthrough effects, although the improvement in saturation current at low temperatures may also decrease due to increased impurity scattering near the Si-SiO_2 interface.

PMOS saturation current improvement at different temperatures is illustrated in Figure 5.29. A 50% improvement in current drive can be obtained at 0.5 μm channel lengths for a supply voltage of 1.5 v, assuming the threshold voltage has been properly chosen.

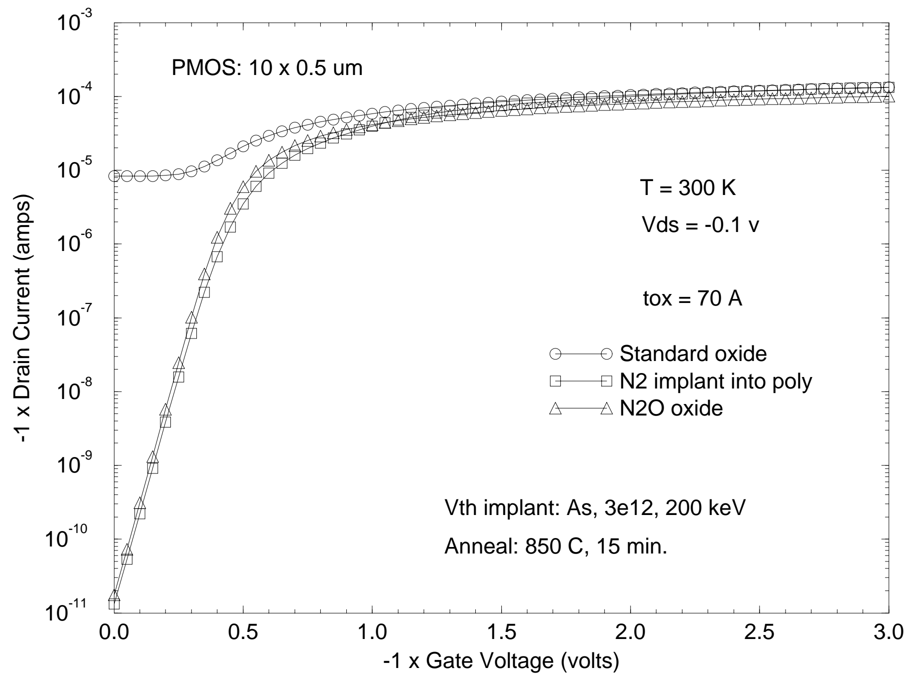

Figure 5.30: I_{ds}-V_{gs} characteristics (log scale) of 10 μm x 0.5 μm PMOS transistors at 300 K using different poly-gate structures. Even for an 850^oC furnace anneal (which is not enough to activate the NMOS dopants), boron penetration is evident.

5.6.1 Boron Penetration Effects

With the need to use a thin gate oxide for sufficient gate control at short channel lengths, rapid thermal annealing must be employed to activate the dopants. Figure 5.30 shows the I_{ds}-V_{gs} characteristics of three different 10 μm x 0.5 μm PMOS transistors. The dopants were activated using a furnace anneal of 850^oC for 15 minutes. Even with this low thermal budget and relatively thick oxide (70 Å), boron penetration occurs. The boron that enters the channel region shorts with the boron present in the source and drain, resulting in high off-state leakage currents. In addition, the boron neutralizes with the arsenic used for threshold adjustment, leading to a lower threshold voltage. By implanting nitrogen into the polysilicon or using

a nitrous oxide gate, boron penetration is suppressed, resulting in low off-state currents. Alternatively, rapid thermal annealing can be used with a standard SiO_2 gate, which was the case in the PMOS results shown earlier in this section. Nevertheless, some boron penetration may have occurred, resulting in the high off-state currents observed at low temperature.

The increase in saturation current at different temperatures for the nitrogen implanted poly wafer is shown in Figure 5.31. An N_{14}^{+} implant of $4x10^{15}cm^2$ at 20 keV was used. The improvement at low temperature is comparable to standard gate oxide results and is actually 20% better for 0.5 μm channel lengths at 78 K. N_2O oxide saturation currents are also comparable to standard oxide results (Figure 5.32) with the additional advantage of low leakage currents (Figure 5.33).

5.6.2 Freezeout Design Issues

The parasitic source/drain resistances associated with different extension implant doses are summarized in Table 5.5. The method described in [35] is used to extract R_{sd} from the linear I_{ds}-V_{gs} curves at different channel lengths. As expected, higher extension doses lead to lower source/drain resistances. Freezeout effects occur at 78 K for NMOS doses below $7x10^{13}cm^2$ and PMOS doses below $4x10^{14}cm^2$. A higher dose is needed to minimize PMOS freezeout because boron diffuses faster than arsenic. This also explains why R_{sd} is higher at the same doses for PMOS versus NMOS transistors.

As seen earlier with the 1 μm CMOS devices (Figure 3.6), the linear drain current is actually reduced at lower temperatures under freezeout conditions (Figure 5.34). With no freezeout, the linear drain current between 150 K and 78 K increases by

Figure 5.31: Improvement in saturation current versus channel length for PMOS transistors with N_2 implanted poly at different temperatures. Results are normalized to 300 K at a drain and gate above threshold bias of 1.5 v. Standard oxide results are shown for comparison.

	Implant Dose cm^2	R_{sd} (ohms)			
		T = 300 K	T = 218 K	T = 150 K	T = 78 K
NMOS (20 keV)	$5x10^{13}$	50	55	60	90
	$7x10^{13}$	60	55	45	40
	$4x10^{14}$	55	50	45	40
	$5x10^{14}$	35	15	25	20
PMOS (10 keV)	$5x10^{13}$	145	145	140	200
	$7x10^{13}$	130	95	100	170
	$4x10^{14}$	120	125	120	100

Table 5.5: Parasitic source/drain resistance at different temperatures for different NMOS and PMOS extension implant doses.

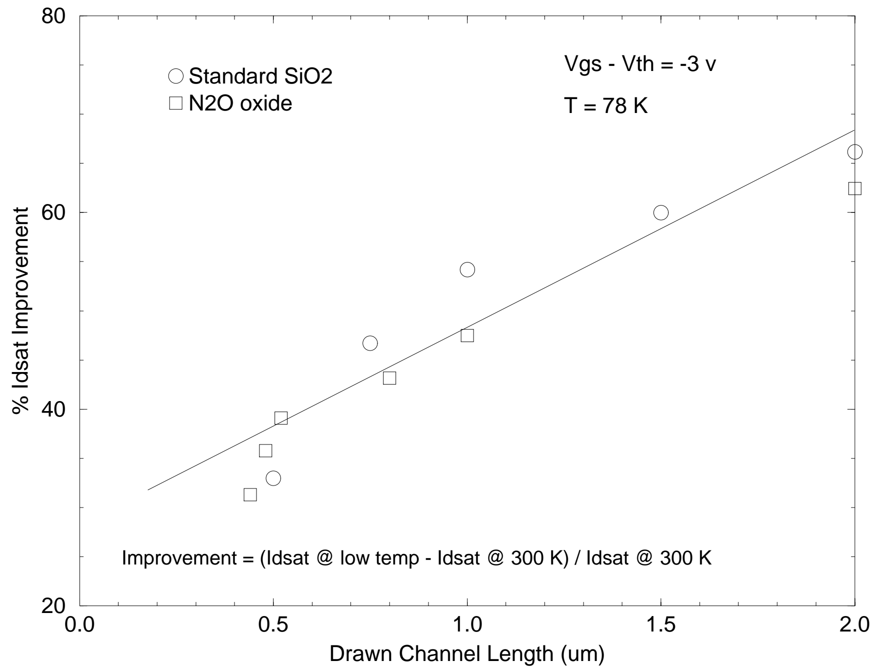

Figure 5.32: Improvement in saturation current versus channel length for PMOS transistors with nitrous oxide gates (N_2O) at 78 K. Results are normalized to 300 K at a drain and gate above threshold bias of 3 v. Standard oxide results are also shown for comparison.

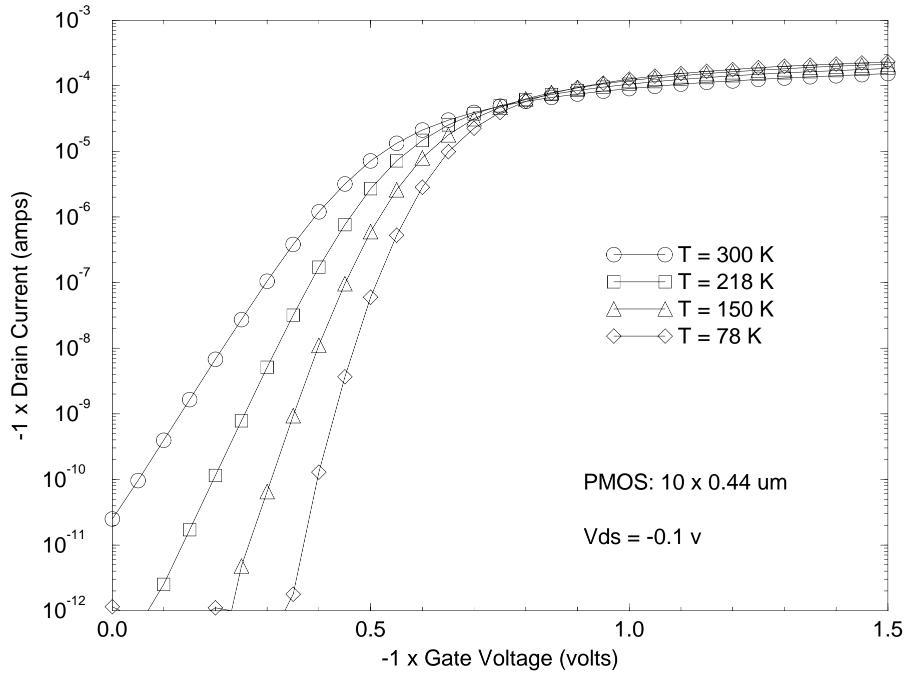

Figure 5.33: I_{ds}-V_{gs} characteristics (log scale) of a 10 μm x 0.44 μm PMOS transistor with N_2O oxide at different temperatures. Unlike standard oxide results, there is no leakage current at low temperature, indicating minimal boron penetration.

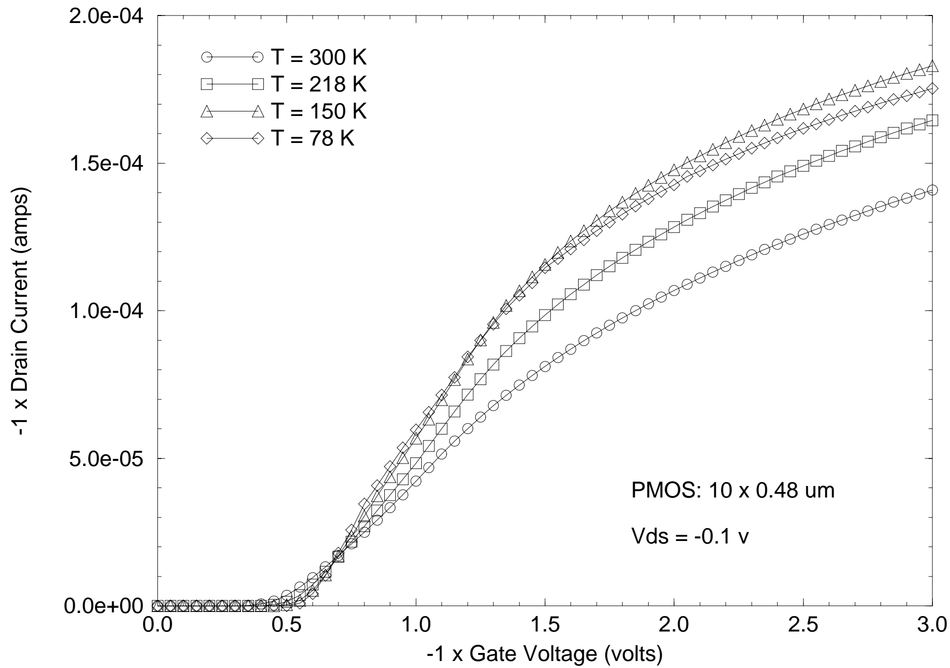

Figure 5.34: Measured I_{ds}-V_{gs} characteristics of a 10 μm x 0.48 μm PMOS transistor at different temperatures (linear scale). Due to freezeout in the extension region, the linear drain current at 78 K is lower than the linear drain current at 150 K for high gate biases.

roughly the same amount as between 218 K and 150 K (see Figure 5.6). Therefore, for optimum low temperature operation at 78 K, the extension dose must be at least $7x10^{13}cm^2$ for NMOS and $4x10^{14}cm^2$ for PMOS.

5.7 Low Temperature Ring Oscillator Performance

The propagation delay of a 61 stage ring oscillator at different temperatures is illustrated in Figure 5.35. The output pad of the buffer is connected to the 50 Ω input of an oscilloscope so that the period of the waveform can be determined. The

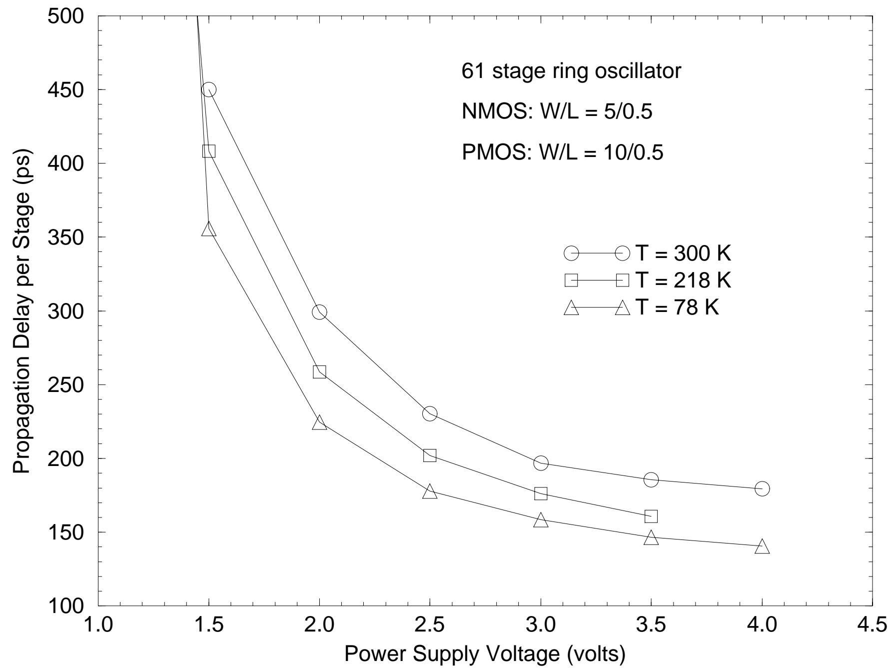

Figure 5.35: Propagation delay per stage of a 0.5 μm 61 stage ring oscillator at 300 K, 218 K, and 78 K.

measured delay is equal to this period divided by two times the number of stages. The expected gate delay is given by equation (5.1). The measured results are higher than the predicted value (Figure 5.36) because only the oxide capacitance was taken into account in the calculation.

Delay improvement for standard and retrograded profiles is shown in Figure 5.37. As predicted by the device results, an additional 35% improvement in delay at 78 K is achievable using retrograded profiling. The actual amounts of improvement at 78 K are lower than the device results because the threshold voltage shift cannot be eliminated from the circuit measurements. The threshold voltage shift can be eliminated by normalizing the 78 K retrograded ring oscillator ($V_{th,n}$ = -$V_{th,p}$ = 0.4 v) to the 300 K standard ring oscillator ($V_{th,n}$ = -$V_{th,p}$ = 0.4 v). Propagation delays

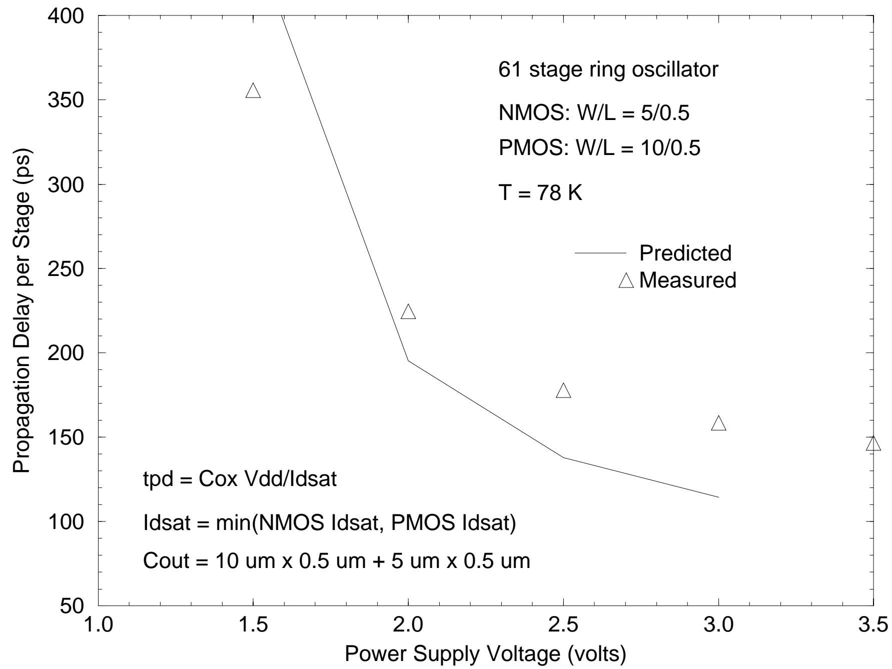

Figure 5.36: Measured and predicted propagation delay per stage of a 0.5 μm 61 stage ring oscillator at 78 K. The output capacitance is assumed to be the input (gate) capacitance of the next stage. The saturation current chosen is the minimum of the NMOS and PMOS value.

Figure 5.37: Improvement in propagation delay versus power supply of 0.5 μm 83 stage, 2-input NOR ring oscillators with different channel profiles at 78 K. Results are normalized to 300 K.

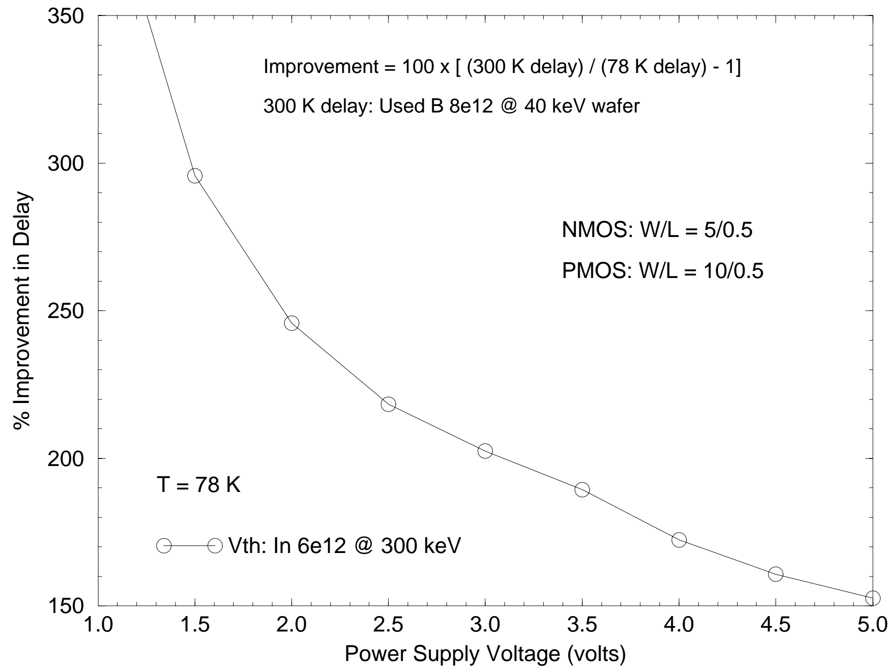

Figure 5.38: Improvement in propagation delay versus power supply of a 0.5 μm 83 stage, 2-input NOR ring oscillator with retrograded channel at 78 K. Results are normalized to a 300 K ring oscillator with a standard channel profile. The amount of delay improvement here includes the temperature change, the use of retrograde profiling, threshold voltage normalization, and improved metal conductivity.

are reduced by more 2x below power supplies of 3 v (Figure 5.38) and reach 3x at 1.5 v.

5.8 Conclusion

NMOS devices, PMOS devices, and CMOS circuits have been successfully fabricated and tested at low temperature. Process splits were made several steps in the process flow in order to evaluated their impact on low temperature performance. By replacing a standard channel profile with a retrograded profile, up to 35% improvement in current drive can be achieved due to the reduction in impurity scattering along

Process Change	Improvement	Reference
None (cooling only)	35%	Figure 5.11
Retrograde channel profile	+35%	Figure 5.16
Shift V_{th} back to room temp	+10%	Figure 5.23
Reduce V_{th} by 50% (maintain same I_{off})	+100%	Figure 5.25
Total improvement:	180%	

Table 5.6: Breakdown of expected improvements and their sources for 0.5 μm CMOS at 78 K and a power supply of 3 v.

the Si-SiO_2 interface. To further improve short channel effects, a halo implant can be added.

At low temperature, the threshold voltage can be reduced well below the room temperature value due to the improvement in subthreshold slope. Up to 60% improvement can be obtained with aggressive scaling of V_{th}. Freezeout effects can be eliminated by increasing the doses of conventional LDD implants and using a dual polarity gates. Although switching to p^+-poly results in possible boron penetration, implanting nitrogen into the polysilicon or using an high-κ gate dielectric can be employed to minimize its impact.

Table 5.6 summarizes the expected improvements that each successive process change will have on the performance of 0.5 μm CMOS at 78 K for a power supply voltage of 3 v. The additional improvement in circuit performance (see Figure 5.38) can be associated with higher metal conductivity. Reducing the power supply will reduce the current levels and increase the propagation delays, but the amount of improvement through cooling will be significantly more (see Figure 5.11).

Chapter 6

Conclusion

6.1 Summary

In this dissertation, device design issues for optimum low temperature CMOS have been presented. The improved current drive and subthreshold slope make low temperature CMOS ideal for high performance, digital applications. The reduction of thermal noise also makes low temperature CMOS attractive for analog circuits, such as low noise amplifiers. Conventional room temperature CMOS generally behaves poorly at lower operating temperatures, leading to a very pessimistic outlook on its usefulness. The sources of this poor performance are carrier freezeout and the increase in threshold voltage. To design better performing devices and circuits at low temperature, TCAD simulation tools can be used. A good fit to measurement results can be achieved by accounting for the temperature dependent modeling parameters. By modifying just a few steps of an existing, room temperature process flow, better low temperature performance can be achieved. Implementation of these minor

changes is more readily acceptable (from an economic standpoint) than proposing a novel device structure or process flow.

Threshold voltage plays an key role in determining performance improvement at low temperature. By not accounting for its increase, devices will be "off" for too much of the voltage swing. At 78 K, the threshold voltage can be reduced to 50% of its room temperature value without leading to an increase in off-state leakage currents. Freezeout effects can be minimized by increasing the dose of the extension implant and switching to dual-polarity gates. Although boron penetration becomes problematic with the use of p^+-poly, nitrogen implantation in the poly or the use of high-κ dielectrics can be used to solve this problem. Further refinements to the channel, such as retrograding the profile, will lead to additional improvements in low temperature performance. Along with proper scaling of the oxide thickness and junction depth, a 2x improvement in current drive and as much as a 3x improvement in circuit performance has been demonstrated at deep-submicron dimensions. Operating temperature can still be used today to scale the performance of most CMOS applications.

6.2 Suggestions for Future Work

Although this work has focused on device and some circuit performance issues at low temperature, several additional items can be addressed to further support the merits of low temperature CMOS. These include the following:

- Low temperature performance at sub-100 nm dimensions. Device design below 0.1 μm is currently underway [36]. The amount of improvement that can be achieved at this dimension through cooling is unknown. As velocity saturation

becomes more and more of a limiting factor (instead of mobility) at deep sub-micron dimensions, low temperature operation may be less attractive at 100 nm and below. Nevertheless, with proper scaling and process modification, a 2-3x improvement in performance may still be possible, as was seen from scaling 1 μm CMOS down to 0.5 μm.

- Low temperature reliability of alternative gate dielectrics. Although high-κ gate dielectrics and nitrogen-into-poly implantation can be used to suppress boron penetration, their impact on device reliability at low temperatures has not been thoroughly studied. Room temperature results are promising [28]-[29], and some low temperature work has been done on reoxidized-nitrided oxides [37], which should be similar to the thermally grown nitrous oxide gates presented here.

- Freezeout modeling. Current TCAD tools do not take Mott's transition into account when dealing with freezeout. Therefore, poor current drive typically results if the option is turned on. Additional parasitic source/drain resistance can be substituted, resulting in a better linear fit, but the subthreshold fit for buried channel devices is poor (Figure 3.16). A concentration dependent freezeout model at low temperature is clearly necessary to model the device behavior from the subthreshold to linear regime. Otherwise, it will not be possible to predict both linear and subthreshold device performance, and circuit performance can be under- or over-estimated.

- Higher level circuit performance. Extraction of the low temperature device characteristics presented in this work to HSPICE should be done so that the

performance of more advanced digital and analog circuits can be studied. Retrograding the doping profile may produce less desirable effects on the circuit level, such as an increased V_{th} shift with back-gate bias. This can lead to poor performance when transistors are stacked together in series.

Bibliography

[1] F.H. Gaensslen, V.L. Rideout, E.J. Walker, and J.J. Walker, "Very Small MOSFET's for Low-Temperature Operation", *IEEE Trans. on Electron Devices*, vol. ED-24, p. 218-229, 1977.

[2] Jack Y.C. Sun, Y. Taur, R.H. Dennard, and S.P. Klepner, "Submicrometer-Channel CMOS for Low-Temperature Operation", *IEEE Trans. on Electron Devices*, vol. ED-34, p. 19-26, 1987.

[3] M. Aoki, S. Hanamura, T. Masuhara, and K. Yano, "Performance and Hot-Carrier Effects of Small CRYO-CMOS Devices", *IEEE Trans. on Electron Devices*, vol. ED-34, p. 8-17, 1987.

[4] James Simmons and Joseph Madden, "Practical HTS/Cryogenic Systems for Wireless Applications", *Microwave Journal, Euro-Global Edition*, vol. 39, p. 124-136, 1996.

[5] J.J. Tzou, C.C. Yao, R. Cheung, and H. Chan, "Hot-Electron-Induced MOSFET Degradation at Low Temperatures", *IEEE Electron Device Letters*, vol. EDL-6, p. 450-452, 1985.

[6] M. Song, K.P. MacWilliams, J.S. Cable, and Jason C.S. Woo, "Bias and Temperature Dependence of Hot Carrier Lifetime from 77 K to 300 K", *IEDM Tech. Digest*, p. 707-710, 1992.

[7] M. Kakumu, D. Peters, H.Y. Liu, and K.Y. Chiu, "Design Optimization for Deep-Submicron CMOS Device at Low Temperature Operation", *IEDM Tech. Digest*, p. 211-214, 1990.

[8] C.Y Hwang, "Extraction of SPICE Parameters for CMOS at 300 K and 77 K", UCLA Master's Thesis, 1992.

[9] R.J. Gillick, "Temperature Dependence of the Hot Carrier Effect in NMOS Transistors", UCLA Master's Thesis, 1995.

[10] C.Y. Hwang, J. Gillick, C. Jenq, B. Hammond, and J. Woo, "Bias Dependent Hot-Carrier Reliability and Lifetime Over a Wide Temperature Range", *Journal de Physique IV (2nd European Workshop on Low Temperature Electronics)*, vol. 6, p. 25-28, 1996.

[11] C. Hu, S.C. Tam, F. Hsu, P. Ko, T. Chan, and K.W. Terrill, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement", *IEEE Trans. on Electron Devices*, vol. ED-32, p. 375-385, 1985.

[12] "TMA MEDICI: Two-Dimensional Device Simulation Program (versions 1.1.3, 2.0.2, 2.3.1, and 4.0.0)", *Technology Modeling Associates, Inc. (now Avant!).*

[13] S.E. Hansen, "Suprem-III User's Manual, Version 8520", Integrated Circuits Laboratory, Stanford University.

[14] "TSUPREM-4: Two-Dimensional Process Simulation Program (version 6.2.1)", *Technology Modeling Associates, Inc. (now Avant!).*

[15] J.T. Watt, "Surface Mobility Modeling", *Computer-Aided Design of IC Fabrication Processes*, Stanford University, 1988.

[16] J.T. Watt, "Modeling the Performance of Liquid-Nitrogen Cooled CMOS VLSI", *Stanford Electronics Lab Technical Report*, No. G725-3, 1989.

[17] "HSPICE User's Manual", *Meta-Software, Inc.*.

[18] B. Cheng and J. Woo, "Temperature-Dependent MOSFET Inversion Layer Carrier Mobility Model for Device and Circuit Simulation", *IEEE Trans. on Electron Devices*, vol. 42, p. 343-345, 1997.

[19] G.G. Shahidi, J. Warnock, S. Fischer, P.A. McFarland, A. Acovic, S. Subbanna, E. Ganin, E. Crabbe, J. Comfort, J.Y. Sun, T.H. Ning, and B. Davari, "High-Performance Devices for a 0.15 μm CMOS Technology" *IEEE Electron Device Letters*, vol. 14, p. 466-468, 1993.

[20] S. Venkatesan, J.W. Lutze, C. Lage, and W.J. Taylor, "Device Drive Current Degradation Observed with Retrograde Channel Profiles", *IEDM Tech. Digest*, p. 419-422, 1995.

[21] C.T. Sah, T.H. Ning, and L.L. Tschopp, "The Scattering of Electrons by Surface Oxide Charges and by Lattice Vibrations at the Silicon-Silicon Dioxide Interface", *Surface Science*, vol. 32, p. 561-575, 1972.

[22] T. Hori, "A 0.1 μm CMOS Technology with Tilt-Implanted Punchthrough Stopper (TIPS)", *IEDM Tech. Digest*, p. 75-78, 1994.

[23] B. Yu, C. Wann, E. Nowak, K. Noda, and C. Hu, "Short-Channel Effect Improved by Lateral Channel-Engineering in Deep-Submicrometer MOSFET's", *IEEE Trans. on Electron Devices*, vol. 44, p. 627-634, 1997.

[24] M. Song, J.S. Cable, Jason C.S. Woo, and K.P. MacWilliams, "Optimization of LDD Devices for Cryogenic Operation", *IEEE Electron Device Letters*, vol. EDL-12, p. 375-377, 1991.

[25] I. Groves, G. Brown, G. Pollack, K. Green, L. Dawson, A. D'Souza, C. Lin M. Song, C. Hwang, J. Woo, and K. MacWilliams, "One-Micrometer, Radiation-Hardened Complementary Metal Oxide Semiconductor for Cryogenic Analog Applications," *Infrared Readout Electronics II*, Proceedings of the SPIE, vol. 2226, p. 72-84, 1994.

[26] G. Guegan, M. Lerme, M. Ada-Hanifi, and D. Moi, "Comparative Study of Surface and Buried Channel 0.25 μm pMOSFETS", *Proceedings of the 26th European Solid State Device Research Conference (ESSDERC '96)*, p. 325-328, 1996.

[27] J. Pfiester, F. Baker, T. Mele, H. Tseng, P. Tobin, J. Hayden, J. Miller, C. Gunderson, and L. Parrillo, "The Effects of Boron Penetration on P^+ Polysilicon Gated PMOS Devices," *IEEE Trans. on Electron Devices*, vol. 37, p. 1842-1851, 1990.

[28] W. Ting, G. Lo, J. Ahn, T. Chu, and D. Kwong, "MOS Characteristics of Ultrathin SiO_2 Prepared by Oxidizing Si in N_2O", *IEEE Electron Device Letters*, vol. EDL-12, p. 416-418, 1991.

[29] G. Lo, W. Ting, J. Ahn, and D. Kwong, "P-Channel MOSFET's with Ultrathin N_2O Gate Oxides", *IEEE Electron Device Letters*, vol. EDL-13, p. 111-113, 1992.

[30] V. Matthews, R. Maddox, P. Fazan, J. Rosato, H. Hwang, and J. Lee, "Degradation of Junction Leakage in Devices Subjected to Gate Oxidation in Nitrous Oxide", *IEEE Electron Device Letters*, vol. EDL-13, p. 648-650, 1992.

[31] A. Ditali, V. Matthews, and P. Fazan, "Hot-Carrier-Induced Degradation of Gate Dielectrics Grown in Nitrous Oxide Under Accelerated Aging", *IEEE Electron Device Letters*, vol. EDL-13, p. 538-540, 1992.

[32] S. Nakayama and T. Sakai, "The Effect of Nitrogen in P^+ Polysilicon Gates on Boron Penetration into Silicon Substrate Through the Gate Oxide", *1996 Symposium on VLSI Technology*, p. 228-229, 1996.

[33] B. Yu, D. Ju, N. Kepler, T. King, and C. Hu, "Gate Engineering for Performance and Reliability in Deep-Submicron CMOS Technology", *1997 Symposium on VLSI Technology*, p. 105-106, 1997.

[34] "Virtuoso Layout Editor (version 4.3.4)", *Cadence Design Systems.*

[35] J.G.J. Chern, P. Change, R.F. Motta, and N. Godinho, "A New Method to Determine MOSFET Channel Length", *IEEE Electron Device Letters*, vol. EDL-1, p. 170-173, 1980.

[36] G. Timp, K. Bourdelle, J. Bower, F. Baumann, T. Boone, R. Cirelli, K. Evans-Lutterodt, J. Garno, A. Ghetti, H. Gossman, M. Green, D. Jacobsen, Y. Kim,

R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampfor, W. Mansfield, S. Moccio, D. Muller, L. Ocola, M. O'Malley, J. Rosamilia, J. Sapjeta, P. Silverman, T. Sorsch, D. Tennat, W. Timp, and B. Weir, "Progress Towards 10 nm CMOS Devices", *IEDM Tech. Digest*, p. 615-618, 1998.

[37] C. Lin, J. Cable, and J. Woo, "Temperature and Electric Field Characteristics of Time-Dependent Dielectric Breakdown for Silicon Dioxide and Reoxidized-Nitrided Oxides", *IEEE Trans. on Electron Devices*, vol. ED-42, p. 1329-1332, 1995.