1/f Noise in MOSFETs with Ultrathin Gate Dielectrics

by

Blaine Jeffrey Gross

S.B. Massachusetts Institute of Technology (1986)

S.M. Massachusetts Institute of Technology (1986)

Submitted in Partial Fulfillment
of the Requirements for the
Degree of
Doctor of Philosophy
in Electrical Engineering and Computer Science

at the

Massachvsetts Institvte of Technology

June, 1992

Signature redacted

Signature of Author _____
Department of Electrical Engineering and Computer Science
May, 1992

Signature redacted

Certified by _____
Professor Charles G. Sodini
Thesis Supervisor

Signature redacted

Accepted by _____
Professor Campbell L. Searle
Chairman, Departmental Committee on Graduate Students

ARCHIVES
MASS. INST. TECH.
JUL 10 1992
LIBRARIES

© Massachusetts Institute of Technology, 1992
All rights reserved

1/f Noise in MOSFETs with Ultrathin Gate Dielectrics

by

Blaine Jeffrey Gross

Submitted to the
Department of Electrical Engineering and Computer Science
on May 13, 1992 in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Abstract

This thesis is concerned with increasing the understanding of the mechanisms involved in $1/f$ noise in MOSFETs with ultrathin gate dielectrics. This is achieved through the characterization of N- and P-MOSFETs with both conventional oxide and reoxidized nitrided oxide (ROXNOX) dielectrics over a broad range of temperature and bias conditions. Characterization of single electron traps in deep-submicron N-MOSFET transistors allows the basic assumptions of previous $1/f$ noise models to be tested. Making reasonable assumptions allows the single electron trap results to be extended to model noise in P-MOSFET transistors.

A portion of this thesis is concerned with the development of an optimized 850°C thin (10 nm) low-pressure-furnace ROXNOX process. Previously, low pressure ROXNOX dielectrics formed at 950°C have been shown to: i) reduce interface state generation and electron trapping under electrical stress [1], and ii) eliminate interface state generation and reduce positive charge buildup under radiation stress [2], as compared to conventional oxide. A drawback of this previous work is that the Dt product, or thermal budget, of the 950°C ROXNOX process is too high for compatibility with modern VLSI processes. In this thesis, the ROXNOX process is demonstrated to be readily scalable from 950°C to 850°C through the use of higher partial pressures in the nitridation and reoxidation steps, resulting in the same quality dielectric formed with a lower Dt product.

The electrical characteristics of thin (10 nm) MOS gate dielectrics formed at 850°C by low pressure furnace nitridation of SiO_2 followed by an oxygen anneal (reoxidation) are described. Electrical characterization of capacitors and transistors with the 850°C ROXNOX dielectric is presented. The reduction of effective mobility with normal field of NMOS transistors with the ROXNOX dielectric is shown to be much less severe than for transistors with conventional oxide, especially at low temperatures. This behavior is absent in the PMOS transistors. PMOS transistors with the ROXNOX dielectric exhibit a similar reduction in mobility with normal field when compared to conventional oxide devices. Reliability of the 850°C ROXNOX dielectric is demonstrated through Fowler-Nordheim and channel hot-electron stressing results.

Deep-submicron self-aligned NMOS transistors with both conventional oxide and with reoxidized nitrided oxide gate dielectrics were fabricated using conventional photolithography and a photo-resist ashing technique at the polysilicon gate level. These devices, which had effective channel lengths and effective channel widths of under 1 μm each, are small enough that the removal of a single electron from the channel due to trapping causes an appreciable change in device current. Furthermore, the small channel area increases the probability that only one electron trap will be active

at a given bias and temperature. These two effects make characterization of single trapping events possible.

Single-electron traps were characterized through measurements over a range of gate bias and temperatures (from 80 K to 350 K). The time constant of the traps was found to be a strong function of temperature, displaying in general activation energies of hundreds of milli-electron volts. The time constant of individual traps was also found to be a relatively strong function of gate bias. Both of these results conflict with past assumptions of the behavior of single traps, an ensemble of which make up the $1/f$ noise spectrum observed in MOSFETs. The occupancy of each single-electron trap is well described by Fermi statistics assuming a single trap energy level with the trap located in the oxide at some small tunneling distance away from the Si/SiO_2 interface. Traps were observed at various distances from the Si/SiO_2 interface and with a variety of trap energies and activation energies. The measured noise spectra of the single electron traps was found to agree with previous estimates of the autocorrelation function of a random telegraph noise process. Trapping of a carrier from the channel perturbs the channel current both by reducing the *number* of carriers flowing in the channel, and by inducing a change in the *mobility* of channel carriers. The effect of the fluctuating trap occupancy to induce mobility fluctuations was found to dominate for the NMOS devices studied. No significant variation of normalized trap amplitude with temperature is observed.

A model of the inversion layer making use of the physics of quantized subbands was applied to modeling the behavior of single-electron traps. This inversion layer model accounts for quantization effects observed in the channel of a MOSFET at low temperatures and/or when the device is biased at a high gate bias. This level of detail is demonstrated to be necessary to accurately model the behavior of single-electron trapping events when the transistor is biased in strong inversion.

$1/f$ noise measurements are performed on N- and P-MOSFET transistors with both conventional oxide and ROXNOX dielectrics in the linear region over a range of bias and temperatures. This is an important contribution due to the lack of published data of $1/f$ noise in MOSFETs varying *both* bias *and* temperature. The importance of proper normalization of $1/f$ measurements for extraction of device parameters is discussed. The normalized drain current noise for both N- and P- MOSFET devices with thin (10 nm) dielectrics is found to be a strongly increasing function of gate bias, in contrast to the constant value observed for thicker oxides, and the constant value expected from the number fluctuation noise model. The behavior of PMOS transistor $1/f$ noise power is found to display a square law dependence on absolute temperature, in contrast to the linear temperature predicted by the Christensson et al. noise model [3]. The behavior of NMOS transistor $1/f$ noise at low gate bias is found to be relatively temperature independent.

A model of $1/f$ noise in MOSFETs which incorporates the measured single-electron trap characteristics is reported. This model predicts the variation of $1/f$ noise *versus* bias and temperature for both PMOS and NMOS devices. The model demonstrates that the noise in the MOSFET is due to the fluctuation of the number of carriers in the channel, and these fluctuations inducing local correlated fluctuations in the channel mobility. The induced mobility fluctuation effect is found to dominate the noise performance of NMOS devices at low gate bias and at low temperature. The induced mobility fluctuation effect is found to be insignificant in PMOS $1/f$ noise behavior.

Thesis Supervisor: Professor Charles G. Sodini

Title: Associate Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to acknowledge my advisor, Professor Charlie Sodini, for his technical, managerial, and personal support during my tenure as a graduate student. I appreciate not only his technical guidance, but also his 'hands-off' style of management which allowed me to explore areas of interest to me. His concern for my future career is also appreciated.

My colleague Kathy Krisch is gratefully acknowledged for her collaboration on the ROXNOX dielectric work over the last 5 years. It was always great to have someone in the office to bounce ideas off of, and Kathy would always take time out to work things through.

I would like to acknowledge my readers: Professor Alan McWhorter is acknowledged for his suggestions about the direction of the research for this thesis over the last year, and his meticulous proof-reading of the manuscript. Professor Dimitri Antoniadis is acknowledged for his technical input and support of this work through his affiliation with the SRC project. He is thanked for serving as a reader on my thesis committee, and providing useful suggestions.

Professors Rafael Reif, Professor Steve Senturia, Professor Marty Schmidt, Professor Jim Chung, and Professor Hank Smith are also acknowledged for their technical suggestions and support provided through their affiliation with the SRC project.

This work has been supported by the SRC under contract # SP-91-080, ONR under contract # N00014–90-J-1296, and IBM East Fishkill under contract # 1622. Partial support was obtained from fellowships from GTE and NCR. Characterization software was provided by Hewlett-Packard, as part of the HP-TECAP University Program.

Raj Jayaraman is acknowledged as my predecessor on the thin-gate dielectric project. Raj brought me up to speed on device measurement and dielectric growth technique when I started graduate school, and so I owe him thanks.

Gregg Dunn is acknowledged for his collaboration on the ROXNOX radiation paper, and for many valuable technical discussions.

The MTL staff is acknowledged for fabrication of most of the devices used in this research. In particular, Octavio Hurtado, Joe DiMaria, and Paul McGrath are acknowledged for technical support involving the low-pressure furnace in the Technology Research Laboratory, and Velma McClure, Brian Foley, Rudy Lia, Rob Cuikay, Bob Machinski, and Jim Bishop are acknowledged for technical support involving MOSFET fabrication in the Integrated Circuits Laboratory.

Professor Pierre Humblet is acknowledged for his assistance in working out the power spectral density of single-electron-trap fluctuations.

Professor Ping Ko, Professor Chenming Hu, and Dr. Zhihong Liu of UC Berkeley, and Professor T.P. Ma of Yale, are acknowledged for their collaboration on a project to verify the reliability of ROXNOX dielectrics compared to other ultrathin dielectrics.

I would like to thank Carolyn Zaccaria for many years of help with dealing with the bureaucracy of MIT. In this same vein, I would also like to acknowledge Kate Paterson and Pat Varley.

I would like to thank my *tovarishiy*: Curtis Tsai, Joe Lutsky, Gee Rittenhouse, Patrice Parris, Andrew Karanicolas, Fritz Herrmann, and Craig Keast for their technical help, and just general amusement. I would also like to thank Julie Tsai, Merit Hung, Ken O, Ken Szajda, Shujaat Nadeem, Rod Hinman, Steve Decker, and Monica Choi for enjoyable discussions.

On a more personal note, I would like to thank Pat and Brian Dixon for their support. They were always ready to lend a hand and it is greatly appreciated. I would also like to thank Lori Martinez and the rest of the Martinez clan for their friendship and support over the last 3 years.

Much of the credit for what I have accomplished so far must go to my parents; for providing me with the values and the basic skills necessary to succeed in life, and for standing behind me in my studies at MIT.

Last and most importantly, I would like to thank my best friend and wife, Elizabeth, for her constant love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Noise

All devices which carry electrical current have noise. In its most basic form, noise current arises from the discrete nature of the conduction processes in nature, due to the fact that electrical current is carried in discrete packets of single electrons. The existence of a well-defined conductance in a device is due to the mean value of a large number of scattering events of all the current carrying particles in that device [5]. The net effect of all these discrete processes is to cause the electrical current through a device to be a time-varying quantity. Current flow is generally considered to be continuous to first order, because the number of electrons involved in carrying a current is generally quite large (i.e., it takes $\approx 6 \times 10^{18}$ electrons/sec for a current of one ampere). For very high performance applications, however, noise can be a dominant concern.

There are many forms of noise in electron devices. The most common of these types is 'thermal noise', also known as 'Johnson noise' or 'Nyquist noise' [6]. Thermal noise arises from discrete scattering events in bulk conductance devices. Another form of noise is shot noise, which was first observed by W. Schottky in vacuum tubes. Shot noise is due to the discrete random emission of single electrons from the cathode to

the anode in tube or other transit time devices [7, 8]. The noise power spectral density (PSD) of both of these noise sources, defined as the amount of ac noise power in the signal per unit bandwidth as a function of frequency, is constant up to a characteristic frequency equal to the inverse of some characteristic time constant. For bulk conductance devices, this time constant can be very fast, corresponding to the mean time between scattering events ($\approx$ femtoseconds). For vacuum tubes or other transit-time devices, this time can be on the order of microseconds or longer, corresponding to the mean transit time of the device. In Figure 1.1a. is shown a graph demonstrating the frequency dependence of the noise power spectral density for both a thermal noise source, and a shot noise source. Shot noise and thermal noise are forms of 'white noise' due to their flat frequency spectrum.

At about the time that noise in electron devices was first being investigated, another form of noise was discovered which had a noise power spectral density inversely proportional to frequency. This form of noise, first reported by Johnson [7], was termed 'flicker noise', and is now commonly known as '$1/f$ noise'. A graph of the noise PSD of a flicker noise source is shown in Figure 1.1b. $1/f$ noise has been most commonly found in devices which are dominated by surface effects [9]. The devices used by Johnson were vacuum tubes in which the cathode was coated with various oxides. McWhorter observed $1/f$ noise in germanium filaments [10]. He attributed the $1/f$ noise in the filaments to electrons tunneling to traps distributed throughout the native oxide on the germanium filament. This mechanism of $1/f$ noise generation is called the McWhorter noise model. One of the most important electron devices today is the MOSFET transistor, which is inherently a surface dominated device. It is not surprising that $1/f$ noise dominates the low-frequency noise characteristics of MOSFET devices.

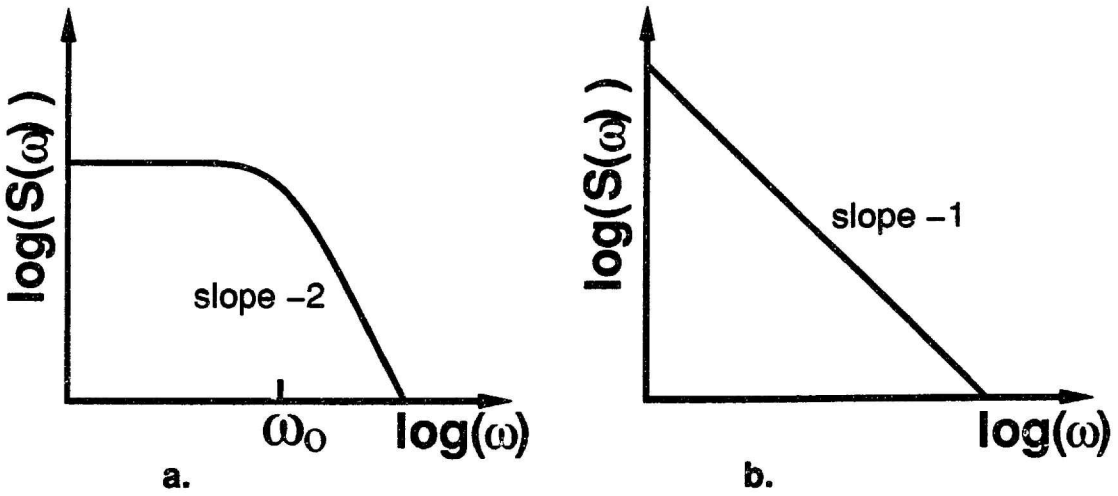


Figure 1.1: Noise PSD for fundamental noise sources. a) Shot or thermal noise PSD. b) Flicker or $1/f$ noise PSD.

1.2 $1/f$ Noise as a Characterization Tool

Traps in the region of the dielectric near the interface between the substrate and the dielectric have been implicated in a wide variety of effects in the operation of MOSFET devices, including $1/f$ noise behavior. Because $1/f$ noise is dependent on the near-interface region of the dielectric, it is believed that accurate modeling of $1/f$ noise can lead to a better understanding of the oxide trap density and therefore allow more insight to be gained into effects such as positive fixed charge near the Si/SiO_2 interface [11], degradation of short-channel MOSFETs under channel-hot electron stress [12 - 14], and the behavior of mobility under high normal fields [15, 16], all of which are of critical importance for high performance MOSFET operation. Furthermore, a strong correlation has been observed to exist between $1/f$ noise and the interface-state density near the silicon band edges [17]. Interface states in this region are important for operation of MOSFETs in moderate inversion [18], and are not detected by the common methods used for characterizing interface states in the

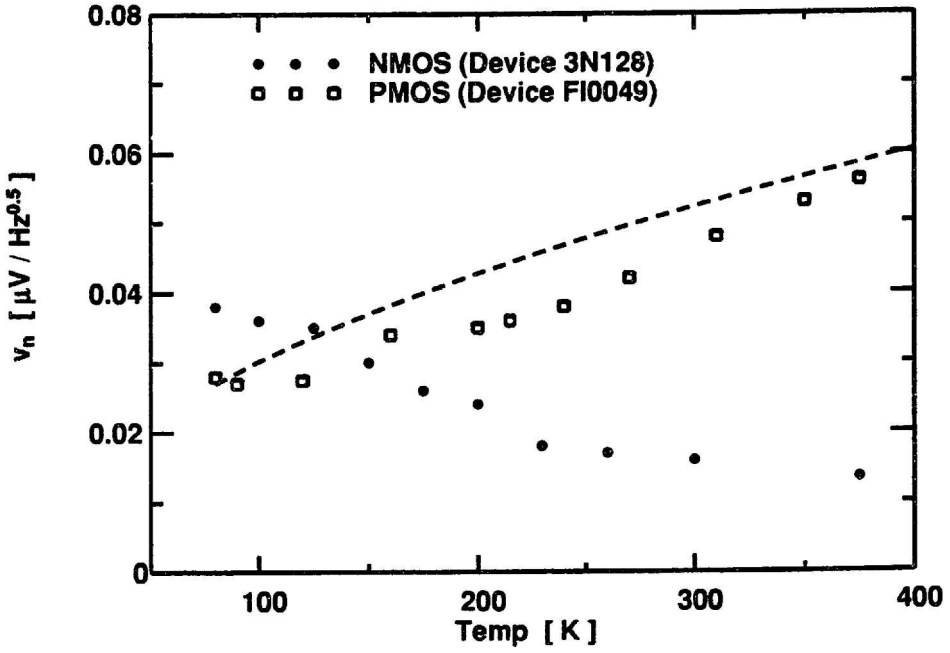


Figure 1.2: Temperature dependence of the equivalent input noise for two different MOSFETs at 50 kHz. The dashed line is the expected noise voltage from the model. Data replotted from Christensson et al. [4].

band-gap, such as weak inversion methods [19], high-frequency C-V [20], or charge-pumping [21].

A problem exists that currently, no single model of $1/f$ noise behavior can explain all of the observed behavior of $1/f$ noise in MOSFETs. The most widely applied model is based on the work of Christensson et al., who first applied the McWhorter noise model to MOSFETs [3]. Their model predicts that input referred $1/f$ noise power will exhibit a linear dependence on temperature. While their model is in qualitative agreement with the observed noise in PMOS transistors, it is inaccurate at modeling NMOS transistors over an extended temperature range (see Figure 1.2) [4]. Even with these shortcomings, the model has been commonly applied even to the present day. One of the major problems of the model is that the assumptions relating to the individual trap kinetics have never been tested in an empirical way.

It is the purpose of this thesis to improve on the $1/f$ noise model for MOSFETs, in particular, for N-channel MOSFETs. This work brings some unique tools to the problem: Reoxidized nitrided oxide (ROXNOX) dielectrics and random-telegraph noise devices.

1.2.1 Reoxidized Nitrided Oxide

The vast majority of studies of noise in MOSFETs have used conventional thermal SiO_2 as the gate dielectric. When slight variations exist in the noise performance among different conventional gate dielectrics, it is difficult to determine exactly what is different about the trapping/de-trapping processes which constitute the observed $1/f$ noise behavior. In effect, one is more likely to expect various conventional oxides to behave the same, than to expect them to behave differently. ROXNOX dielectrics have been observed to have a higher level of $1/f$ noise compared to conventional oxide, but in other ways to behave quite similarly to conventional oxide [22]. Thus we can assume that to first order, the differences in $1/f$ noise properties between devices with these two dielectrics are due to an increase in near-interface trap density.

The use of $1/f$ noise as a dielectric diagnostic tool may enable more insight to be gained into the nitridation and reoxidation processes. Recent work on the improvement of mobility of ROXNOX transistors following irradiation and anneal has demonstrated the utility of this technique [23].

1.2.2 Random Telegraph Noise Devices

As the size of a MOSFET is reduced, the number of traps contributing to $1/f$ noise in a device decreases. From estimates of trap density obtained from $1/f$ noise measurements, it can be determined that devices of size $\approx 1 \times 1$ μm will have only one trap active on average at a given time. Ralls et al. [24] was the first to report

the observation of single trapping events in MOSFETs with W/L=0.1 μm/1.0 μm. Since then there have been a number of additional reports on these trapping events [25 - 27]. Characterization of these single trapping events allows an accurate model of single-trap kinetics to be formulated.

A large number of papers in the recent past have contributed to what one researcher has called "...a long-running and rather sterile debate over 'mobility-fluctuation' *versus* 'number-fluctuation' models" [25]. Few results subject to unambiguous interpretation have arisen, mainly because most workers have concentrated on characterizing devices whose noise performance is due to a large ensemble of traps; even though the $1/f$ noise characteristic is quite distinct, the characteristic is also relatively featureless compared to the details of the individual traps. This thesis addresses these issues by carefully characterizing both the noise performance of large geometry devices, as well as the details of the individual traps which make up the ensemble of traps. Relating the noise of the trap ensemble to the details of the individual traps is an important contribution to the $1/f$ noise problem.

1.3 Summary of Results

The objective of this work was to improve the understanding of $1/f$ noise in MOSFET transistors. This was accomplished both by comprehensive characterization of MOSFET $1/f$ noise in the linear region over an extended range of bias and temperature, and by characterization of single-electron traps in deep-submicron MOSFETs. The characterization of single-electron traps makes it possible to test the most basic assumptions of current $1/f$ noise models. A secondary objective of this work was the development of a lower processing temperature optimized ROXNOX process.

The accomplishments of this thesis fall into three categories: the development of an optimized ROXNOX process, the characterization and modeling of single-electron

traps in deep-submicron MOSFETs, and the characterization and modeling of $1/f$ noise in conventional MOSFET devices.

The optimization procedure suggested in [28] was applied to developing a lower temperature ROXNOX process. An 850°C furnace ROXNOX process was developed which exhibited almost complete suppression of interface state generation under high field stress for capacitors and reduced transconductance degradation behavior under channel hot-electron stress for transistors, as compared to conventional oxide. The ROXNOX dielectric exhibited 20% lower mobility for NMOS devices and 10% lower for PMOS devices. Characterization results of effective mobility over a range of gate bias and temperature are reported.

Deep-submicron devices were fabricated, and used to characterize single-electron traps in the MOSFET channel. The individual capture and emission processes are shown to be well modeled as ideal Poisson processes. The capture and emission times of the single-electron traps in the channel are found to have a large temperature activation energy. The measured noise spectrum of the single-trap devices is found to agree with estimates of the autocorrelation function for the random telegraph process from [25]. The magnitude of the drain current change caused by the trapping of a single electron is observed to be much larger than the current change expected for the removal of a single electron from the conducting channel. This result indicates that the fluctuation of trap occupancy induces fluctuations in the mobility of carriers in the channel. The induced mobility fluctuation if found to be largest at low gate bias; at high gate bias screening acts to reduce the scattering cross section of the trapped charge. A qualitative model of trap fluctuation amplitude *versus* gate bias based on single-electron-trap characterizations is presented.

The single-electron traps are modeled physically using a multiphonon emission model of capture and emission [29]. The multiphonon emission mechanism models

the temperature activation behavior of the traps very well. Modification of current theories to account for variation of tunneling probabilities with gate field, also allows the gate bias dependency to be modeled well. A model of the inversion layer which accounts for quantization effects at high field and low temperature is applied to the single-electron-trap model.

The $1/f$ noise behavior of conventional N- and P-MOSFET devices in the linear region was characterized over a range of gate bias and temperature. This is an important result because of the lack of $1/f$ noise results in the literature presenting a comprehensive study of the variation of $1/f$ noise with both bias *and* temperature. The $1/f$ noise PSD of PMOS devices is found to follow a square-law dependence on temperature. This is a result which has not previously been recognized, although the data of previous workers fits this dependence [4]. In contrast to this, the NMOS $1/f$ noise PSD at low bias is found to be relatively temperature independent. Both N- and P-MOSFET devices with ROXNOX dielectrics were characterized, and were shown to have $1/f$ noise behavior similar to, although larger in magnitude than, that observed for oxide.

The $1/f$ noise characteristics of the conventional MOSFET devices is modeled using an ensemble of single-electron traps. The measured single-trap results conflict with past assumptions about the characteristics of single-electron traps, especially with regard to the measured temperature activated behavior of the traps, and the measured rapid variation of the trap time constant with bias and temperature. This is an important result, because the characteristics of an ensemble of traps depend critically on the properties of the individual traps in the ensemble. The behavior of NMOS device $1/f$ noise is modeled by the dominant induced mobility fluctuation effect of the individual traps in the dielectric. The behavior of PMOS device $1/f$ noise is modeled well by assuming that the induced mobility fluctuation effect of single-hole

traps is negligible. The model not only predicts the noise performance of the devices characterized for this work, but also models well the noise results of Christensson et al. [4].

The utility of using $1/f$ noise to characterize the near-interface region of the dielectric was demonstrated by a study investigating an anomalous mobility increase observed in ROXNOX devices subjected to ionizing radiation.

1.4 Organization of the Thesis

In Chapter 2 the processing techniques used to fabricate the devices in this thesis are discussed. In Chapter 3 a discussion of the measurement techniques used for characterizing the devices is presented. In Chapter 3 there is also a thorough discussion of the measurement circuits used for characterizing both $1/f$ noise and the single-electron-trap devices. A discussion of the 850°C ROXNOX process and the optimization method used for its development is presented in Chapter 4.

The single-electron-trap measurements are presented in Chapter 5. Here an inversion layer model based on the physics of quantized subbands is also discussed. In Chapter 6 the $1/f$ noise results are presented. The model of Christensson et al. [3] is examined in detail. The model for single-electron-trap behavior is used to develop a noise model applicable to large geometry MOSFETs. The conclusions of this thesis and suggestions for future work are presented in Chapter 7.

Chapter 2

Device Processing Techniques

A number of device structures were fabricated and characterized for the results presented in this thesis. Polysilicon-gate capacitors were used as a 'short-loop' process for reoxidized nitrided oxide gate dielectric optimization. NMOS and PMOS transistors were used for characterizing the threshold voltage, mobility, and $1/f$ noise. Small geometry NMOS transistors were used to characterize single-electron trapping phenomena. The details of the fabrication of these structures are given below.

2.1 Gate Dielectric Fabrication

The majority of the dielectrics reported in this thesis were grown in the Technology Research Laboratory (TRL) of the Microsystems Technology Laboratory (MTL) at M.I.T., in a special low-pressure oxidation furnace. Details of the operation of this furnace have been given elsewhere [22]. These details of operation will be reviewed here along with some modifications.

2.1.1 Low-Pressure Furnace

The low-pressure (LP) furnace is a three-zone, resistively heated furnace. The tube is composed of double-walled quartz, for high purity dielectrics. Ball joint seals are located at the gas inlet and at the vacuum outlet connection near the load-end

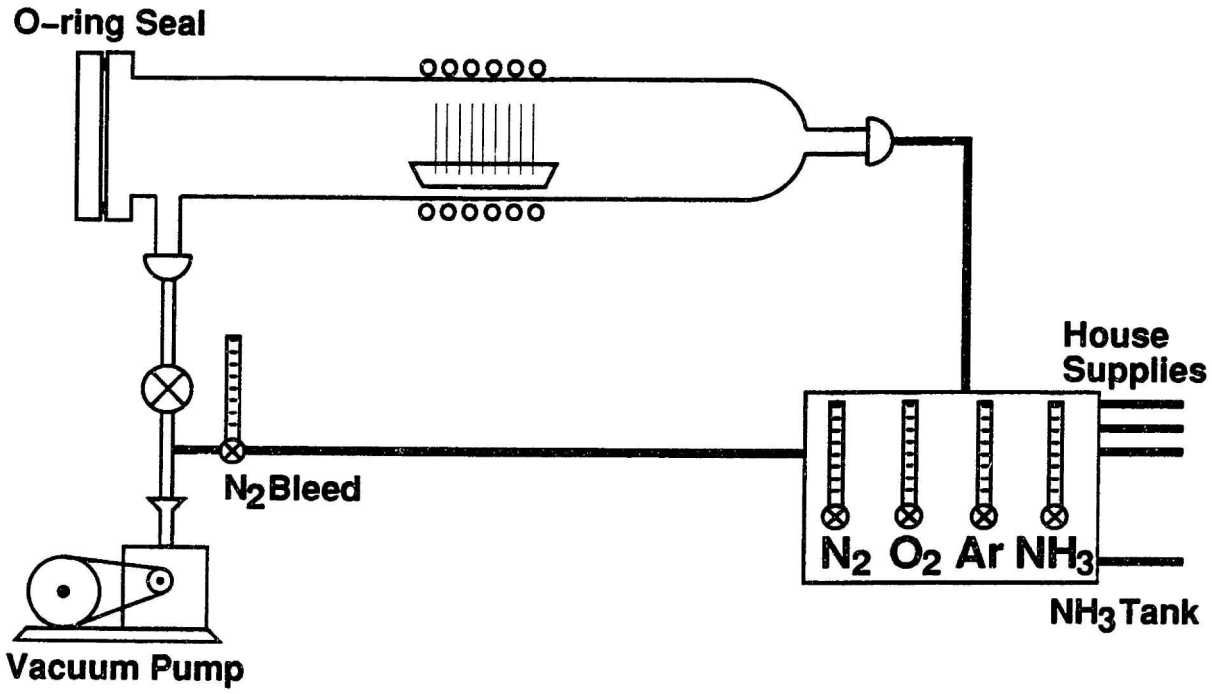


Figure 2.1: Schematic diagram of the LP furnace system. The N_2 bleed gas line is used to dilute the gases exhausted from the furnace before they enter the pump.

of the tube (see Figure 2.1). The door to the furnace is composed of aluminum plate and is mated to the furnace with an O-ring seal. The furnace is maintained at low pressure with a roughing pump. The base pressure of the system is in the 10 mTorr range. In general, the equipment used for the LP furnace system is similar to the equipment used for low-pressure chemical vapor deposition (LPCVD) in common use in the semiconductor industry. The furnace is capable of operating in a temperature range of up to 950°C in low pressure or atmospheric operation. Higher temperatures than this cause the door to be too hot to handle, and risks damaging the O-ring seal. The temperature of the furnace is periodically calibrated using a thermocouple junction. Careful records are kept of oxide thickness from each run to insure that temperature drift is not significant from run to run.

A gas jungle supplies the furnace with NH_3, O_2, N_2, and Ar gas. The NH_3 gas comes from a bottled 99.995% anhydrous source. The other gases are supplied by the

house supplies which are liquid source, and are rated at better than 99.999% at the source. The flow of each of the process gases is adjusted using flowmeters. All of the gas lines open to a manifold which is evacuated to base pressure when the furnace is evacuated to base pressure. When the furnace is idled, a low rate of nitrogen is flowed through the furnace, to prevent the backstreaming of pump oil from the vacuum system into the oxidation tube.

The pressure in the furnace is controlled by a large valve with a smaller bleed line connected in parallel to bypass this valve. Adjusting the large valve allows gross control of the pressure, while adjusting the smaller bypass valve allows fine control of the pressure of the furnace. During the time that O_2 or NH_3 is being flowed, a low flow rate of N_2 is added to the vacuum line to dilute the process gases. This avoids the deleterious effects of 100% O_2 and NH_3 on the pump and pump oil. For atmospheric pressure runs, a check valve on the door is used to exhaust the process gases from the tube to the house exhaust system. A scavenger system enclosing the load-end of the tube allows even atmospheric NH_3 runs to be processed.

Immediately prior to gate oxidation, the wafers were cleaned in an RCA clean, during which an HF dip is performed to remove the sacrificial oxide layer. The HF dip is performed until the wafers sheet (display hydrophobic effects), then they are overetched for an additional 30 seconds. Sheeting is observed when the oxide is removed from the dice streets and the device active areas, and the underlying hydrophobic silicon repels the HF solution immediately when removed from the acid bath. After being spun dry, the wafers are loaded into the oxidation furnace. Generally, more than one dielectric run is performed in a given day. For many of these cases, only one RCA clean is performed for both runs. One set of wafers is loaded into the furnace, and the other set is left in an enclosed carrier box. The amount of time wafers were left after cleaning and before being loaded in the furnace was in all

cases less than 6 hours. Note that immediately after the RCA clean, the wafers are somewhat passivated by a chemical oxide grown during the final step of the clean (the ionic clean part, performed in a hydrochloric acid/peroxide solution). This chemical oxide is estimated to be on the order of 20 Å thick from ellipsometry measurements. In any event, comparison of results of dielectrics grown immediately after RCA cleaning with those performed some time after RCA cleaning, demonstrates that allowing the wafers to wait some time after the RCA clean has a negligible effect on dielectric properties.

The wafers are loaded into the oxidation furnace at a rate of ≈9"/min. Nitrogen is flowed during loading, and the temperature of the furnace is held at 850°C. After the wafers are positioned in the center of the furnace, the door is fastened in place, and the tube is evacuated to base pressure. This step is performed to remove any contamination of the ambient which may occur from backstreaming of air into the furnace during the time that the door is off. To ramp to operating temperature, the pressure of the tube is adjusted to 500 mTorr by adjusting the N_2 flow, the temperature is turned to the proper set point, and the tube is allowed 20 minutes to come to the proper temperature. This has been found to be a sufficient amount of time by temperature vs time calibration measurements of the tube.

Changes in ambient pressure in the tube are accomplished by flowing N_2 until the proper pressure is reached and then switching over to the proper process gas. Ramping between process pressures always took less than 5 min. Between each process step, the tube is evacuated to base pressure. This step was found to be necessary by initial experiments on fabricated reoxidized nitrided oxide dielectrics using a diluted process gas at atmospheric pressure, instead of simply low pressure. It became evident that even long timed purges of the tube between oxidation and nitridation steps was not sufficient to remove all traces of O_2 from the tube. This oxygen contamination

prevented the formation of true nitrided oxides, and instead, dielectrics resembling reoxidized nitrided oxides were obtained, even if a reoxidation step was not performed. This information is not meant to imply that dilute nitrided oxide processes are not possible, but simply that with the equipment in use for this study, development of a fully atmospheric dilute nitrided oxide process is not practical.

Following conclusion of processing, the wafers are unloaded from the furnace at a rate of 9"/min. The wafers are allowed to cool for 5 minutes after removal from the tube before being handled. The wafers are then taken to the Integrated Circuits Lab (ICL) in a clean room box for LPCVD deposition of undoped polysilicon. In instances when two dielectric runs are performed in the same day, the first run out of the tube is usually held until the second run is complete. This was usually only one or two hours but wafers have waited up to 6 hours before polysilicon deposition. No adverse effects were observed for wafers waiting for up to 6 hours for polysilicon deposition. Occasionally, due to equipment downtime, wafers have waited over 24 hours between conclusion of oxidation and deposition of polysilicon. All these wafers have had nearly zero yield.

2.1.2 ICL Control Oxide

For comparison to oxides fabricated in the TRL LP furnace, dielectrics consisting of the 230 Å control oxide from the ICL baseline CMOS process were also fabricated. These dielectrics were fabricated at 950°C [30].

2.2 Polysilicon-Gate Capacitor Process

LOCOS-isolated n^+ polysilicon-gate capacitors with the various gate dielectrics were formed on n-type 10-20 Ω·cm (100)-orientation 4-inch silicon substrates. The polysilicon and the back side of the wafer were simultaneously heavily doped with

phosphorus at 925°C from a POCl source. During measurement, contact to the capacitor is made directly to the substrate from the back side and directly to the polysilicon top-plate with a probe. All capacitors received a final anneal treatment in forming gas (20% H_2/80% N_2) at 400°C for 15 minutes. These capacitors were used for capacitance-voltage and constant-current stress measurements. A process flow of the polysilicon-gate capacitor process is given in Appendix A.

2.3 CMOS Transistor Process

LOCOS-isolated n^+ polysilicon-gate NMOS and PMOS transistors with various gate dielectrics were formed in twin-well, 11-mask, 1.75 μm CMOS process. The wells were formed in a p-type epitaxial layer on a p^+ substate. All substrates were 4-inch (100)-orientation wafers. The process is identical to the process described in [30], with the exception of the omission of the p-channel threshold adjust implant. The omission of this implant causes the p-channel devices to be surface-channel devices, in contrast to the buried-channel devices obtained in the standard process. The channel implant doses were the same for each type of device used in this study. The implant doses are designed for tailoring the threshold voltage for the standard process, which uses a 230 Å gate dielectric [30]. This means that the threshold voltages for the 100 Å process are lower than is optimal for conventional CMOS. However, this does allow the impact of the dielectric variation on device characteristics to be more readily determined. BPSG is used as a dielectric isolation layer between the polysilicon and the Al/1%Si metalization. All transistors received a final anneal treatment in 20% forming gas at 400°C for 5 minutes. These transistors were used for low-field mobility, channel hot-electron stress, and $1/f$ noise characterization. The process flow of the CMOS process used is included in Appendix B.

2.4 Deep-Submicron NMOS Transistor Process

LOCOS-isolated n^+ polysilicon-gate NMOS transistors with various gate dielectrics were formed on p-type 20-40 Ω·cm silicon substrates using a standard four-mask process. This 4-mask process is simply a subset of the full CMOS process described above. All substrates were 4-inch (100)-orientation wafers. The device array layout contains devices with drawn channel lengths from 0.5 μm to 1.6 μm in increments of 0.1 μm and drawn channel widths down to 0.4 μm. A photoresist ashing technique, similar to that used by Chung et al. [31], was used at the gate-definition layer to allow deep-submicron channel lengths to be obtained. BPSG was used as a dielectric isolation layer between the polysilicon and the Al/1%Si metalization. All transistors received a final anneal treatment in 20% forming gas at 400°C for 5 minutes. The transistor process was not optimized for the deep-submicron feature size, but none the less, devices with effective channel length of 0.2 μm were obtained. The process flow of the deep-submicron NMOS process used is included in Appendix C.

Chapter 3

Measurement Techniques

3.1 Electrical Characterization Overview

The measurement system for MOS capacitors and transistors is shown schematically in Figure 3.1. All of the measurement instruments were interfaced, via a HPIB interface bus, to a HP200 series computer running TECAP [32] measurement software. The computer was used to control the measurement instruments and to perform data analysis. An HP4140 picoammeter was used to measure the quasi-static capacitance-voltage characteristic of MOS capacitors. An HP4275 LCR meter was used to measure high-frequency capacitance. An HP4145 semiconductor parameter analyzer was used to measure current-voltage characteristics of transistors and capacitors, to provide a constant bias during high-field stress measurements, and to provide an external bias to the HP4275 LCR meter. An HP3185 spectrum analyzer was used for $1/f$ noise measurements. An HP3165 dynamic signal analyzer was used for data capture for measurement of single-electron-trap characteristics. Custom-built low-noise amplifier circuits were used for $1/f$ noise measurements and for measurement of single-electron-trap characteristics. The custom-built amplifier circuits are described in Sections 3.4.1 and 3.4.2 respectively.

A probe station enclosed in a light-tight box was used for the C-V measurements.

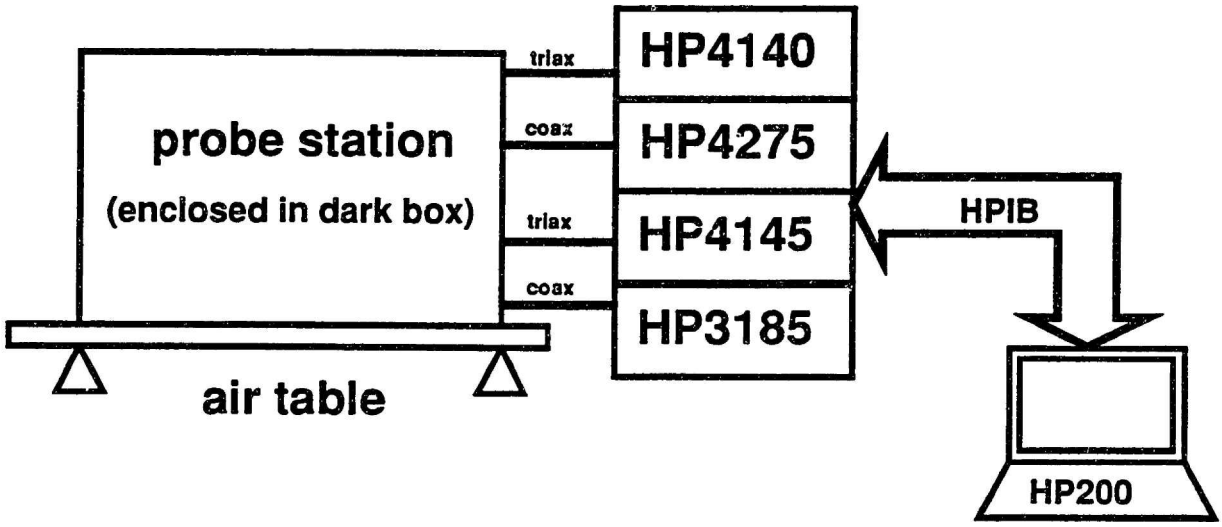


Figure 3.1: Diagram of the measurement system used to characterize MOS capacitors and transistors.

The probe station was located on an air-isolation table to minimize interference from vibrations. For many of the transistor measurements, the door to the light-tight box was left open. Comparison with measurements performed with the box closed demonstrated that the effect on device characteristics was negligible. For low temperature measurements, a MMR LTMP4 system was used, providing measurement temperatures from 80 K to 400 K.

Further details of each of these measurements are given in the following sections.

3.2 Capacitor Measurements

The details of the capacitance measurement technique can be found in [28]; I will briefly outline the techniques below.

3.2.1 Quasi-Static C-V

The quasi-static C-V characteristic was measured using the internal supply of HP4140 picoammeter with a ramp rate of 100 mV/sec. The C-V characteristic was measured from inversion to accumulation on polysilicon gate capacitor structures. Contact to the substrate was usually made via the back side of the wafer. The voltage ramp was connected to the substrate, and the current from the gate electrode was monitored. This technique minimized the effects of leakage currents from the probe-station vacuum chuck.

The capacitor is biased in inversion and a light is briefly (< 10 sec) turned on to create an inversion layer. The voltage is then swept, and the current from the gate electrode monitored. The quasi-static capacitance is then determined from solving the equation

$$I_g = \frac{\partial}{\partial t}(C(V)V) = C(V)\frac{\partial V}{\partial t} \tag{3.1}$$

for the capacitance as a function of the gate bias $C(V)$.

3.2.2 High-Frequency C-V

The HP4275 multifrequency LCR meter was used at a measurement frequency of 100 kHz and a signal amplitude of 30 mV. The bias-voltage for the HF-CV measurement was supplied externally by the HP4145 semiconductor parameter analyzer. The C-V characteristic was measured from inversion to accumulation. At the start of the measurement, when the device is biased in inversion, a light is briefly turned on to create an inversion layer, and avoid the effects of deep depletion [11].

3.2.3 Device Parameter Determination

Oxide Thickness

The oxide thickness was determined using the high-frequency capacitance in accumulation, $C_{ox(HF)}$, and the drawn capacitor dimensions, using the relation

$$t_{ox} = \frac{A\epsilon_{Si}}{C_{ox(HF)}}, \tag{3.2}$$

where A is the device area. The smallest capacitor structure characterized on each die was 100×100 μm; hence, any error due to using drawn dimensions instead of effective dimensions will be on the order of 1%. The bulk index of refraction of nitrided oxides and reoxidized nitrided oxides formed by 60 minute 1000°C 100% atmospheric NH_3 anneals has been shown to be within a few percent of the refractive index of conventional oxide [33]. Because the refractive index is ideally related to the square root of the dielectric constant, a dielectric constant of $\epsilon_{ox} = 3.9\ \epsilon_o$ was assumed regardless of whether the dielectric was pure oxide or some form of nitrided oxide. Oxide thickness characterized in this way yielded results consistent with the values calculated by ellipsometry.

Substrate Doping

The substrate doping N_d is determined from the measured steady-state value of capacitance in depletion, and the oxide thickness. The measured capacitance in depletion, C_M, is related to the depletion capacitance C_D by the relation

$$\frac{1}{C_M} = \frac{1}{C_D} + \frac{1}{C_{ox}}. \tag{3.3}$$

Hence, we can write

$$C_D = \left(\frac{1}{C_M} - \frac{1}{C_{ox}}\right)^{-1}. \tag{3.4}$$

The value of C_D above can be solved for assuming the depletion region is at its maximum equilibrium width $x_{d_{max}}$. The depletion capacitance can be written in the depletion approximation, assuming uniform substrate doping and assuming that in inversion the Fermi level is pinned at the conduction band, as

$$C_D = \sqrt{\frac{qN_d\epsilon_{Si}}{2(\frac{E_g}{2} + \phi_F)}}, \tag{3.5}$$

where E_g is the bandgap of silicon, and ϕ_F is the bulk Fermi level relative to midgap. The value of ϕ_F is related to the substrate doping and the intrinsic carrier concentration n_i by the relation

$$\phi_F = \frac{kT}{q} \ln \frac{N_d}{n_i}. \tag{3.6}$$

Equation (3.5) can be solved transcendentally for N_d.

Flatband Voltage

The flatband voltage V_{FB} is a unique function of the gate-substrate work function difference ϕ_{MS} and any oxide charges in the MOS system [11],

$$V_{FB} = \phi_{MS} + \frac{Q_{ox}}{C_{ox}}, \tag{3.7}$$

where Q_{ox} is the effective fixed positive charge in the dielectric referred to the Si/SiO_2 interface. The capacitance of the substrate at the flatband condition, C_{FB}, is a unique function of the extrinsic Debye length of the substrate

$$C_{FB} = \frac{\epsilon_{Si}}{\lambda_{ext}} = \sqrt{\frac{q^2N_d\epsilon_{Si}}{4kT}}. \tag{3.8}$$

It is now possible to determine the gate voltage at which the flatband condition occurs, by finding the measured capacitance value which satisfies the relation

$$C_{M,FB} = \left(\frac{1}{C_{ox}} + \frac{1}{C_{FB}}\right)^{-1}. \tag{3.9}$$

By knowing the value of work function for the polysilicon one can characterize how much oxide charge is in the dielectric. Using a value of $\phi_M = 0.42$ eV relative to the midgap level, which is similar to the value of ϕ_M reported by Hickmott et al. for heavily phosphorus doped polysilicon, gave consistent results for our control oxides independent of oxide thickness [34].

The value of V_{FB} was used to calculate the amount of fixed charge in the dielectric using Equation (3.7). Note that in dielectrics with a high density of interface states the value of flatband voltage determined from the above analysis will be in error. No effort was made to account for this.

Interface-State Density

The interface-state density was calculated using the high-low C-V method [11]. The basis for this measurement lies in the observation that for quasi-static C-V, the effective bulk capacitance C_B is the sum of the effective interface-state capacitance C_{it} and the depletion capacitance, i.e.,

$$C_B = C_{it} + C_D. \tag{3.10}$$

The measured quasi-static capacitance is therefore

$$C_{M(LF)} = \left(\frac{1}{C_{ox}} + \frac{1}{C_{it} + C_D}\right)^{-1}. \tag{3.11}$$

The high-frequency capacitance, in contrast, is largely unaffected by the interface-state density, since the interface states do not respond to the high-frequency signal. Hence we can write the measured high-frequency capacitance as

$$C_{M(HF)} = \left(\frac{1}{C_{ox}} + \frac{1}{C_D}\right)^{-1}. \tag{3.12}$$

Using these two relations, we can solve for the interface-state capacitance as

$$C_{it} = \left(\frac{1}{C_{M(LF)}} - \frac{1}{C_{ox}}\right)^{-1} - \left(\frac{1}{C_{M(HF)}} - \frac{1}{C_{ox}}\right)^{-1}. \tag{3.13}$$

This equation yields a mapping of interface-state density to gate voltage. To determine the mapping of gate voltage to surface potential ψ_s requires performing a numerical integration from $V_{FB} \equiv (\psi_s = 0)$, as is described in Nicollian and Brews [11]. In this thesis, interface state values are generally given at the gate bias for which the surface potential corresponds to the midgap level.

Current-Voltage Characterization

An HP4145 semiconductor parameter analyzer was used to measure the current-voltage (I-V) characteristics of capacitors. The I-V measurements were used to confirm that the current through the capacitors was area, and not perimeter, dominated. Past experience has demonstrated that perimeter current can indicate thinning of the dielectric along the edge of the capacitor, a symptom of the Kooi effect [35]. The existence of premature breakdown is often evident from the I-V characteristics. For all of the dielectrics in this thesis, the I-V characteristics are described well by the Fowler-Nordheim characteristic.

3.3 Transistor Measurements

The HP4145 semiconductor parameter analyzer was used to measure the I-V characteristics of MOSFETs. A number of data extraction routines were implemented in Hewlett-Packard's TECAP program [32]. They are described in the next section.

3.3.1 V_t Extraction

Threshold voltages are extracted using the method of Sun and Plummer [36]. A program was integrated into TECAP to extract the point of maximum slope of a linear region I_{ds} vs V_{gs} plot, and then extrapolate from that point using the maximum slope to a threshold voltage. The program yields the maximum transconductance g_m

in addition to the threshold voltage. The minimum HP4145A step-size of 10 mV was used when characterizing V_t to obtain good resolution.

3.3.2 ΔL and ΔW Extraction

The effective channel length of a series of devices was found by measuring the linear region drain-source resistance of a number of devices with identical drawn widths and various drawn lengths following the method of Chern et al. [37]. The drain-source resistance R_{DS} can be thought of as the sum of the linear region channel resistance R_{CH} and the parasitic source-drain series resistance R_S arising from resistance in source-drain diffusion, contacts, and metal lines, so that

$$R_{DS} = R_{CH} + R_S. \tag{3.14}$$

The linear region drain current can be written as

$$I_{DS} = \frac{W}{L_D - \Delta L}\mu V'_{ds} C_{ox}(V_{gs} - V_t - \frac{V'_{ds}}{2}), \tag{3.15}$$

where L_D is the drawn channel length, ΔL is the change in channel length due to imprecision in gate definition plus lateral encroachment of the source-drain diffusions, and V'_{ds} is the voltage across the channel region,

$$V'_{ds} = V_{ds} - I_{ds}R_S. \tag{3.16}$$

We make the assumption that the value of ΔL is identical for devices in close proximity to each other; i.e., those devices which are on the same die. The channel resistance can now be defined as

$$R_{CH} = \frac{V'_{ds}}{I_{ds}} = \frac{L_D - \Delta L}{W\mu C_{ox}(V_{gs} - V_t - \frac{V'_{ds}}{2})}. \tag{3.17}$$

Plotting R_{DS} vs L_D for various values of $(V_{gs} - V_t)$, a set of curves is obtained which intersect at the point $R = R_S$ and $L_D = \Delta L$.

In practice, the lines for each $(V_{gs} - V_t)$ value do not all intersect at a single point. Lines intersecting at a single point implies a well-behaved device. To find the single point of intersection from scattered data, I have implemented a program in TECAP to give the average of all the insections of each set of 2 lines, to obtain a set of R_S and ΔL values. The program gives the mean and standard deviation of this set of data. The standard deviation is generally less than 10% of the mean value of R_S and ΔL. For the results presented on deep-submicron MOSFET devices, the values of ΔL were extracted using the four smallest device lengths which yielded working transistors.

The difference between drawn channel width W_D and effective channel width is found in a similar way. The difference ΔW is found by comparing the values of linear region drain current at a given gate bias for devices with various drawn channel widths, but the same channel length. This proceeds directly from Equation (3.15) above, replacing W by $(W_D - \Delta W)$.

For the extractions for the deep-submicron MOSFETs, no effort was made to account for variations in parasitic resistance among devices with different drawn channel widths. The reason for this is that the fraction of parasitic resistance due to the resistance of the source-drain diffusions should scale to first order with device width. Indeed, the linear region drain current of a series of devices with the same drawn channel length at constant gate bias increased linearly with drawn channel width, as long as the device widths were not greatly different. For the results presented here, values of ΔW were extracted using the 4 smallest drawn widths which yielded working transistors.

3.3.3 Small-Signal Operating Point Parameters

To normalize the $1/f$ noise measurements, it is necessary to characterize the small-signal values of device transconductance and drain conductance at the given

bias points. A program was written to determine these parameters using the 4145 semiconductor parameter analyzer. The gate bias and in turn the drain bias are each varied by ± 10 mV about the given bias. Following each change in the gate or drain bias the device was allowed to stabilize for 1 sec before the drain current was recorded. From the variation in the drain current with the small changes in drain voltage, the drain conductance of the device is determined. Similarly, from the variation of the drain current with the gate bias, the transconductance is determined. These parameters are used to determine the normalized voltage gain of the device for $1/f$ noise measurements.

3.3.4 Low-Temperature Measurements

An MMR LTMP-4 microprobe station was used to characterize transistor parameters, $1/f$ noise, and single-electron traps over an extended range of temperatures. The LTMP system works using the Joule-Thompson effect, by which the temperature of gas at high pressure is lowered when released through a nozzle [38]. The LTMP system consists of a vacuum chamber and a refrigerator unit which is inserted into the vacuum chamber. Four microprobes allow the user to probe devices on the refrigerator stage. The stage is large enough (1 cm$\times$1.5 cm) to allow single die to be mounted. The vacuum is necessary to ensure good thermal insulation of the refrigerator from the surrounding ambient.

The LTMP-4 refrigerator works by flowing high-pressure (1800 psi) N_2 gas into the refrigerator. The gas is released from pressure under the refrigerator stage, and this cooler gas flows through channels back over the inlet line, cooling the high-pressure inlet gas, before being evacuated to the external house vacuum. In this way, the temperature of the stage can be regeneratively lowered all the way down to liquid nitrogen temperature. A heater on the refrigerator stage allows the temperature to

be raised to any temperature above the minimum, from 80 K up to 400 K.

A diffusion pump system was used to maintain the vacuum level in the chamber at the 1.0 mTorr level. This level of vacuum was required to minimize the amount of heat leakage to the LTMP refrigerator. The diffusion pump had the tendency to backstream small amounts of oil, which was evident from the appearance of a microscopic oil film on the device under test (DUT) with time. By comparison of devices measured with the oil film, with devices measured in a regular probe station, the effect of the film on device characteristics was found to be negligible. It was necessary to use IR shielding to maintain proper refrigerator operation. The IR shielding used was a mylar film coated with aluminum. This film was wrapped around the refrigerator, leaving only the refrigerator stage exposed, and resulted in an drastic improvement in refrigerator performance.

3.4 Noise Measurements

3.4.1 $1/f$ Noise Measurement System

A diagram of the noise measurement system is shown in Figure 3.2. Battery bias is used for low noise operation. Wirewound potentiometers were found to have very low drift and hence low noise. Capacitors were chosen for use based on their leakage characteristics; leaky capacitors will introduce excess noise into the measurement system. Polycarbonate and polyester film capacitors, found to have the lowest leakage currents, were used to fabricate the measurement system.

An Analog Devices OP-37 ultralow-noise operational amplifier was used to amplify the noise signal (amplifier A1 in Figure 3.2). The gain of the noise system was periodically calibrated to be 100 by adjusting resistor R1. A lead network was used as series compensation for this amplifier to improve the dynamic performance and to

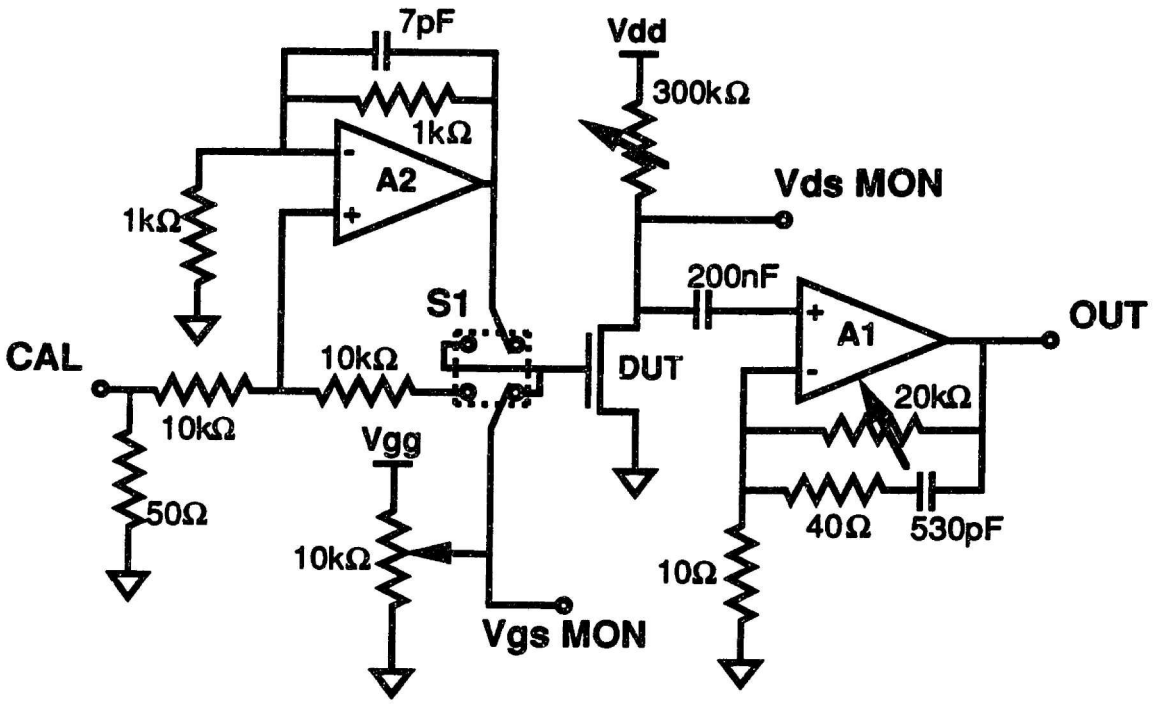


Figure 3.2: $1/f$ noise measurement system. The switch $S1$ is a dual-pole, dual-throw switch used for calibration.

maximize the bandwidth of the system [39]. A general drawback of lead compensation is that it increases the sensitivity to noise in a system [39]. This is not a drawback in our case, but actually an advantage because we are characterizing noise *power spectral density*, whereas generally in feedback systems it is integrated noise power which is a concern. An HP3585 spectrum analyzer was used to characterize the noise PSD at each measurement point. The HP3585 uses internal averaging to average 100 readings at a given frequency when determining noise PSD. The minimum calibrated input signal level for the HP3585 is listed as 30 nV/$\sqrt{\text{Hz}}$, with an ultimate resolution, or noise floor, of 10 nV/$\sqrt{\text{Hz}}$, for a measurement bandwidth Δf of 3 Hz. To be more precise, readings of the HP3585 between 10 nV/$\sqrt{\text{Hz}}$ and 30 nV/$\sqrt{\text{Hz}}$ will not be to the same high precision as measurements in the higher ranges.

The gain of the measurement system was calibrated using a built-in calibration circuit (the left half of Figure 3.2). The dual-pole, dual-throw switch S1 is thrown to connect amplifier A2 to the gate of the DUT. Amplifier A2 is an Analog Devices AD-744 precision BiFET operation amplifier and is connected in a summing configuration. The output of the amplifier A2 is the sum of the CAL input and the VGS bias value. An ac signal is input to the CAL input, and the output of amplifier A1 is measured. The gain of the amplifier is determined from the ratio of the VDS MON and OUT values across a range of frequencies, and is calibrated to a voltage gain of 100 for all of the measurements reported in this thesis.

To reduce the number of cables connected to the noise measurement circuit housing, the CAL and VGS MON leads are multiplexed with a switch to a single BNC jack. Similarly, the OUT and VDS MON leads are multiplexed with a switch to a single BNC jack.

Static proved to be a major problem when working with the noise measurement system, especially when used in conjunction with the low temperature probe station. This problem was alleviated to a large extent by using a grounding wrist strap and by installation of a switch to short circuit all the leads of the DUT together whenever connecting or disconnecting equipment, or when probing a new device. Another switch allowed the DUT to be connected to the HP4145 for I-V measurements without changing cable connections. Whenever the DUT was switched between the noise measurement system and the HP4145, the DUT terminals were grounded together.

The input equivalent noise for the noise measurement system depends a great deal on the output impedance of the DUT. For a nominal value of g_{ds} of 1.0×10^{-3} S, the input equivalent noise of the system is estimated to be a maximum of 5 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ at 30 Hz. Note that the noise contributed by the V_{gs} bias resistor is negligible. It can be shown that the *drain*-referred noise of the input resistor in the linear region follows

the relation

$$\overline{v_d^2} \simeq 4kT(10k\Omega)\left[\frac{V_{gg}V_{gs}V_{ds}}{(V_{gs}+V_{gg})^2(V_{gs}-V_{th})}\right]^2. \tag{3.18}$$

For the values of $V_{gg} = 12$ volts, $V_{ds} = 0.2$ volts, and worst case values of $V_{th} = 1$ volt and $(V_{gs} - V_{th}) = 0.2$ volts, the above equation yields $\overline{v_d} = 1$ nV/$\sqrt{\text{Hz}}$. Higher values of V_{gs} yield lower values of drain referred noise. This value of noise can be neglected when compared to the larger noise contributed at the drain node from the amplifier system. We can obtain a worst case performance of the noise system by analyzing the lowest current devices. The largest geometry (lowest current) devices measured were 20 μm/5 μm which, for the NMOS (PMOS) case, have an output resistance of $\approx$3 kΩ ($\approx$8 kΩ) at $V_{gs} - V_t = 0.5$. For this value of output resistance, the noise system has a maximum input-referred noise, or minimum detectable signal level, of 4 nV/$\sqrt{\text{Hz}}$ (6 nV/$\sqrt{\text{Hz}}$) at 1000 Hz. These estimates were made using worst case values of noise performance listed in the specification sheets for this op-amp. The $1/f$ corner frequency for this amplifier is listed as 140 Hz. The $1/f$ noise floor of the system can be estimated by using the published input noise sources at 30 Hz, resulting in a maximum input-referred noise for the system of 8 nV/$\sqrt{\text{Hz}}$ (19 nV/$\sqrt{\text{Hz}}$) at 30 Hz. The input equivalent noise for the measurement system typically yields a 3 dB improvement in noise floor, but a much greater improvement in measurement accuracy, given the fact that the HP3585 is uncalibrated for use below 30 nV/$\sqrt{\text{Hz}}$. Note that for all the devices reported in this thesis, the noise was well above the $1/f$ noise floor of the measurement system.

The main problem affecting the noise measurement was external interference. This interference comes from sources such as fluorescent lights, cathode-ray tubes, and machinery in and around the laboratories. The most pervasive problem was interference from 60 Hz and its harmonics. An attempt was made to shield the the 60 Hz interference by using a grounded dark box and allowing the HP3585, the measurement

circuit, and probe station to float independent of the dark box, similar to the approach used by Jayaraman [22]. Much better results were obtained by grounding the HP3585 and the measurement system to the dark box. While this approach helped somewhat, it did not eliminate the problem entirely. This problem was especially evident when measuring device drain voltage noise directly, bypassing the amplifier circuit. It was clear that bringing the high-impedance drain node all the way out to the measurement instrument aggravated the problem of 60 Hz interference. Using the measurement circuit effects an impedance transformation: the node connected to the HP3585 is the low-impedance op-amp output node, which reduces the pickup of 60 Hz interference. As final precautions, noise measurements were not performed at frequencies near harmonics of 60 Hz, and all of the measurement equipment connected together on the same HPIB bus were connected to a single power strip.

For all of the noise data in this thesis, the normalization of the device noise PSD is defined as

$$\text{Normalized Drain Current PSD} \equiv (V_{gs} - V_{th})^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2}, \tag{3.19}$$

which yields a value of noise PSD which is in units of [volts^2/Hz]. This normalization is chosen because it yields a quantity which has all the explicit bias dependencies normalized out, making it most useful for making comparisons between different types of devices. Note that the noise measurement above yields the drain voltage noise PSD $S_{V_{ds}}(f)$. We can write the normalized drain current PSD in terms of $S_{V_{ds}}(f)$ as

$$(V_{gs} - V_{th})^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2} = (V_{gs} - V_{th})^2 \frac{g_{ds}^2 S_{V_{ds}}(f)}{I_{ds}^2}. \tag{3.20}$$

The values of g_{ds} and I_{ds} are extracted as described in the Section 3.3.3 above.

Quite often in the literature, the value of gate-referred noise is given when reporting noise characterization results. The reasons for this is that the gate-referred noise

is of interest to circuit designers and is generally relatively bias independent. The bias independence assumption proceeds directly from (3.20) above as

$$\text{Normalized Drain Current PSD} = (V_{gs}-V_{th})^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2} = (V_{gs}-V_{th})^2 \frac{g_m^2 S_{v_{gs}}(f)}{I_{ds}^2} \simeq S_{v_{gs}}(f), \tag{3.21}$$

demonstrating that at low bias levels in the linear region the gate-referred noise PSD is approximately equal to the normalized drain current noise PSD defined above. Care must be taken in reporting gate-referred noise values for devices with ultrathin dielectrics, as the assumptions which lead to the above results are violated [40]. We can quantify the assumption used in reporting gate-referred noise PSD by rewriting (3.21) as

$$S_{v_{gs}}(f) = \left[\frac{I_{ds}^2}{g_m^2(V_{gs}-V_t)^2}\right] \cdot (V_{gs}-V_t)^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2}. \tag{3.22}$$

For an ideal device in the linear region, the factor in brackets in Equation (3.22) will be approximately equal to 1. In fact, for an ideal device in the saturation region, the factor in brackets in Equation (3.22) will also be equal to 1. For a device with an ultrathin gate dielectric (or any device operated at high values of gate overdrive), the effect of mobility reduction at high normal gate field [41] will become important, causing a reduction in the value of g_m, and the term in brackets in Equation (3.22) will not be equal to 1. This effect is especially important to consider when evaluating reports of $1/f$ noise in dielectrics when the value of gate-referred noise is reported.

For each bias point, the noise power spectral density was measured at 17 distinct frequencies logarithmically spaced between 20 Hz and 50 kHz. The data is plotted as normalized drain current noise, and a line was fit to the data on a log-log plot. This fits the data to the equation

$$\frac{S_{I_{ds}}(f)}{I_{ds}^2} = \frac{A}{f^\gamma}, \tag{3.23}$$

where the slope of the line gives γ, and the intercept at $f = 1$ Hz gives A. The average

error of the fit can be determined by calculating the mean square-deviation of the data points from the line. This standard deviation will correspond to the average distance a data point is away from the the fit line, and was generally 5%.

3.4.2 Single-Electron Traps

Measurement System

The measurement system used for characterization of the random telegraph noise (RTN) signals is shown schematically in Figure 3.3.

The amplifier circuit consists of a transresistance first stage, amplifying the drain-source current of the DUT by a factor of 10,000 Ω. Following this is an amplifier with a gain of 100, for a total transresistance gain of 10^6 Ω. The amplifier system was equipped with a offset cancellation circuit, to compensate for drift in the amplifier output. An HP3561 Dynamic Signal Analyzer was used for data acquisition. This instrument allows up to 40960 time-data points to be acquired at one time.

Analog Devices OP-37 ultralow-noise operational amplifiers were used in the forward signal path (amplifiers A1 and A2 in Figure 3.3). An Analog Devices AD744 precision BiFET operational amplifier was used in the offset cancellation circuit (amplifier A3 in Figure 3.3). This op-amp was chosen because of its very low offset voltage, and very high differential input impedance ($\approx 10^{12}$ Ω). The offset cancellation circuit determines the low-frequency cutoff of the amplifier transfer function. The measured relaxation time of the offset cancellation circuit was approximately 5 seconds. A switch was closed during changes in bias to allow the measurement system to quickly settle to its operating point. The on-state resistance of the switch is 10 kΩ and effectively reduces the time constant for integrating offset charge on capacitor C_c.

The input impedance of the measurement system is calculated to be 10^{-2} Ω. The

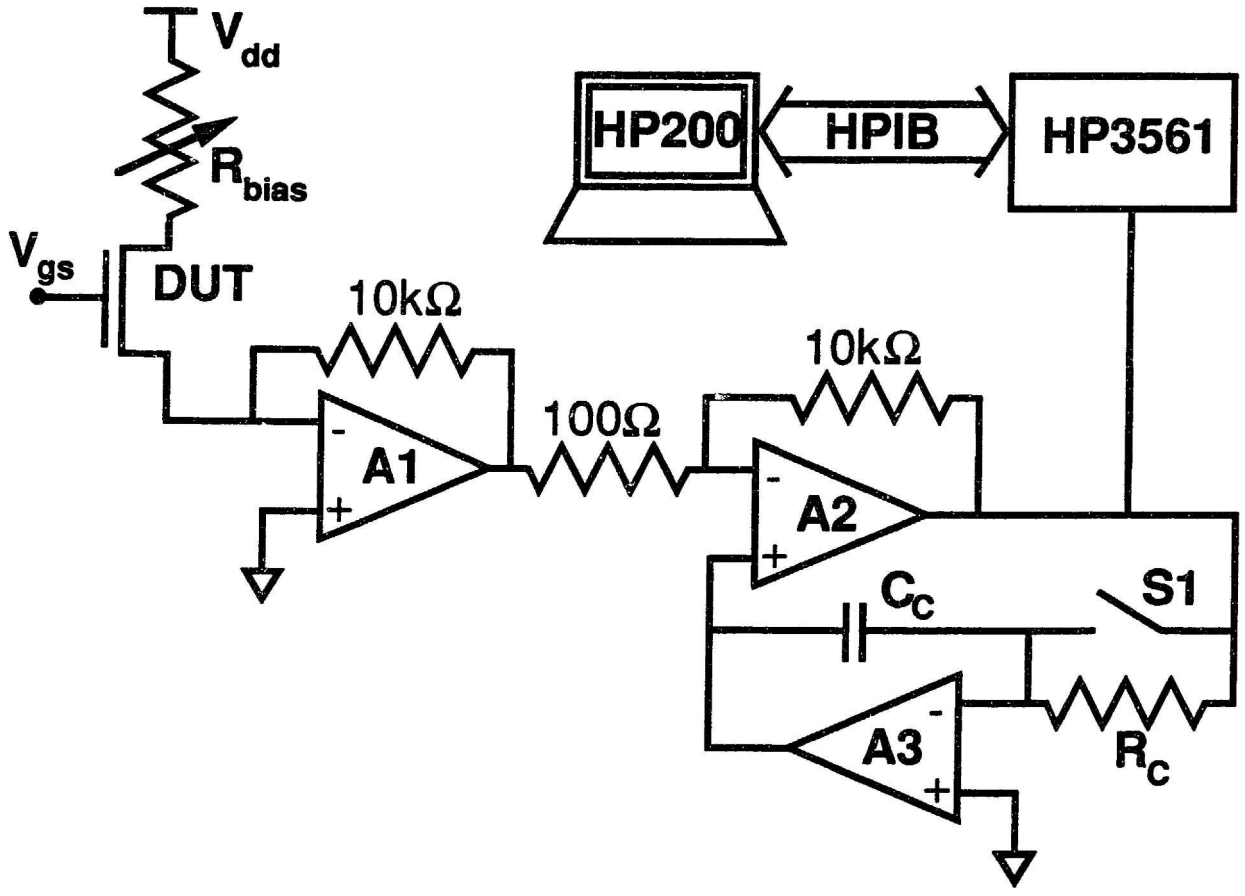


Figure 3.3: RTN measurement system. The switch $S1$ has an on-state resistance of 10 kΩ.

offset voltage of the first-stage op-amp will make the input of the amplifier system be at a nonzero level. An offset trimming resistor was used to adjust the input voltage to zero, which corresponds to setting the input offset to zero. The AD-OP37 is rated to produce up to 10 mA of output current before some loss of linearity occurs. From the data sheets, the absolute maximum current that the op-amp can drive into a 100 Ω load is 40 mA. These conditions mean that we prefer to operate below a input current (DUT drain-source current) of $\approx$100 μA, with a hard limit occurring at 400 μA. These limits are compatible with the linear region drain currents of the very small geometry MOSFETs to be characterized.

A 500 Ω resistor is tied across the differential inputs of the first stage of the amplifier as part of a compensation network (not depicted) to improve the amplifier stability. A single-pole filter was used at the LNA output to reduce excess high-frequency noise. The single-pole filter could be continuously adjusted between 1.6 kHz and 16 kHz. The worst case minimum amplitude resolution of the measurement system is calculated to have an rms value of $<$ 1 nA over the measurement band from 0.1 Hz to 16 kHz based on the specification sheets for the AD-OP37. This worst case occurs with the base time length of the HP3561 set to any a value less than 32 msec. For base times longer than this, the signal will be band-limited to the Nyquist rate, and the minimum detectable signal level will be decreased. Traps which cause a change in drain current as small as 1 nA have been characterized.

The device is biased by a resistive network connected to batteries for low noise. The drain bias resistor R_{bias} has a significant effect on the measurement of the magnitude of the drain current fluctuations, which is accounted for by the technique described here. All of the measurements reported were performed with the device biased in the linear region, where the channel can be modeled as a resistance R_{ds}. The fluctuations in drain current can be modeled as a fluctuation in the channel resistance

by making use of the relation

$$R_{ds} = R_{DS} + \delta R. \tag{3.24}$$

The total current fluctuation caused by the drain-source resistance fluctuation can be written as

$$\delta I_{meas} \simeq \frac{\delta R}{R_{DS} + R_{bias}} I_{ds} = \frac{\delta R}{R_{DS} + R_{bias}} \frac{V_{dd}}{R_{DS} + R_{bias}}. \tag{3.25}$$

At each value of gate bias, the drain bias resistor R_{bias} is adjusted to set the drain-source voltage V_{ds} to be a constant value. This has the effect of setting the ratio of the device resistance to the total resistance to be

$$\frac{R_{DS}}{R_{bias} + R_{DS}} = \frac{V_{ds}}{V_{dd}}. \tag{3.26}$$

Using this relation, the drain current fluctuation for the case of a constant drain voltage bias can be inferred as

$$\delta I_{ds} = \frac{\delta R}{R_{DS}} I_{ds} = \frac{V_{dd}}{V_{ds}} \delta I_{meas}. \tag{3.27}$$

The relation (3.25) was verified by measurement of a single-electron trap under both resistive bias, and using a low-impedance voltage source bias.

Capture/Emission Time Extraction

The HP3561 dynamic signal analyzer has a buffer which allows the acquisition and storage of up to 40280 points in time. The data is organized in the instrument's memory in 40 records of 1024 data points each. The time base of the machine is set to be the length of time for one record of data. Hence the minimum time base of 4 msec corresponds to a total time data capture length of 160 msec for all 40 records. The time base used for the RTN measurements varied from the minimum value of 4 msec to a maximum of 5 sec. The HP3561 is capable of time base length of up to 39062.5 sec, for a total time capture length of approximately 18 days.

The mean time in the high state, $<\tau_{high}>$, and the mean time in the low state, $<\tau_{low}>$, for a given trap are extracted from the RTN signal-vs-time data capture using a HP200 series computer. The data is down-loaded from the HP3561 and is operated on by a thresholding operation. A hysteresis value of 0.5 mV is used during the thresholding operation to reduce the effects of random noise causing false transitions.

A number of routines were implemented in the data extraction program to improve the signal-to-noise ratio of the data. The simplest of these is a binomial low-pass filter. This is an averaging operation of each data point with the point in time before it and with the point in time after, with relative weights of 0.5, 0.25, and 0.25 respectively. This eliminates the highest frequencies of random noise from the time-capture. Performing N iterations of this filter on the data produces the net result of a $2N^{th}$-degree binomial filter, which in the limit of a large N will result in a Gaussian frequency response [42].

For many of the traps characterized, the measured trap amplitude was only marginally higher than the noise of the measurement system (i.e., the signal to noise ratio, S/N $\simeq$ 10 dB). For trap amplitudes of lower signal-to-noise ratio than this, the extraction of time constants was prone to error, and so these results were discarded. This could introduce an unintentional statistical bias into the distribution of trap amplitudes characterized towards higher values.

Chapter 4

Reoxidized Nitrided Oxide

In a direct extension of the work of Yang [28] and Jayaraman [22], who developed a 950°C reoxidized nitrided oxide (ROXNOX) process, a lower temperature ROXNOX dielectric has been developed. The motivation for developing a lower temperature ROXNOX process is to make the ROXNOX dielectric more easily integratable into current VLSI processes. The main benefit of using a ROXNOX dielectric in a VLSI process is that it gives much higher reliability for only a 20% reduction in channel mobility. Reliability here is defined as resistance to degradation under channel hot-electron stress (CHE) and resistance to degradation under radiation stress.

The main benefit of the various nitrided oxide and ROXNOX dielectrics fabricated en route to the optimum 850°C process is the varied properties of these dielectrics in terms of interface-state density, oxide trap density, and oxide fixed charge density.

4.1 The ROXNOX Process

The nitridation of silicon dioxide, first introduced by Ito [43] and Naiman [44], has been shown by several authors to provide resistance to interface-state generation under electrical stress [45], to reduce sensitivity to radiation [46], and to produce a barrier to various dopants and contaminants [47]. However, the furnace nitridation

process, which is typically performed in an anhydrous NH_3 ambient at atmospheric pressure and 1000°C for about 1 hour, is known to cause both a high fixed charge density [48] and a large number of electron traps [49] in the resultant dielectric. These traps and charges result in shifted threshold voltages, degraded inversion layer mobilities [15], and reduced stability. The reoxidation of nitrided oxides [45, 50] improves dielectric stability by eliminating the electron traps while maintaining reduced interface state generation under electrical stress.

The reoxidation of nitrided oxides formed by heavy, high-temperature (above ≈900°C), atmospheric furnace nitridations does not reduce fixed charge densities to levels required by high performance scaled MOSFETs with high inversion layer mobilities [15]. Therefore we have investigated the use of light, low pressure (LP) furnace nitridations (0.01 atm., typically), which differ from atmospheric furnace nitridations only in that the increase and subsequent turnaround in fixed positive charge and interface-state density occur more gradually with nitridation time. A number of other researchers have pursued the use of rapid thermal processing (RTP) to similarly obtain light nitridations at atmospheric pressure [51 - 54]. The reoxidation of LP nitrided oxides has been found to be necessary to eliminate the electron traps as well as to provide the improved suppression of interface-state generation under electrical stress [22].

Future generation CMOS processes will require PMOS transistors with heavily boron doped polysilicon gates [55, 56]. These devices can be subject to the penetration of boron through the dielectric to degrade the device characteristics [57]. Reoxidized nitrided oxide has been shown to reduce the amount of boron penetration for these PMOS devices [58 - 60]. The full effect of this process on transistor characteristics has not yet been determined.

Results of the optimization of a 950°C ROXNOX process and electrical charac-

terization of 950°C ROXNOX capacitors and transistors through C-V, I-V, constant-current stress, channel hot-electron stress, and inversion layer mobility measurements have been presented [1]. The 950°C ROXNOX process has the drawback of a processing temperature which is too high for compatibility with existing VLSI processes. We have therefore investigated the nitridation and reoxidation processes at lower temperatures.

In this chapter, results of a furnace 850°C optimized ROXNOX process are presented which demonstrate suppression of interface-state generation and electron trapping under Fowler-Nordheim stress and a factor of five improvement in resistance to degradation under channel hot-electron stress.

4.2 ROXNOX Process Optimization

The behavior of thermal nitridations can be explained qualitatively as follows. For the purposes of this discussion, *stability of the interface* will pertain to resistance to interface-state generation under Fowler-Nordheim stress. For very *light* nitridations (low NH_3 pressure, low temperature, and/or short time) the Si/SiO_2 interface is less stable, exhibiting greatly increased interface-state generation with constant current stress. For progressively heavier nitridations, the stability of the interface reaches a minimum and slowly begins to increase. At the same time for progressively heavier nitridations, the amount of positive fixed charge and the number of electron traps in the bulk of the dielectric increases. For a *heavy* enough nitridation, the amount of positive fixed charge introduced into the bulk of the dielectric reaches a maximum, with heavier nitridations resulting in less fixed charge. We call this point *turnaround*, and say that lighter nitridations than this are in the *pre-turnaround* range.

It has been found that even though all nitridation conditions in the pre-turnaround range result in an inferior dielectric compared to conventional SiO_2 (in terms of higher

fixed charge, more electron traps and reduced interface stability), reoxidizing these nitrided oxides removes the electron traps, and increases the stability of the interface to a level greater than that of the original oxide [1]. The improvement of interface stability occurs much more rapidly with reoxidation than does the removal of electron traps in the bulk of the dielectric. Furthermore, the improvement of interface stability with reoxidation reaches a maximum which is dependent only on the degree of *nitridation.* With very long reoxidations, dielectric regrowth begins to occur and the dielectric's electrical properties begin to revert back to those of conventional SiO_2. The dielectric regrowth process occurs much more slowly than the improvement in interface stability with reoxidation; this yields a broad window for the choice of reoxidation conditions to achieve the maximum improvement in interface stability for any given nitridation.

Yang et al. developed an optimization technique based on the properties of reoxidized nitrided oxides described above, using plots of degradation of interface-state density and trapped charge versus initial fixed charge to define a *trade-off region* of optimal 950°C ROXNOX performance [1]. The trade-off primarily occurs in the selection of the proper nitridation conditions to minimize generation of interface states under stress at the cost of increased fixed charge in the dielectric (and consequently lower transistor mobility). The selection of the reoxidation conditions is dictated by the requirement that the reoxidation be sufficiently heavy to remove electron traps and to provide the suppression of interface-state generation, but light enough to prevent dielectric regrowth.

Yang's method is used here to develop an optimal 850°C ROXNOX dielectric. Initial experiments with 850°C nitridations demonstrated that the trends present in previous work at 950°C [1] are also present at the process temperature of 850°C. These nitridation experiments also demonstrated that the entire pre-turnaround range

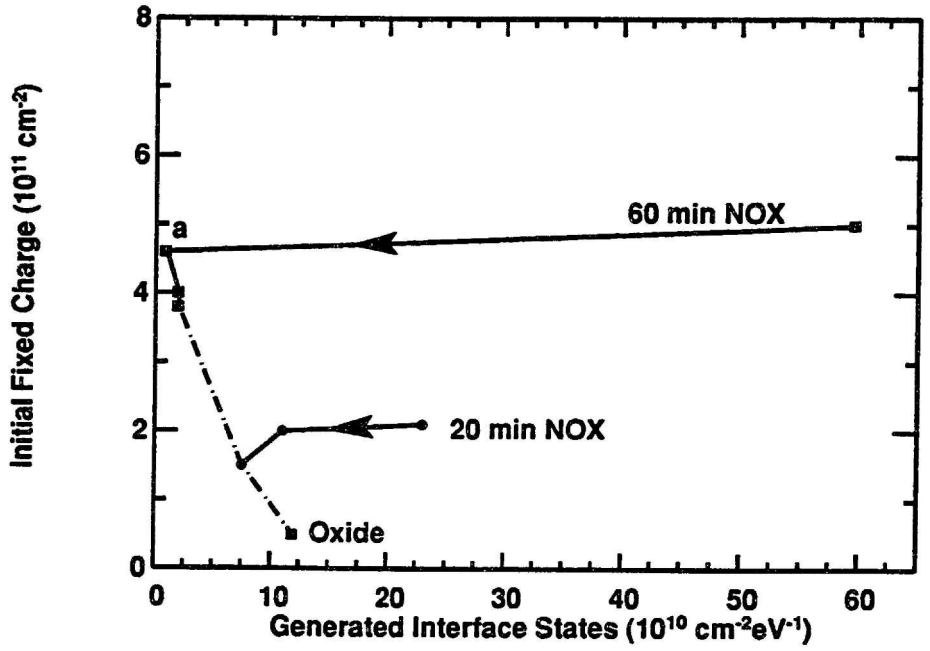


Figure 4.1: Interface-state generation *versus* initial fixed charge. NOX = 0.1 atmosphere LP NH_3 anneal of oxide. The arrows indicate increasing reoxidation time. The dashed line defines the trade-off region.

of electrical characteristics can be accessed by nitridations at 850°C by varying the nitridation pressure (up to atmospheric) and by varying the time (up to 3 hours). A representative set of optimization curves are shown in Figure 4.1 and Figure 4.2.

Interface-state generation and charge trapping were characterized under positive gate bias constant-current stress at 10 μA/cm^2 (to a final fluence of 0.1 coulomb) on 100×100 μm capacitors. Both before and after stress, quasi-static and high frequency C-V measurements were made from which midgap interface-state densities and flatband voltages were extracted using the high-low C-V method [20]. Oxide charge values were determined from the flatband voltage by assuming all the charge to be located at the Si/SiO_2 interface. The upper-rightmost point of each curve is the nitrided oxide with no reoxidation (NOX), and the arrows indicate the direction of increasing reoxidation.

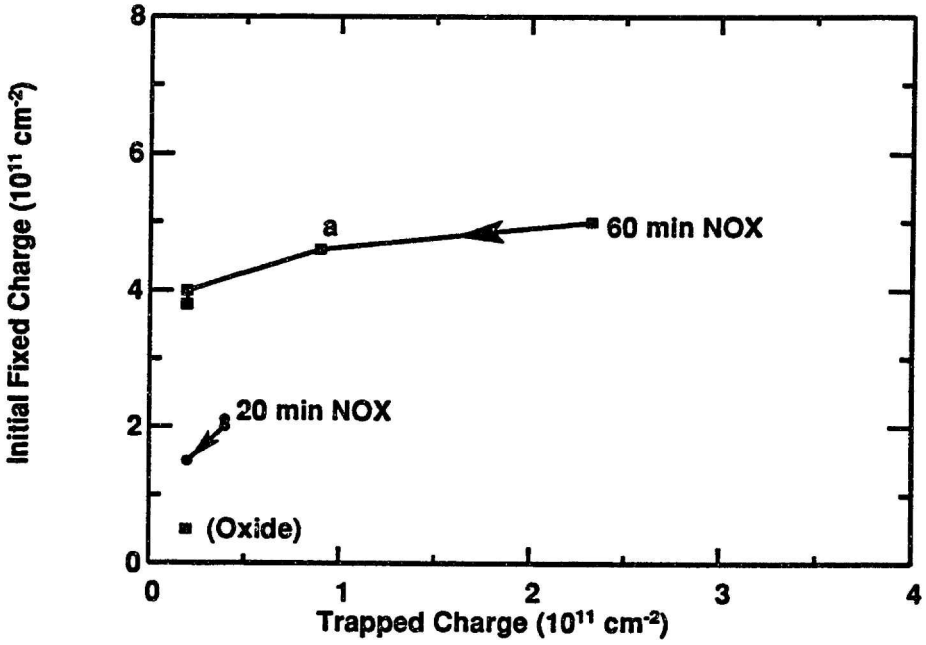


Figure 4.2: Charge trapping *versus* initial fixed charge. The oxide point is net hole trapping, others are net electron trapping. The arrows indicate increasing reoxidation time. Stress current = 10 μA/cm^2 to a fluence of 0.1 C/cm^2. T_{ox}=10 nm. Capacitor area =10^{-4} cm^2.

Figure 4.1 demonstrates that the suppression of interface-state generation with reoxidation reaches a minimum dependent on the degree of nitridation, as discussed above. Here we use time as the parameter to control the degree of nitridation, but varying temperature or NH_3 pressure produces similar results. The locus of points of minimum interface-state generation plotted vs fixed charge concentration forms the *trade-off region* mentioned above. Comparison of Figure 4.1 and Figure 4.2 demonstrates that the suppression of interface-state generation occurs much more rapidly with increasing reoxidation than the removal of electron traps (compare point 'a' in Figures 4.1 and 4.2). The x axis of Figure 4.2 represents electron trapping, except for the oxide case, where it represents trapped holes. Only one axis is used for simplicity.

The processing conditions for our optimal 850°C dielectric are shown in Table 4.1. The optimal processing conditions are those which result in the best device

NAME	CONDITIONS
CONTROL OXIDE	55 MINUTE ATM OXIDATION
NOX	1 HOUR 0.1 ATM NITRIDATION OF OX
ROXNOX	3 HOUR ATM REOXIDATION OF NOX

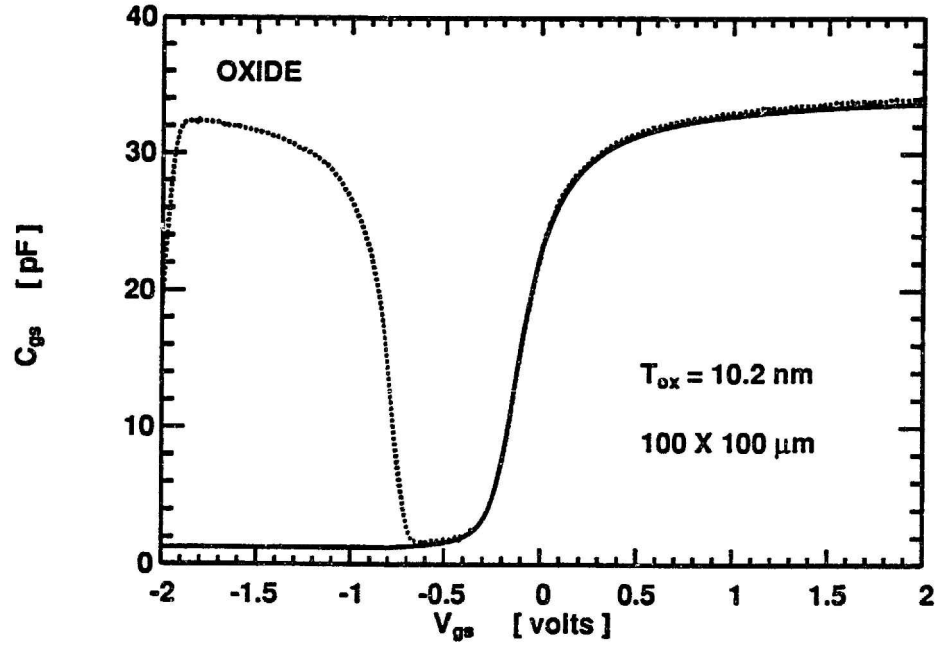
Table 4.1: Fabrication steps for optimal 850°C ROXNOX process. T=850°C for all gate processing steps. The 10 nm oxide was grown in 100% dry O_2 ambient. Nitridation ambient was 100% anhydrous NH_3 . All dielectrics received a final 30 minute N_2 anneal.

performance (lower fixed charge ⇒ higher transistor mobility) while preserving the properties of optimal reliability (minimal charge trapping and interface-state generation). A heavier nitridation than the one chosen would have resulted in a dielectric with higher initial fixed charge. A lighter nitridation than the one chosen would have resulted in a dielectric with lower initial fixed charge, but with less suppression of interface-state generation with high field stress.

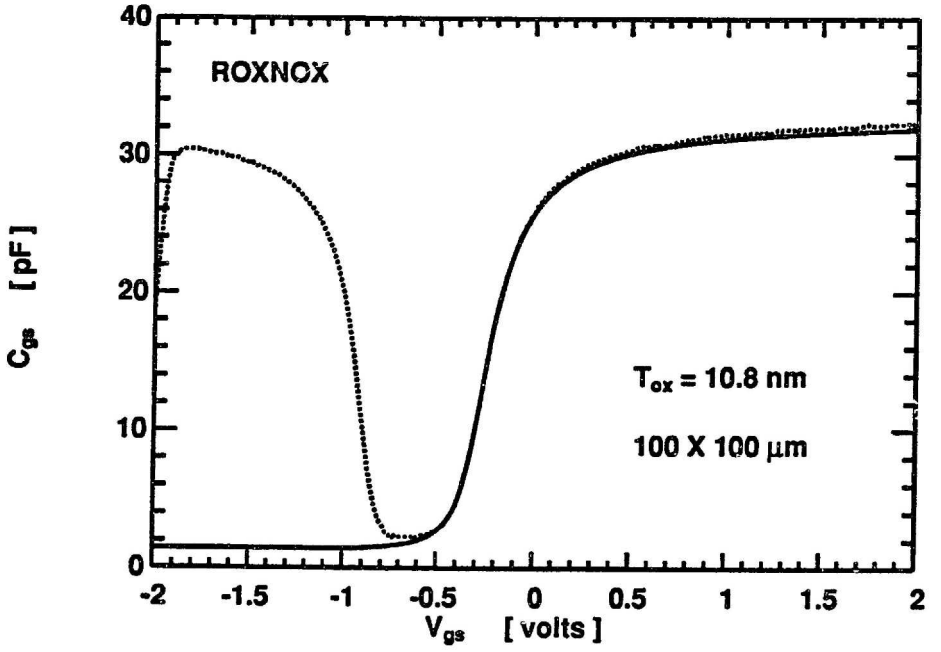
4.3 ROXNOX Results

4.3.1 Capacitor Results

High-frequency and quasi-static C-V curves for the oxide and optimized ROXNOX dielectrics are shown in Figure 4.3. The substrate of these devices was doped n-type, yielding PMOS capacitors. Note that inversion occurs at low values of gate-to-substrate voltage, and accumulation occurs at high values of gate-to-substrate voltage. The C-V characteristics of all the dielectrics reported in this thesis were well behaved.



(a) Oxide dielectric.



(b) ROXNOX dielectric.

Figure 4.3: High frequency (solid line) and quasi-static capacitance (dotted line) for (a) oxide and (b) ROXNOX dielectrics. The midgap D_{it} values extracted from these curves are approximately 1×10^{10} $cm^{-2}eV^{-1}$ for the oxide device, and 3.5×10^{10} $cm^{-2}eV^{-1}$ for the ROXNOX device. $N_d = 2\times10^{15}$ cm^{-3}.

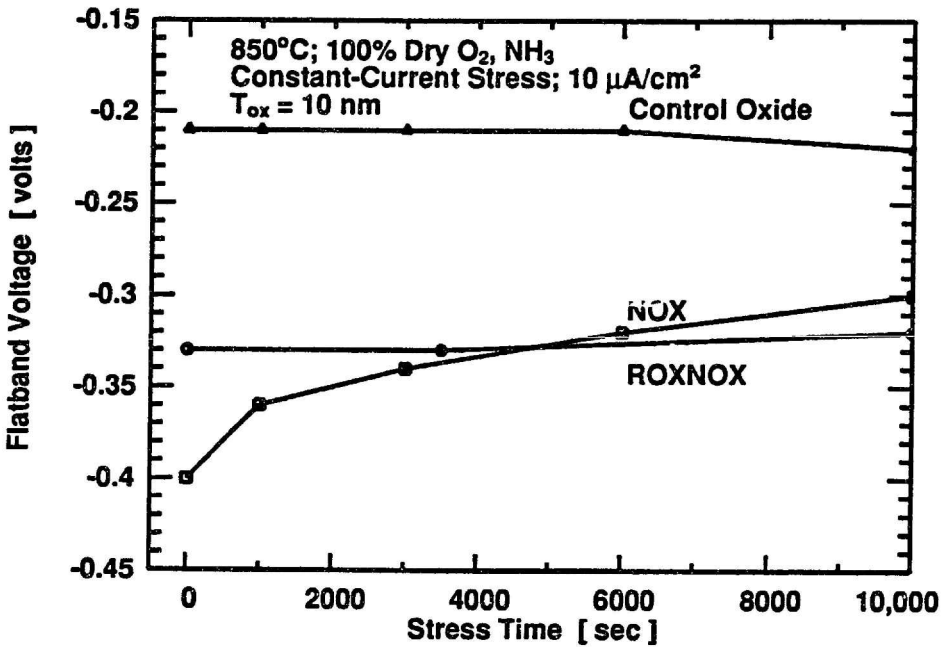


Figure 4.4: Flatband voltage vs positive gate bias constant-current stress time as extracted from C-V measurements. Stress current density = 10 $\mu A/cm^2$. Capacitor area = 10^{-4} cm^2.

Interface-state generation and charge trapping were characterized under constant-current stress as described in the previous chapter. Here, the stress was periodically interrupted to obtain quasi-static and high frequency C-V measurements from which midgap interface-state densities and flatband voltages were extracted. The results of the control oxide and the nitrided oxide without reoxidation (NOX) are presented for comparison.

As shown in Figure 4.4, pre-stress values of the flatband voltage for ROXNOX and NOX are shifted negative with respect to oxide, indicating increasing amounts of positive fixed charge with increasing nitridation. During stress, the flatband voltage for NOX shifts positively, indicating the presence of electron traps. By contrast, the reoxidation of this nitrided oxide (ROXNOX) results in the virtual elimination of

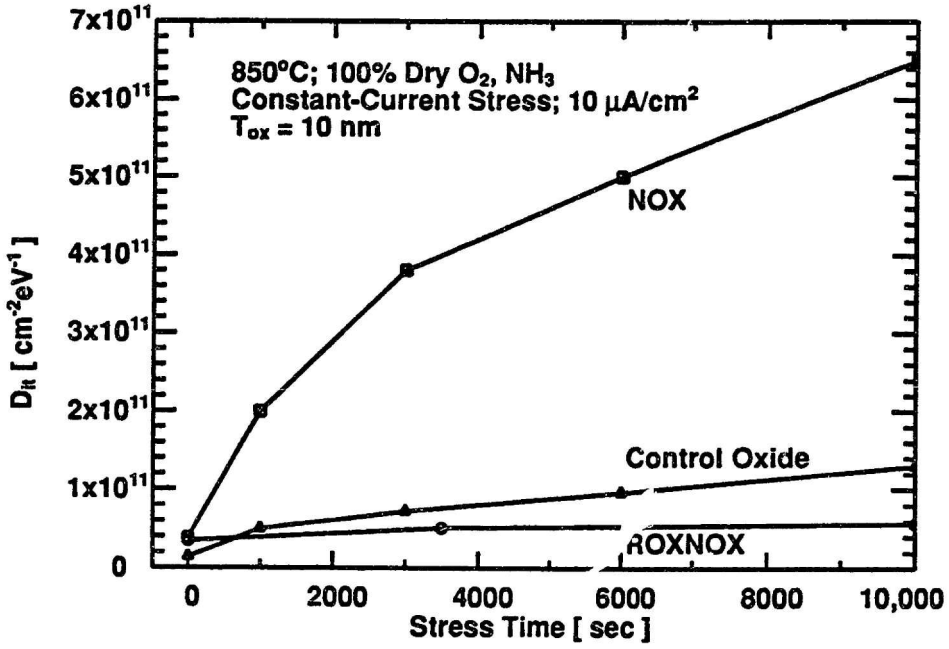


Figure 4.5: Midgap interface-state density *versus* positive gate bias constant-current stress time as extracted from C-V measurements. Stress current density = 10 μA/cm^2. Capacitor area = 10^{-4} cm^2.

electron traps, indicated by the small flatband voltage shift with stress.

In Figure 4.5, the pre-stress values of midgap interface-state density are seen to increase slightly with nitridation. Under stress, the nitrided oxide (NOX) has increased generation of interface-states as compared to the control oxide. Reoxidation of this nitrided oxide (ROXNOX in Figure 4.5) reduces the pre-stress value of interface state density to an intermediate value between NOX and oxide. However, the reoxidation also results in almost total suppression of interface-state generation with stress. It is this property of suppression of interface-state generation with stress which is a primary advantage of ROXNOX dielectrics over conventional oxide.

Results of a recent study of the reliability of various gate dielectrics, including both the 850°C ROXNOX dielectric reported here and the 950°C ROXNOX dielectric of [1], have been reported. This study demonstrates that ROXNOX dielectrics have

DIELECTRIC	NMOS	PMOS
OXIDE	540	190
ROXNOX	425	170

Table 4.2: Peak effective mobility (in units of $cm^2/V \cdot s$) for NMOS and PMOS transistors with ROXNOX dielectric as compared to conventional oxide. W/L = 70 μm/70 μm, $|V_{DS}|$ = 50 mV.

significantly better reliability for time-dependent dielectric breakdown (TDDB) [61] and better reliability under dynamic stress [62], both as compared to conventional oxide.

4.3.2 NMOS and PMOS Transistor Results

Linear Region Mobility

Inversion layer mobility measurements were made on n- and p-channel devices with a W/L ratio of 70 μm/70 μm. Mobility values were extracted from linear region I_{DS} vs V_{DS} curve for V_{DS} = 50 mV. The effective mobility is defined as

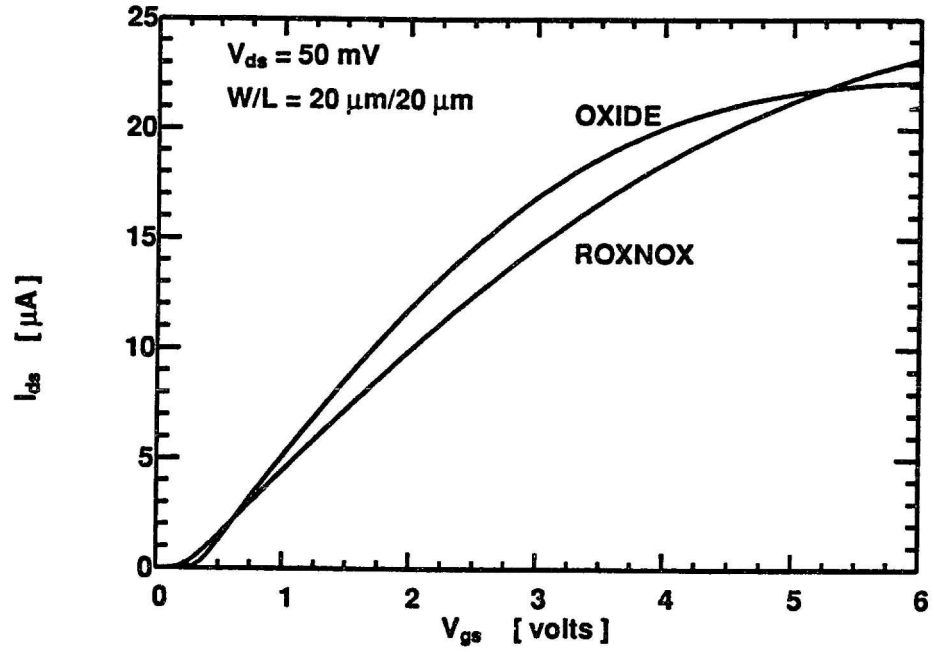
$$\mu_{eff} = \frac{I_{DS}}{\frac{W}{L} C_{ox}(V_{GS} - V_{th})V_{DS}}. \tag{4.1}$$

The peak effective mobility occurs at the point of maximum slope of the I_{DS} vs V_{GS} curve. The threshold voltage V_{th} was found by linear extrapolation of the I_{DS} vs V_{GS} curve from the maximum slope point to $I_{DS} = 0$ [36]. Note that this equation will slightly overestimate the value of channel mobility due to an underestimation of the mobile inversion charge density Q_n [63].

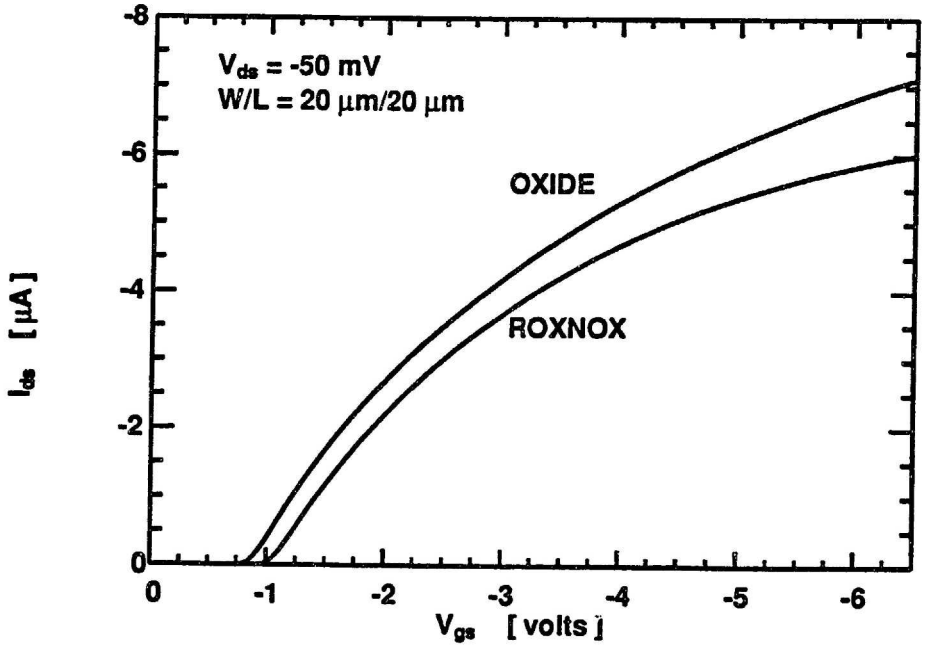
The results, summarized in Table 4.2, show that the optimized ROXNOX dielectric has reduced inversion layer mobility compared to oxide in both n- and p-channel devices. The percentage degradation, $\Delta\mu_{eff}/\mu_{eff}$, in n-channel devices, however, is more severe than in p-channel devices (21% versus 11%). It has recently been proposed [15] that the observed degradations in electron and hole mobilities can be explained by considering the combined effects of fixed charge and electron traps at the Si/SiO_2 interface. Since nitridation introduces positive fixed charge, both electron and hole mobilities are degraded by coulombic scattering. In addition, the presence of interfacial electron traps, which are believed to be located at an energy near the conduction band of Si and to be charge neutral when empty [46], further reduces the effective electron mobility as a result of reduced mobile channel charge due to trapping and coulombic scattering by the trapped inversion layer electrons. Hole mobility is not affected by these electron traps since the traps are empty and hence neutral when a p-channel device is biased in inversion.

A listing of the mean and standard deviations of the threshold voltages for the ROXNOX process for a typical CMOS wafer is given in Table 4.3. The values for control oxide are also listed for comparison. Note that the standard deviation of ROXNOX threshold voltages is comparable to the standard deviation of the conventional oxide thresholds, indicating that the furnace ROXNOX process does not contribute to device non-uniformity. The data for this table was characterized by the MTL staff on the HP4062 automatic wafer probe system as a standard post-processing step. 52 of the total of 60 dice on each wafer are routinely characterized. The remaining 8 dice are located at the corners of the wafer, and are not tested.

A plot of drain current vs gate voltage in the linear region for an NMOS transistor is shown in Figure 4.6a. An interesting feature of this curve is that the drain current for the ROXNOX device exceeds that of the oxide device at high gate bias, corre-



(a) NMOS Current.



(b) PMOS Current.

Figure 4.6: Linear region drain current of (a) oxide and (b) ROXNOX devices. Note enhanced current in NMOS ROXNOX devices at high gate field. For both NMOS and PMOS Higher Q_f of ROXNOX dielectric shifts the threshold voltage negatively, and reduces the low field mobility. Dielectric thickness is 10 nm.

Dielectric	Device Type	Drawn Size [μm]	V_{th} [volts]	σ [volts]
OXIDE	NMOS	20 × 5	0.313	0.0053
ROXNOX	NMOS	20 × 5	0.239	0.0039
OXIDE	NMOS	20 × 2	0.306	0.0052
ROXNOX	NMOS	20 × 2	0.224	0.0048
OXIDE	PMOS	20 × 5	-1.085	0.0073
ROXNOX	PMOS	20 × 5	-1.302	0.0061
OXIDE	PMOS	20 × 2	-1.067	0.0047
ROXNOX	PMOS	20 × 2	-1.269	0.0078

Table 4.3: Mean and standard deviation σ of V_{th} across a CMOS wafer. 52 dice are characterized on each wafer. The remaining 8 dice on each wafer are located at the corners, and are not characterized. Oxide T_{ox} = 9 nm. ROXNOX T_{ox} = 11 nm.

sponding to a higher effective mobility. This behavior is absent in the PMOS devices (Figure 4.6b.), which exhibit a uniformly lower drain current. It should be pointed out that this behavior of effective mobility at high gate bias for NMOS transistors has recently been reported by a number of other researchers working on rapid thermal processed nitrided oxide and reoxidized nitrided oxide gate dielectrics [53, 54, 64, 65]. The mobility model proposed in [15] does not explain this behavior at high gate bias where coulombic scattering is no longer the dominant scattering mechanism. The explanation of the mobility behavior of ROXNOX devices is beyond the scope of this thesis, and so will be left for future work.

Low Temperature Characteristics

The linear region drain current of NMOS and PMOS transistors was characterized over a range of temperatures from 80 K to 350 K and the effective mobility was

extracted using Equation (4.1). A plot of the peak effective mobility is shown in Figure 4.7. Note that the temperature dependence of the oxide mobility is close to the $T^{-1.5}$ dependence commonly assumed for inversion layer mobility.

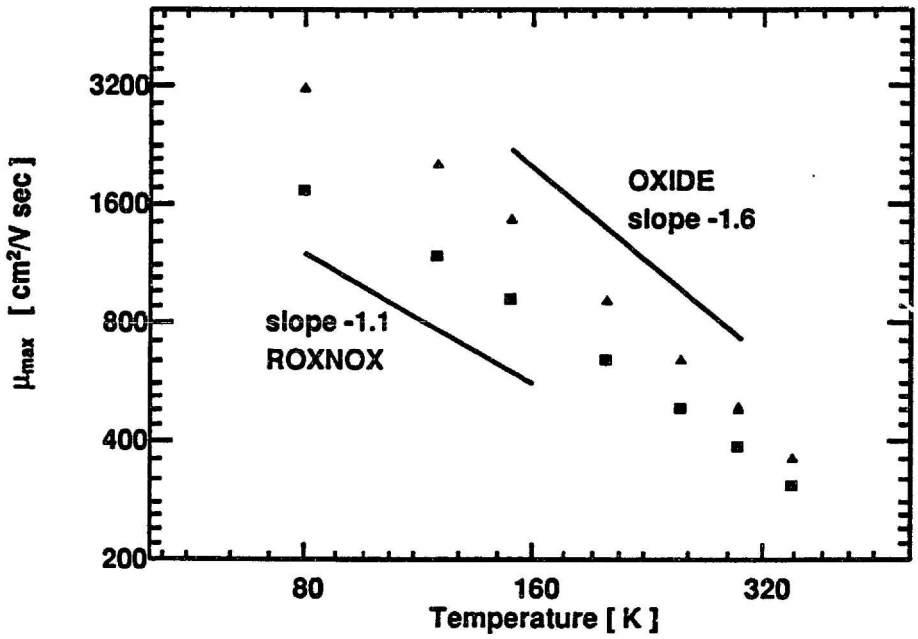
The behavior of the effective mobility *versus* effective normal field for these devices was extracted from the linear region drain current of these devices using Equation (4.1). The inversion layer carrier density was assumed to fit the relation

$$Q_{inv} = C_{ox}(V_{gs} - V_{th} - \frac{V_{ds}}{2}). \tag{4.2}$$

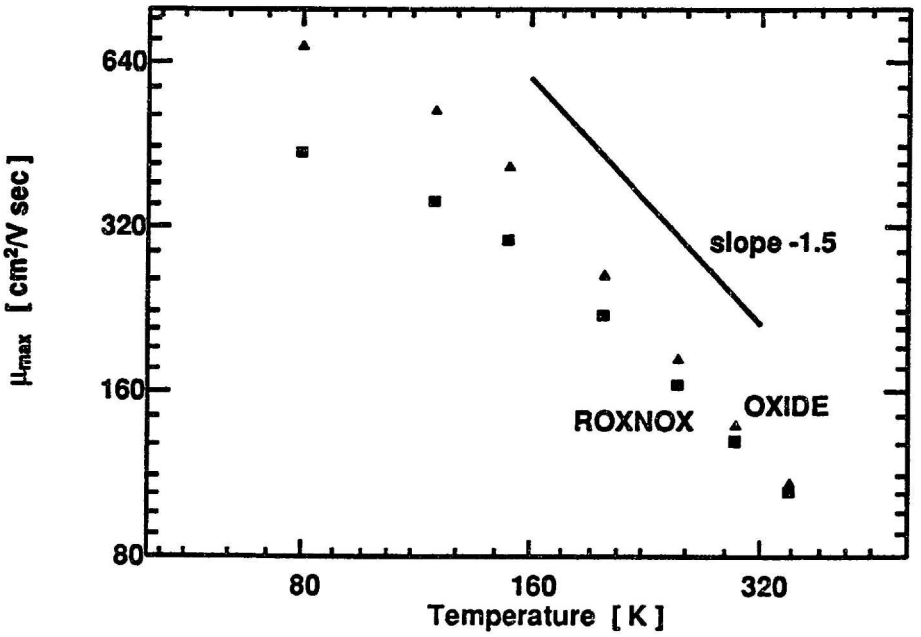
Note that this estimate will introduce a small error into the extraction of effective mobility. This error will be largest for low gate biases, and becomes negligible at biases well above threshold [63]. The average normal field affecting the inversion layer carriers can be written as

$$E_{eff} = \frac{1}{\epsilon_{Si}}(Q_B + \eta Q_{inv}), \tag{4.3}$$

where Q_B is the integrated depletion charge, and η is an empirical constant which has the value of 0.5 for NMOS devices, and 0.33 for PMOS devices [36]. The same value was assumed for Q_B for both oxide and ROXNOX. The difference in thermal budget for the dielectric growth step is assumed to produce a negligible difference in integrated depletion charge. Plots of effective mobility *versus* average normal field are shown in Figure 4.8. The oxide curves are shown by solid lines, and the ROXNOX characteristic is shown by the dashed line. Note that the effective mobility at high average normal field is higher for ROXNOX devices than for the oxide devices, as expected from the linear region drain current results, Figure 4.6, and as has been reported by other investigators [53, 54, 64, 65]. This behavior is not observed for the PMOS devices. Note also that at low temperatures, the difference between the ROXNOX and oxide effective mobility becomes greater at low fields. This could be

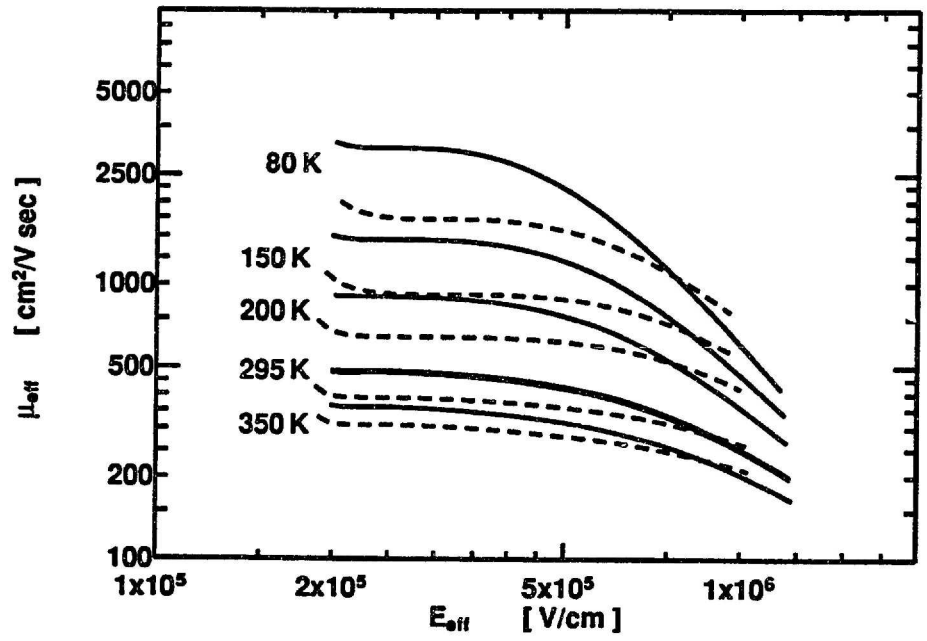


(a) NMOS Mobility.

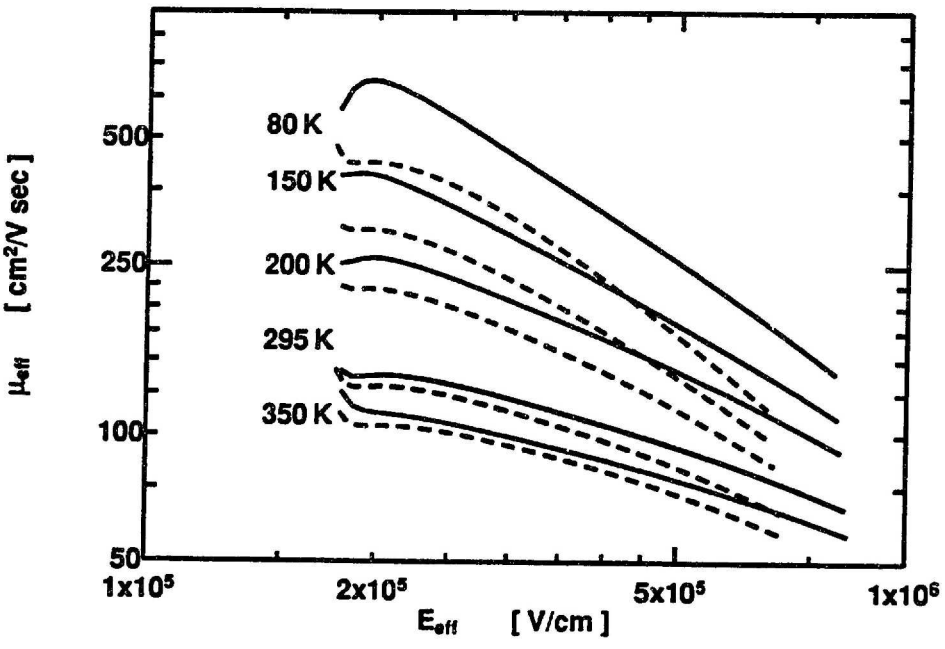


(b) PMOS Mobility.

Figure 4.7: Peak effective mobility *versus* temperature for (a) NMOS and (b) PMOS devices. W/L=70 μm/70 μm. Oxide T_{ox}=9 nm. ROXNOX T_{ox}=11 nm.



(a) NMOS



(b) PMOS

Figure 4.8: Plot of effective mobility (μ_{eff}) vs effective normal electric field for (a) NMOS and (b) PMOS devices. Oxide curves are solid lines. Corresponding ROXNOX curves shown by dashed lines.

due to scattering from fixed charge in the oxide being relatively more important at low temperatures [66]. The ROXNOX dielectric exhibits a higher density of positive fixed charge.

4.3.3 $1/f$ Noise Results

$1/f$ Noise was measured for both NMOS and PMOS transistors. The results of the NMOS noise measurements are shown in Figure 4.9a. Note that the input-referred noise of the ROXNOX dielectric is somewhat higher than that of the control oxide. Also note that the input-referred $1/f$ noise data can be fit to a curve

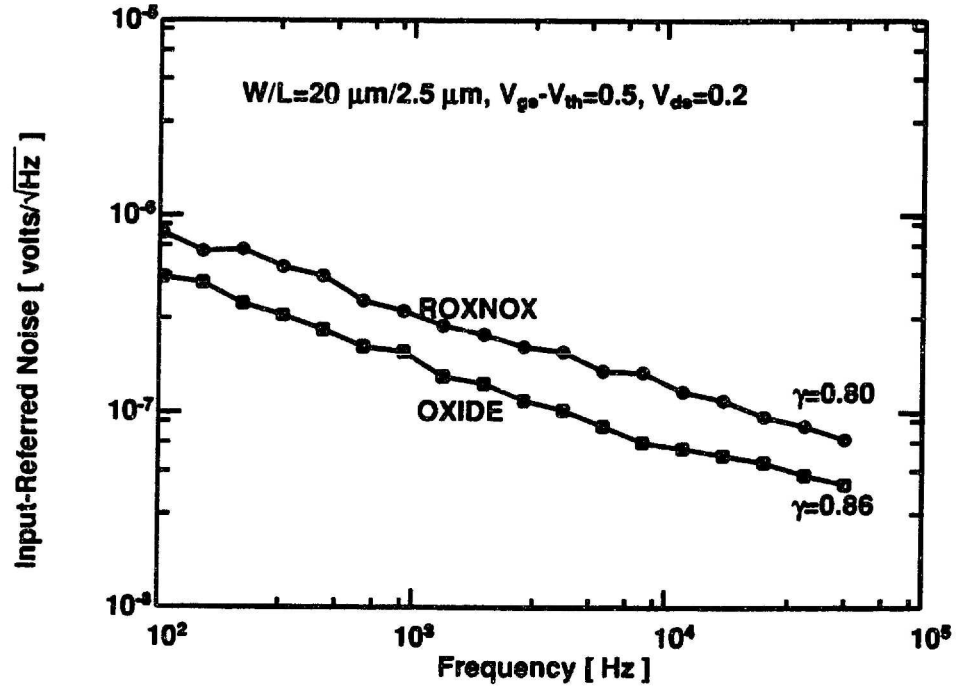
$$S_{VG}(f) = \frac{A}{f^{\gamma}} \tag{4.4}$$

where $S_{VG}(f)$ is the power spectral density of the input-referred $1/f$ noise, A is a constant dependent on device parameters, and γ is a constant which is less than 1.0 for the ROXNOX data. The higher value of input-referred noise voltage for the ROXNOX devices indicates the presence of a greater number of electron traps in the dielectric near the interface, and with energy levels near the conduction band of the silicon. The γ value less than 1.0 is due to non-uniformity of the trap density in the oxide [67], and will be discussed further in Chapter 6.

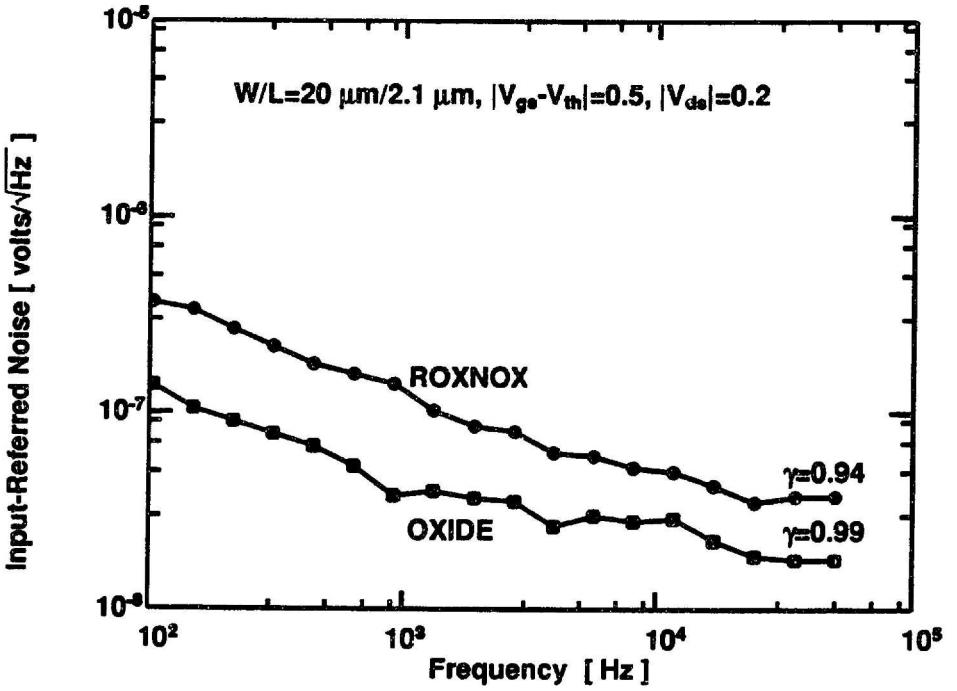
The $1/f$ noise results for the PMOS devices are shown in Figure 4.9b. Note that the noise value for the ROXNOX devices is higher than for the oxide devices. The flattening of the noise curves at high frequency is an artifact due to the noise at those frequencies being comparable to the noise of the measurement system. Further results of $1/f$ noise measurements are found in Chapter 6.

4.3.4 Channel Hot-Electron Stressing Results

NMOS transistors with an effective channel length L_{eff} = 0.9-1.5 μm, a width W = 20 μm, and gate dielectric thickness T_{ox} = 10 nm were used for channel hot-



(a) NMOS noise.



(b) PMOS noise.

Figure 4.9: Input-referred noise ($<v_{gs}^2>^{\frac{1}{2}}$) of oxide and ROXNOX devices for (a) NMOS and (b) PMOS devices. Dielectric thickness is 10 nm.

electron stressing experiments. Due to differences in inversion layer mobilities and in effective channel length, the drain voltage during stress was modified to obtain similar stressing conditions for each device [68]. Thus each device was stressed at a gate voltage corresponding to the peak in substrate current for a range of drain voltages (6.0, 6.5, and 7.5 V). The duration of the stress was 2 hours. The transfer characteristics at $V_{DS} = 0.1$ V were taken before and after stress and the relative change in the maximum linear region transconductance $\Delta g_m/g_{m0}$ was taken as a measure of the degradation. Threshold shifts ΔV_{th} with source-drain leads in forward and reverse configurations were negligible in these experiments.

The results of the stressing experiments are shown in Figure 4.10. Degradation in ROXNOX devices is considerably less than in oxide devices over the range of drain voltages tested, verifying that the reduced interface-state generation and charge trapping measured in capacitors translates to greater resistance to CHE stressing measured in transistors.

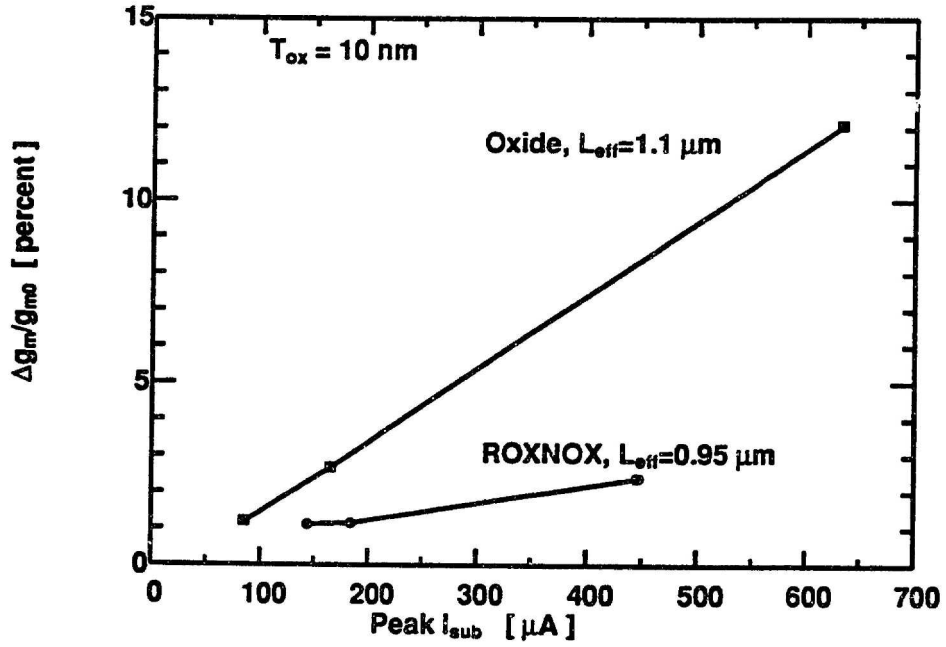


Figure 4.10: Linear region transconductance degradation in NMOS oxide and ROX-NOX devices. The devices were biased at the gate voltage corresponding to the peak in substrate current for a range of drain voltages (6 to 7.5 V). Stress time = 2 hours.

Chapter 5

Single-Electron Traps

5.1 Introduction

If the channel area of a MOSFET is made small enough, a very small number of traps is present to cause noise in the device characteristic. Likewise, keeping the ratio of channel length to channel width the same, the channel resistance will remain constant to first order. Hence, as the channel area is reduced, the current flowing in the MOSFET will be carried by fewer and fewer channel carriers. For devices with channel length and channel width on the order of 1 μm each, it becomes likely that there may be only a single trap active in the channel at one time. For these devices, the trapping and de-trapping from this trap can cause a large percentage change (0.01% or larger) in the drain current. Devices of these dimensions have been fabricated, and produce a drain current which discretely switches between two states as a single trap captures and emits. The discrete switching characteristic is commonly known as a random-telegraph signal (RTS) [25].

Drain current traces from a typical single-electron trap are shown in Figure 5.1. The change in drain current is plotted versus the measurement time. The three traces are all from a single device, taken at three values of gate bias. We interpret the data to demonstrate that the trap has captured an electron when in the low state and has

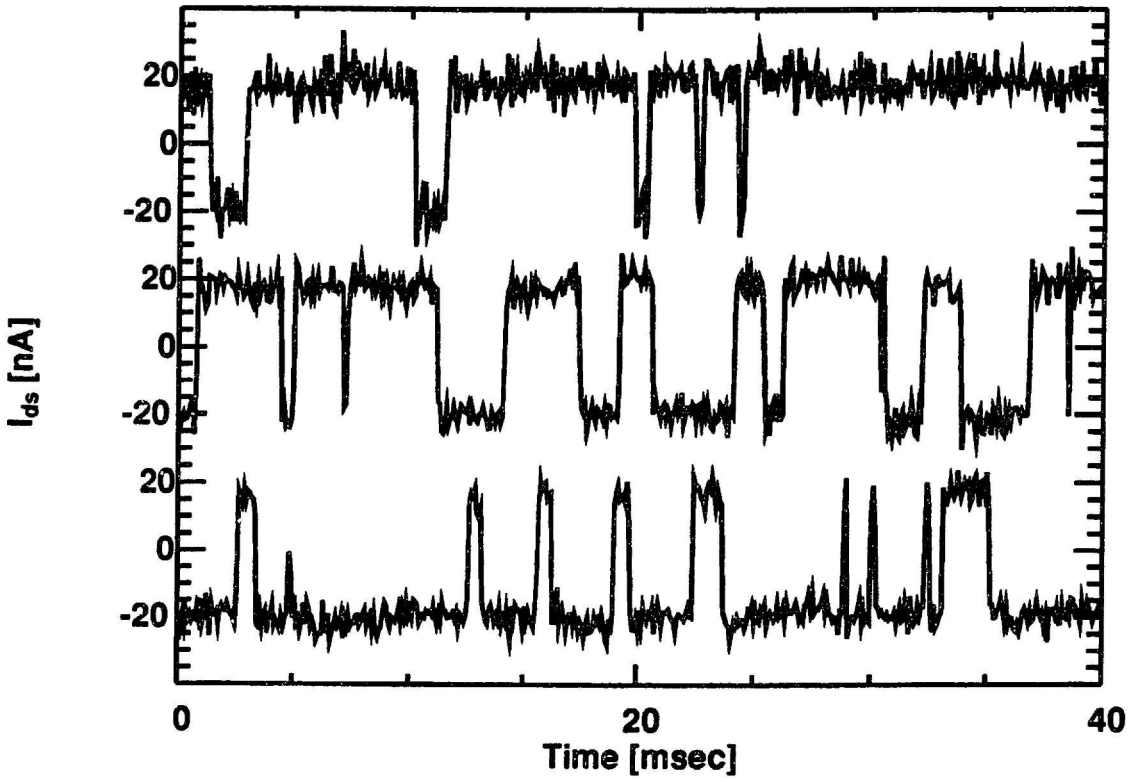


Figure 5.1: Typical random-telegraph noise signal. Top trace: V_{gs} = 1.4 V. Middle trace: V_{gs} = 1.6 V. Bottom trace: V_{gs} = 1.8 V. V_{ds} = 0.2 V for all traces.

emitted an electron when in the high state. As the gate bias is increased, the energy of the trap in the oxide is lowered relative to the Fermi level in the silicon substrate and so the trap is full more of the time, as can be seen from the bottom trace of Figure 5.1.

Recently, there has been a strong interest in the study of single-electron trapping in very small geometry MOSFETs. Ralls et al. were the first to report the observation of single trapping events in MOSFETs with W/L=0.1 μm/1.0 μm. They used metal-gate transistors and characterized the switching in channel resistance by using a lock-in amplifier system. All of the measurements reported by this group were performed at temperatures well below room temperature [24].

Since then a great deal of work has been reported by Kirton et al. (An excellent

review of their results can be found in [25].) Kirton et al. have used a number of different techniques for characterizing RTS. They were the first to report that RTS's interarrival times possess a exponential probability distribution. They were also the first to report the strong temperature activation observed for the capture and emission processes, as well as reporting results of traps which are active at room temperature [25]. Kirton et al. model the capture and emission processes as occurring through a multiphonon emission process. This model will be discussed more thoroughly in Section 5.4.2.

Recent work by Hung et al. has sought to characterize the effect of a single trap on the mobility of a deep-submicron MOSFET at room temperature [26]. Another report by the same group presented results of hot-electron stressing of these small devices. This report demonstrated that single-electron traps were generated by channel hot-electron stressing [69]. Other workers have reported on the relationship between random-telegraph noise and quantum-transport in deep-submicron devices operating at low temperatures [27]. Recent work on the amplitude of the random telegraph signals has been presented in [70]. These results will be discussed in the trap amplitude section, 5.2.4.

In this chapter, results of the characterization of single-electron traps in deep-submicron MOSFETs will be presented. First, measured results will be presented on devices with both conventional oxide and reoxidized nitrided oxide dielectrics. Then, following a brief discussion of a model of the inversion layer using the physics of quantized subbands, a discussion of current models of single-electron-trap behavior will be presented. It is shown that modeling the capture and emission processes using a multiphonon emission mechanism explains the behavior of trap time constant *versus* bias and temperature very well. A model of single-electron-trap characteristics is presented which will be used in Chapter 6 to model the ensemble of traps responsible

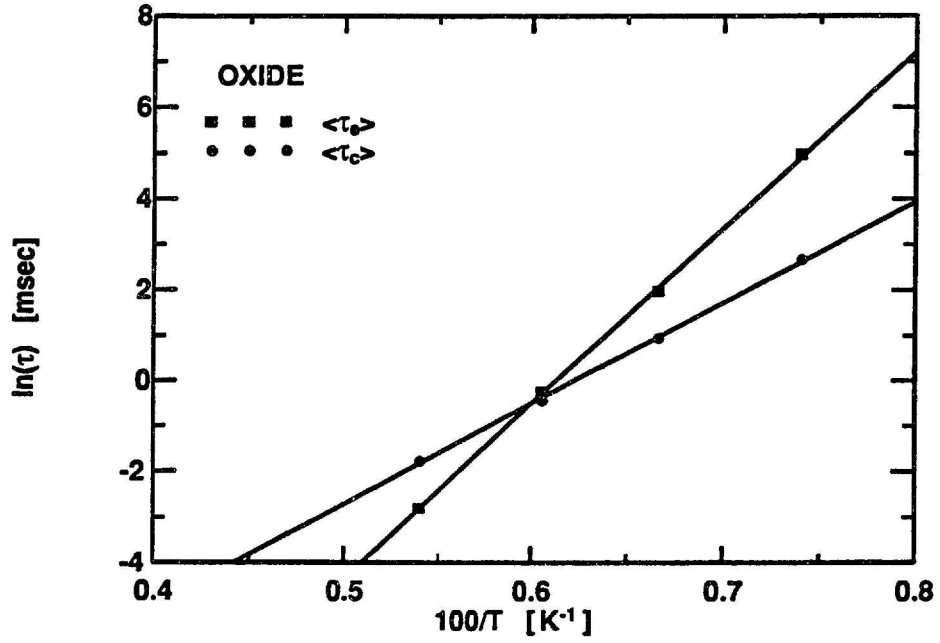
for $1/f$ noise in MOSFETs.

5.2 Single-Electron-Trap Measurements

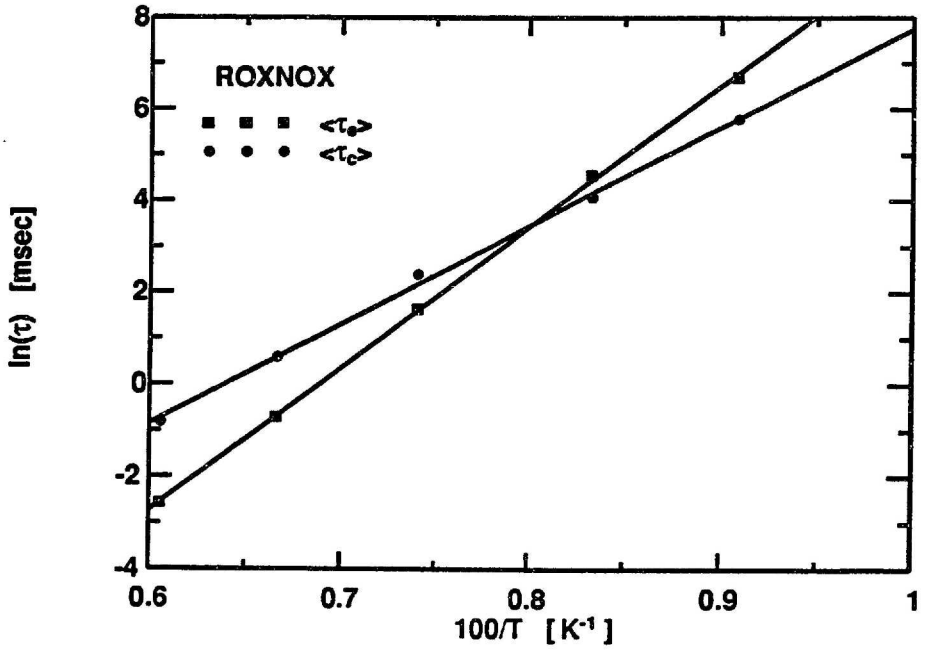
5.2.1 Capture and Emission Time

The mean capture time $<\tau_c>$ and the mean emission time $<\tau_e>$ for a given trap are extracted from the RTS signal-vs-time data capture using a HP200 series computer. On the order of 100 transitions were captured in a single trace to compute the mean capture and emission times. A threshold value halfway between the high and low state amplitudes is used to determine the mean amount of time spent in the high and the low states. The mean time the trap spends in the low state is interpreted to be equal to the mean emission time, and the mean time the trap spends in the high state is equal to the mean capture time. This interpretation is justified since trapping of an electron from the conduction band is expected to reduce drain current; hence the mean time in the low state is the mean time it takes for the full trap to emit an electron, i.e. the mean emission time. The strongest argument in favor of this interpretation is the behavior of the trap occupancy under applied bias, as seen in Figure 5.1, and as will be discussed near the end of this section.

In Figure 5.2a. is shown an Arrhenius plot of average capture and emission times for a conventional oxide trap. Both the capture and emission times show a very strongly temperature activated behavior, as has been reported by other authors [25]. In Figure 5.2b. is shown an Arrhenius plot for a typical reoxidized nitrided oxide trap. The traps for this dielectric show a similar temperature activation behavior to that demonstrated by the oxide traps. It is worth noting that the density of traps in the reoxidized nitrided oxide devices was larger ($\approx 5\times$) than in the oxide devices, as expected from the higher $1/f$ noise exhibited by this dielectric [71]. The higher tra-



(a) Oxide dielectric.



(b) ROXNOX dielectric.

Figure 5.2: Arrhenius plot of capture and emission times for a typical single-electron trap.

density complicated the measurement because devices which had only one trap active at a given bias, temperature, and device size were much harder to find. The problem was alleviated somewhat by measuring ROXNOX devices with smaller channel areas.

Detailed balance applies to the steady-state occupancy of the trap. A quasi-Fermi level $\overline{\mu_t}$ can be defined to describe trap occupancy. The trap quasi-Fermi level is based on the observation that in steady state, the trap will be full for a constant percentage of the time. Stated another way, observing the trap at any instant, there is a constant percentage chance that the trap will be full. The detailed balance condition can then be written as [25],

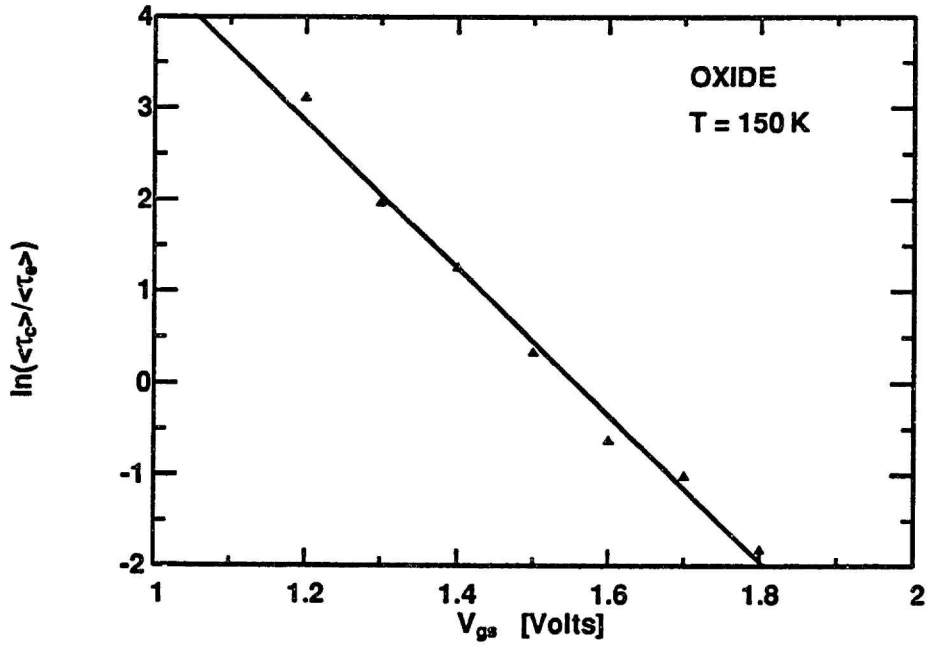
$$\frac{<\tau_c>}{<\tau_e>} = g\, e^{-\frac{(\overline{\mu_t}-E_t)}{kT}} \tag{5.1}$$

where g is the trap degeneracy factor, and E_t is the trap energy. A plot of the logarithm of the ratio of capture time to emission time for the traps shown in Figure 5.2 is shown in Figure 5.3. The ratio of the capture and emission times fits the form of Equation (5.1) very well. The negative slope of the curve in Figure 5.3 provides a *a posteriori* justification for calling the mean time spent in the low state the mean emission time, and the mean time spent in the high state the mean capture time. As the value of gate bias is increased, for example, the trap is full more of the time, and hence its emission time is increasingly longer than its capture time. It is expected that the curves in Figure 5.3 would have a positive slope if the opposite interpretation were true.

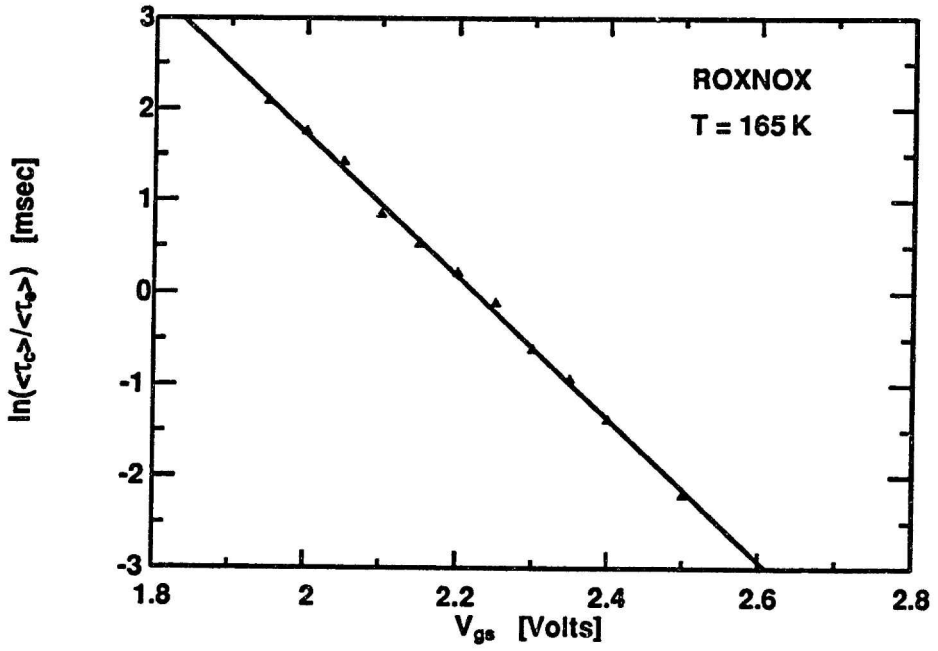
An ideal Poisson process of rate λ has an exponential distribution of transition times. For a given Poisson process at any given time, we can write the probability density function for the time until the next transition as [72]

$$\Pr(t) = \lambda e^{-\lambda t}. \tag{5.2}$$

Alternatively, Equation (5.2) can be thought of as the probability density that the



(a) Oxide dielectric.



(b) ROXNOX dielectric.

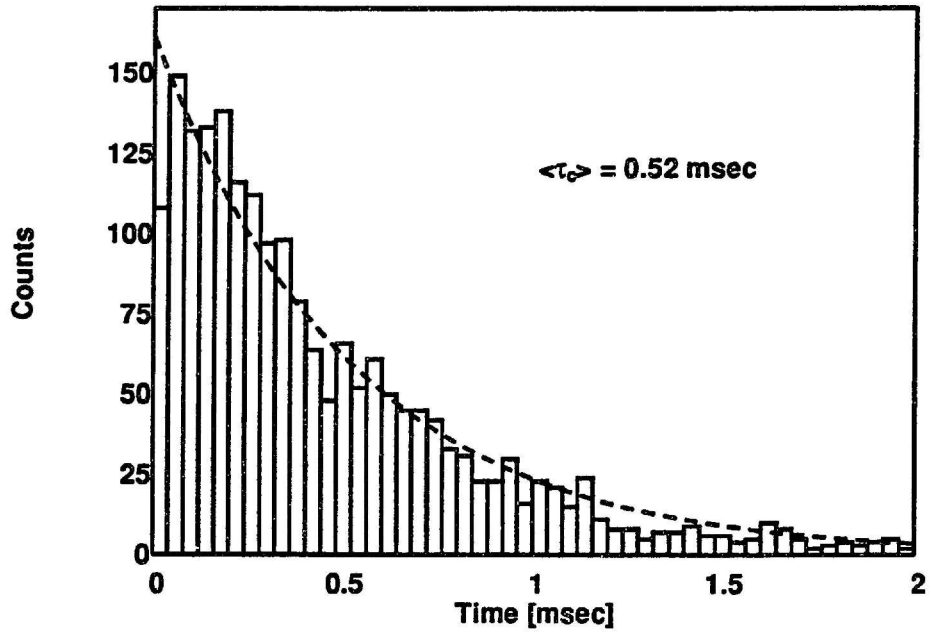
Figure 5.3: Ratio of capture time to emission time plotted vs V_{gs}. Note that the capture time grows increasingly short at high bias, as the trap spends more of the time in the full state.

Poisson process will remain in a given state for an amount of time t. The mean transition time, or the mean time spent in one state, for this process is simply λ^{-1} [42]. A histogram of the occurrences of specific capture and emission times for a typical single trap is shown in Figure 5.4. These histograms correspond to the probability distribution of the capture and emission processes. The dashed line in the figure is calculated from Equation (5.2) using the mean capture and emission times to determine estimates for the rate constants,

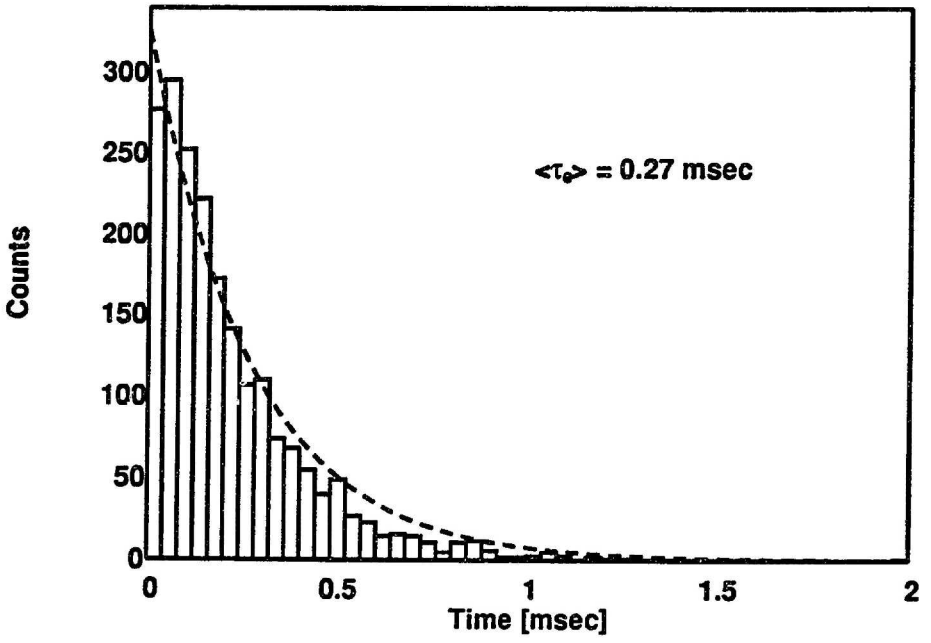
$$\lambda_{est} = \frac{1}{<\tau>}. \tag{5.3}$$

From the exponential characteristic of these histograms, it is apparent that the capture and emission processes for a single trap can be modeled very well as Poisson processes. This is an important result, because the Poisson random process is a simple and well understood process from a theoretical point of view. To be rigorous, it is necessary that the process be independent from one time interval to the next in order to be modeled as an ideal Poisson process. Examination of the drain-current *versus* time characteristic (Figure 5.1 for example), indicates that this is a reasonable assumption.

Note that both of the histograms in Figure 5.4 have a dip near $t = 0$. The high-frequency pole of the measurement system acts as a filter to smooth out the fastest transitions of the random telegraph signal. For this reason, the number of fast transitions is slightly under-counted, resulting in the dip in the histogram. The under-counting of the fast transitions will also result in an error in the estimate of mean capture and mean emission times. For the histograms shown, this error is estimated to be less than 10%.



(a) Capture time distribution.



(b) Emission time distribution.

Figure 5.4: Histogram of capture and emission times. The dashed line is the expected distribution given the values of mean capture time and mean emission time. Note the dip in the characteristic near $t = 0$ due to the high-frequency cutoff of the measurement system.

5.2.2 Functional Form

The variation of the capture and emission time for all of the single-electron traps were very well behaved. The emission time is well described by the equation

$$<\tau_e>= [e(x, V_{gs})]^{-1} \, e^{\frac{\overline{\mu_t}-E_t}{kT}} \, e^{\frac{E_g}{kT}}, \tag{5.4}$$

where $e(x, V_{gs})$ is a generalized emission rate parameter, and $\overline{\mu_t}$ is the trap quasi-Fermi level. For the traps characterized in this work, the trap quasi-Fermi level is within a few kT of the trap energy. Making use of the detail balance condition of Equation (5.1), we can write the capture time as

$$<\tau_c>= [c(x, V_{gs})]^{-1} \, e^{\frac{E_g}{kT}}, \tag{5.5}$$

where $c(x, V_{gs})$ is the generalized capture rate parameter, which satisfies the relation

$$c(x, V_{gs}) = \frac{e(x, V_{gs})}{g}. \tag{5.6}$$

This behavior is evident from Figures 5.3, which demonstrates that both the capture and emission times have a large activation energy, and from Figure 5.1 which demonstrates that the ratio of capture to emission times is proportional to a Fermi factor. Note that this functional form is similar to the form of the model proposed by Kirton et al. [25].

As the gate bias is varied, the effective trap energy changes due to the bending of the oxide bands. The variation of E_t with gate bias with the MOSFET biased in strong inversion can be described to first order by

$$\frac{\partial E_t}{\partial V_{gs}} = -q\frac{x_t}{t_{ox}}, \tag{5.7}$$

where x_t is the distance the trap is away from the silicon/dielectric interface, t_{ox} is the dielectric thickness, and q is the unit electron charge. The distance the trap is

away from the Si/SiO_2 interface can be determined from plots like Figure 5.1 of the logarithm of the ratio of mean capture to mean emission times vs V_{gs}. If we assume for the moment that the change in trap quasi-Fermi level with respect to gate bias is negligible, the slope of this curve is equal to

$$\frac{\partial ln\left(\frac{<\tau_c>}{<\tau_e>}\right)}{\partial V_{gs}} \simeq \frac{1}{kT}\frac{\partial E_t}{\partial V_{gs}} = -\frac{q}{kT}\frac{x_t}{t_{ox}}. \tag{5.8}$$

The higher the gate bias, the better the approximation that the trap quasi-Fermi level is pinned at a constant value. This approximation will be reevaluated in the context of the inversion layer simulation discussed in Section 5.3.

The above equations describe most of the trends present in the data. For the traps characterized in this work, the generalized rate parameters $e(x, V_{gs})$ and $c(x, V_{gs})$ are seen to be much stronger function of gate bias than expected from any of the models in the literature [25]. Discussion of this result is deferred to Section 5.4.3.

In Table 5.1 are listed the distance from the Si/SiO_2 interface, x_t, the trap activation energy E_a, and the trap energy E_t for a number of traps found in conventional oxide and reoxidized nitrided oxide. The values in this table are comparable to the trap parameters reported by other authors [25 - 27]. A discussion of the model used to explain the functional form of the data is given in Section 5.4.

5.2.3 Noise Spectrum of Single Traps

Noise Power of Single Trap

The low frequency noise power of a single trap can be readily derived from first principles. We choose a very low frequency to analyze, so that we can insure that the trap will fill and empty an arbitrarily large number of times. When the trap fills and empties much more rapidly than the frequency of interest, the details of the filling and emptying are unimportant.

I.D.	Device Type	W/L (μm)	Temp. (K)	V_{gs} Range (V)	E_a (meV)	x_t (Å)
A	OX	1.6/0.3	80-100	1.0-1.6	93.1	7.5
B	OX	0.8/0.4	120-185	1.1-1.7	285	9.8
C	OX	0.8/0.5	150-200	0.7-1.2	389	16.0
D	OX	0.8/0.4	150-200	0.9-1.8	328	14.3
E	OX	1.1/0.3	250-310	2.2-3.0	616	9.8
F	OX	0.8/0.5	250-340	1.0-1.7	636	16.9
G	ROXNOX	0.75/0.3	110-165	1.8-2.5	241	11.2
H	ROXNOX	0.55/0.3	150-200	1.9-2.4	371	23.5
I	ROXNOX	0.55/0.3	165-200	1.0-1.3	459	41.9
J	ROXNOX	0.75/0.3	200-250	0.5-0.8	570	34.8
K	ROXNOX	0.55/0.3	215-295	0.2-0.7	561	23.9
L	ROXNOX	0.75/0.3	310-355	0.6-1.2	755	28.5

Table 5.1: Model parameters extracted from RTS data

Consider a trap which causes the current in an arbitrary device to be +A/2 when full and -A/2 when empty. Furthermore assume that the trap is on average full a fraction f of the time (and empty for a fraction $1-f$ of the time). The mean value (dc component) of the current in our device is found to be

$$\frac{A}{2}f - \frac{A}{2}(1-f) = \frac{A}{2}(2f-1). \tag{5.9}$$

The current noise power is obtained by subtracting the average value of the device current from the actual value and squaring the result. When the trap is full, the noise power is

$$P_{full} = \left\{\frac{A}{2}[1-(2f-1)]\right\}^2 = [A(1-f)]^2. \tag{5.10}$$

Likewise, the amount of noise power when the trap is empty is

$$P_{empty} = \left\{\frac{A}{2}[-1-(2f-1)]\right\}^2 = (Af)^2. \tag{5.11}$$

The average noise power for each of the two states above is simply the noise power in that state times the fraction of the time the device is in that state. The total average current noise power of the device is simply the sum of the average noise power when

the trap is full plus the amount when empty,

$$P_{noise} = fP_{full} + (1 - f)P_{empty}, \tag{5.12}$$

which, using the above equations, is equal to

$$P_{noise} = A^2 f(1 - f). \tag{5.13}$$

Note that the noise power above is an estimate of the total noise power over the entire spectrum. The noise power spectral density is proportional to this value. An interesting result of this development is that the traps which possess the highest noise power are those which are full approximately half of the time. Traps which are nearly always full, or nearly always empty, contribute a negligible amount of noise power.

In general, when the occupancy of the trap is described by a Fermi function, the value of f in the above equation is simply the Fermi function for the trap. This is the same as the result reported by Sah [73] for modeling the noise contributed by individual traps.

Noise Spectrum of Single Trap

The autocorrelation function for a random process is simply a relation of how the state of a system at a given time relates to the state of the system at some earlier or later time. A random telegraph signal whose transitions are modeled as ideal Poisson processes has an autocorrelation function $R(\tau)$ of the form

$$R(\tau) = f(1 - f)A^2 e^{-(\lambda_h + \lambda_l)|\tau|}, \tag{5.14}$$

where A is the trap amplitude, f is the fraction of the time in the low state, λ_h is the rate from the high state to the low state, and λ_l is the rate from the low state to the high state. A derivation of the autocorrelation function for a random telegraph noise

process is given in Appendix D. We can rewrite Equation 5.14 in terms of the time constants as

$$R(\tau) = f(1-f)A^2 e^{-(\frac{1}{\tau_h}+\frac{1}{\tau_l})|\tau|}, \tag{5.15}$$

making use of the relations

$$\tau_h = \frac{1}{\lambda_h},$$
$$\tau_l = \frac{1}{\lambda_l}. \tag{5.16}$$

Intuitively, Equation (5.15) tells us that knowing the state of the system at a given time, the state of the system at some later time τ will likely be the same as long as

$$\tau \ll \frac{1}{\lambda_l + \lambda_h}. \tag{5.17}$$

By contrast, a white noise process has an autocorrelation function which is a unit impulse,

$$R_{white}(\tau) = \delta(\tau), \tag{5.18}$$

so that knowledge of the state of the noise process at any given time gives no information about the noise at any other time.

The noise power spectral density for a random process can be calculated from the autocorrelation function for that random process by [42]

$$S(\omega) = \int_{-\infty}^{+\infty} R(\tau) e^{-j\omega t} d\tau. \tag{5.19}$$

For the autocorrelation function (5.15), the noise spectrum of the random-telegraph signal is calculated to be

$$S_{RTS}(\omega) = A^2 f(1-f) \frac{2\frac{1}{\lambda_h+\lambda_l}}{1+\left(\frac{\omega}{\lambda_h+\lambda_l}\right)^2}, \qquad -\infty \leq \omega \leq +\infty. \tag{5.20}$$

Note that this quantity is the two-sided spectral density, existing for both positive and negative frequencies. The actual measurements of noise spectral density do not

distinguish between positive and negative frequencies, so that the measured RTS noise power spectrum is two times the above value,

$$S_{RTS}(\omega) = A^2 f(1-f)\frac{4\frac{1}{\lambda_h+\lambda_l}}{1+\left(\frac{\omega}{\lambda_h+\lambda_l}\right)^2}, \qquad 0 \leq \omega \leq +\infty. \tag{5.21}$$

The Wiener-Khinchin theorem states that the spectrum of a sample of a noise process will approach the spectrum of the true noise process [calculated from Equation (5.19) above] as long as the sample time is long enough [74]. This theorem allows us to estimate the noise power of a random process by measurement over a finite length of time. The measured noise power spectral density for a typical single trap is shown in Figure 5.5a. A sample of the time domain data is shown in Figure 5.5b. The dotted line is Equation (5.21) using the values of the parameters shown, demonstrating good agreement between the theory and the measured data. All of the traps characterized have spectra which fit Equation (5.21).

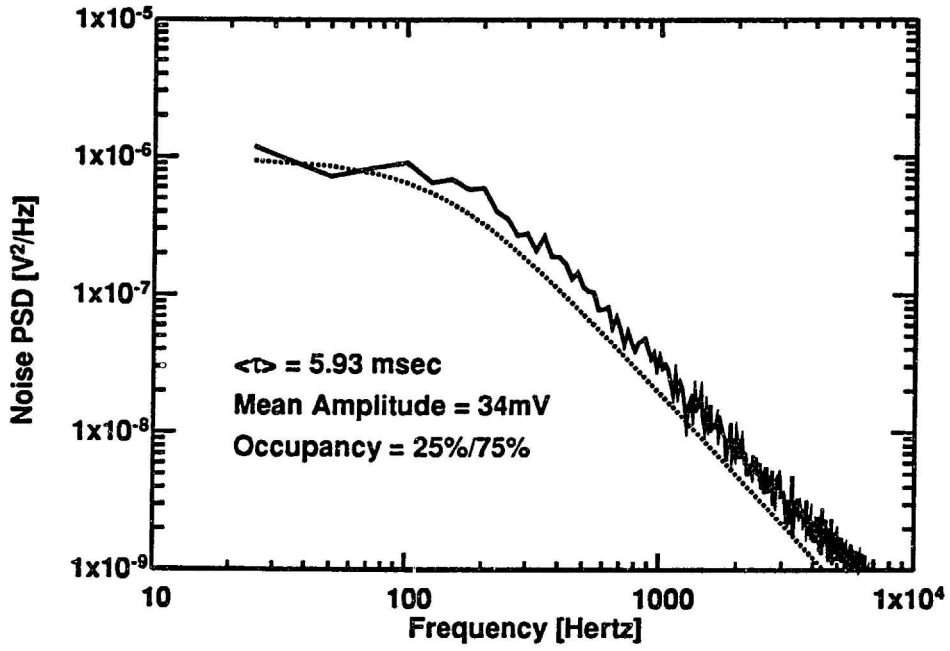
It is useful to define an effective time constant for the single electron trap as

$$\tau = \frac{1}{\lambda_h + \lambda_l}. \tag{5.22}$$

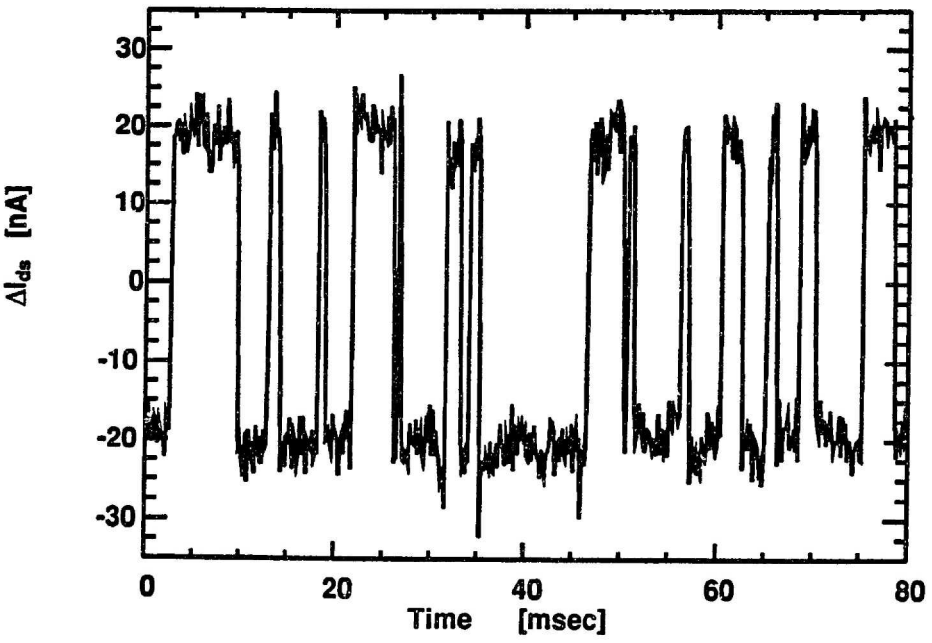
This allows us to write the single-trap spectrum of Equation (5.21) in the simplified form of

$$S_{RTS}(\omega) = A^2 f(1-f)\frac{4\tau}{1+(\tau\omega)^2}. \tag{5.23}$$

Writing the noise spectrum of a single trap in this form will simplify the calculation of large geometry device $1/f$ noise in the next chapter, with no loss of generality.



(a) Frequency domain.



(b) Time domain.

Figure 5.5: Measured spectrum of a single RTS trap. The dotted line is the expected noise power of the trap given its amplitude, occupancy, and time constant.

5.2.4 Amplitude of Single-Trap Fluctuation

Although a number of papers have been presented reporting on the behavior of trap amplitude [26, 70], none of them have been able to accurately account for all aspects of the amplitude characteristics [25]. In this section the change in drain current due to trapping is shown to be composed of two parts: a part due to the change in the number of mobile channel carriers, and a change in the mobility of channel carriers caused by the trapped charge acting as a Coulombic scatterer. Furthermore, the mobility fluctuation effect is shown to dominate. A qualitative model of scattering by a screened Coulombic scatterer is applied to explain the observed amplitude dependence on gate bias.

Results

An *ideal number fluctuation* can be defined as the situation which occurs in a MOSFET where every free carrier in the inversion layer will contribute the same amount to the current flowing through the device. The drain current flowing in an n-channel MOSFET transistor in the linear region can be written as

$$I_{ds} = \frac{W}{L} V_{ds} \mu_{eff} |Q_n|, \tag{5.24}$$

where W and L are the device width and length respectively, and Q_n is the inversion layer charge density. The fractional change in the drain current ΔI_{ds} caused by the trapping/removal of a single electron from the inversion layer, when normalized by I_{ds}, can be written as the magnitude of the electronic charge normalized by the total integrated inversion layer charge,

$$\frac{\Delta I_{ds}}{I_{ds}} = \frac{\frac{W}{L} V_{ds} \mu_{eff} \frac{q}{WL}}{\frac{W}{L} V_{ds} \mu_{eff} |Q_n|} = \frac{q}{WL|Q_n|}. \tag{5.25}$$

The normalization by drain current yields a dimensionless quantity which is more applicable to the $1/f$ noise results and model reported in Chapter 6.

In real devices Equation (5.25) does not hold [25]. The primary reason for this is that the capture and emission of carriers to and from traps in the oxide causes a change in the mobility of carriers in the channel. It is well known that charges in the oxide act as Coulombic scatterers to affect the mobility of channel carriers [66]. When the occupancy of a single trap changes, its charge state also changes. Hung et al. treat this situation by assuming the fluctuation in trap occupancy will cause a uniform global fluctuation in both the channel mobility and the channel charge [26],

$$\Delta I_{ds} = \frac{W}{L} V_{ds}(|Q_n|\delta\mu + \mu_{eff}\delta Q_n). \tag{5.26}$$

Normalizing by the drain current yields

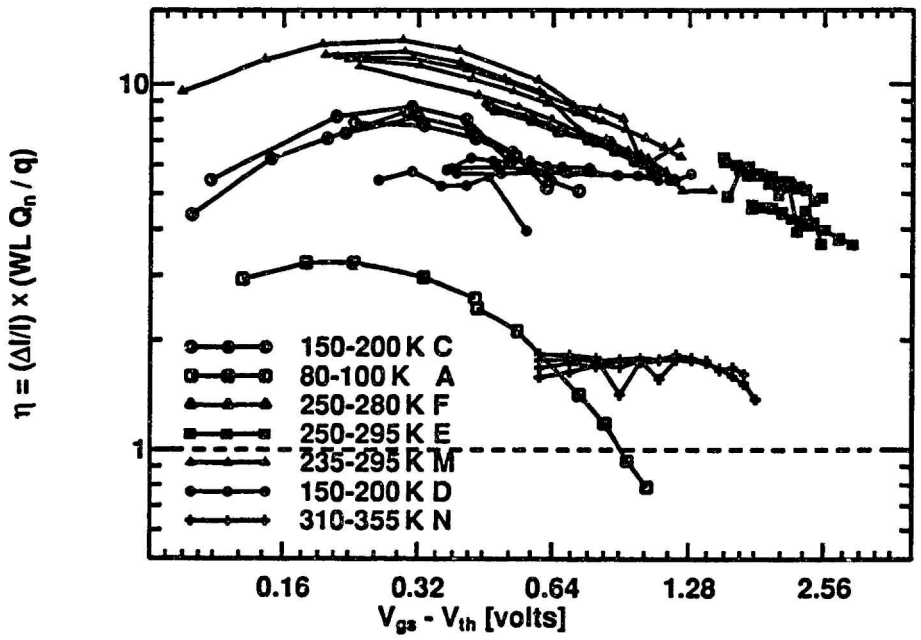
$$\frac{\Delta I_{ds}}{I_{ds}} = \frac{\delta Q}{|Q_n|} + \frac{\delta\mu}{\mu_{eff}}. \tag{5.27}$$

The first term is the ideal number fluctuation term, Equation (5.25), and the second term is an induced mobility fluctuation term. To determine the relative strength of each term in contributing to the current fluctuations, it is useful to divide the normalized drain current fluctuation by normalized charge fluctuation. This yields a quantity η defined as

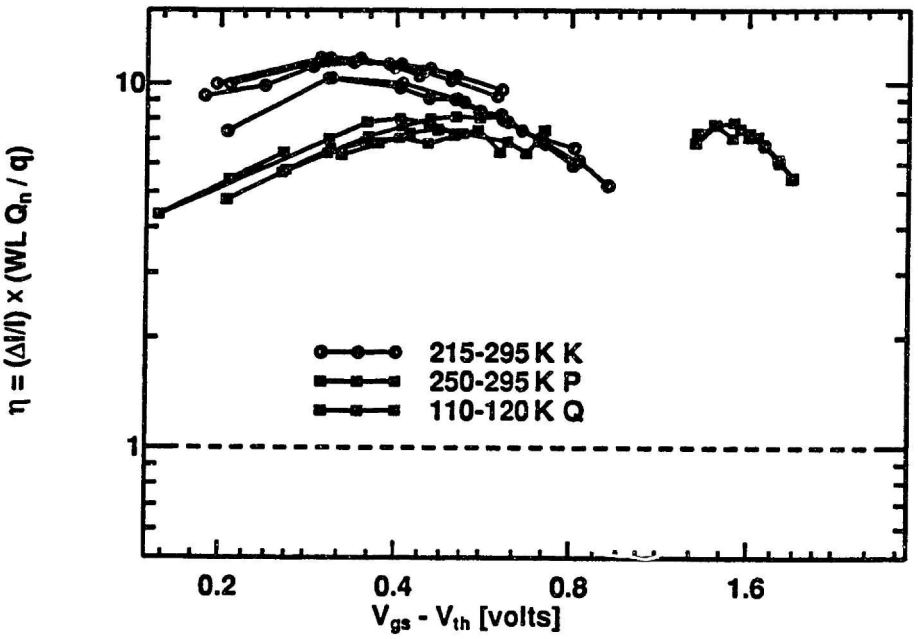
$$\eta = \frac{\frac{\Delta I_{ds}}{I_{ds}}}{\frac{q}{WL|Q_n|}} = 1 + \frac{\frac{\delta\mu}{\mu_{eff}}}{\frac{q}{WL|Q_n|}}, \tag{5.28}$$

which is equal to unity if the drain current fluctuation is due only to an ideal number fluctuation (i.e., if $\delta\mu = 0$).

In Figure 5.6a. are shown plots of trap η-factor *versus* gate bias for various oxide traps at various temperatures. The dashed line indicates the expected η-factor for traps which are due to ideal number fluctuations. A similar plot for ROXNOX devices is shown in Figure 5.6b. For both plots in Figure 5.6 the trap η-factor is seen to be much greater than one over the whole range of bias for nearly all of the device characteristics plotted, which demonstrates that the effect of single-trap capture and



(a) Oxide devices.



(b) ROXNOX devices.

Figure 5.6: Magnitude of normalized drain current fluctuation of single-electron traps divided by normalized charge fluctuation. Note that all of the curves are significantly higher than 1, indicating the dominant role played by channel mobility fluctuations.

emission on channel *mobility* is in most cases much more important than the effect on the number of carriers in the channel.

If the traps we are characterizing are neutral when empty, it is expected that the magnitude of mobility will be reduced when the trap is full; hence $\delta\mu$ in Equation (5.28) will be positive (i.e., $\delta\mu$ will have the same sign as the charge fluctuation term δQ). For a trap which is positively charged when empty, it is expected that $\delta\mu$ will be negative when the trap is full. As the inversion charge density increases, the effect of a single Coulombic scatterer on the channel current is reduced as the trap becomes effectively screened by the mobile inversion layer charge [75]. Therefore, for all cases, we expect the $\delta\mu$ term to strongly decrease with higher gate bias.

An interesting feature of both the oxide and ROXNOX devices is that from device to device there is a considerable spread in the magnitude of normalized drain current fluctuations due to single traps. This can be seen by the spread in the η-factors of the data in Figure 5.6, as well as in data reported by Kirton and Uren in their review article on single-electron-trap fluctuations (Figure 16 in [25]). As was pointed out in Chapter 3, the measurement technique used here could introduce an unintended bias towards traps which cause larger drain current fluctuations. This is not believed to be a critical problem, since the traps which cause the largest drain current fluctuations will also have the largest noise power, and so should be relatively more important in determining the $1/f$ noise performance of MOSFETs.

Drain Current Fluctuation Model

Equation 5.27 above is useful at an intuitive level but seems to obscure some of the details of the single-trap fluctuation behavior. For example, it is expected that the mobility of the channel far removed from the single-electron-trap fluctuation will be unaffected by the charge fluctuation; in Equation 5.27, the mobility fluctuation

is assumed to be a global effect, affecting all regions of the channel equally. This is not to say that the net velocity of carriers in other regions of the channel will be unaffected; it very well may be. However, any change of net velocity of carriers in other regions of the channel will be solely due to changes in the electric field in those regions. A model of drain current fluctuations in the MOSFET channel region which treats the fluctuations as a purely localized effect is presented here.

Since we are interested in modeling the single-electron-trap fluctuations in the linear region of the MOSFET, it is useful to model the channel as a conducting sheet of conductivity μQ_n and to model the fluctuations caused by charge trapping as a local change in the channel conductivity. In this context it is possible to treat both the change in the number of channel carriers and the local induced mobility fluctuations. The advantage of modeling the channel as a conductance is useful from the point of view that the final result is insensitive to the boundary conditions at source and drain; that is, the model is equally applicable to biasing the drain of the device with either a current source or a voltage source.

The main points of the model will be presented here and then discussed in more detail below. A plan view schematic diagram of the single-trap drain current fluctuation model is shown in Figure 5.7.

1. When a single oxide trap captures an electron, the conductivity near the trap is reduced due to a) the reduction of local density of channel carriers, and b) the reduction of mobility of carriers in the channel due to scattering from the trapped charge.

2. The charge which is trapped is assumed to come from the mobile inversion layer charge in the region near the trap; $\delta Q_T = -\delta Q_n$, where Q_T is the area density of trapped charge.

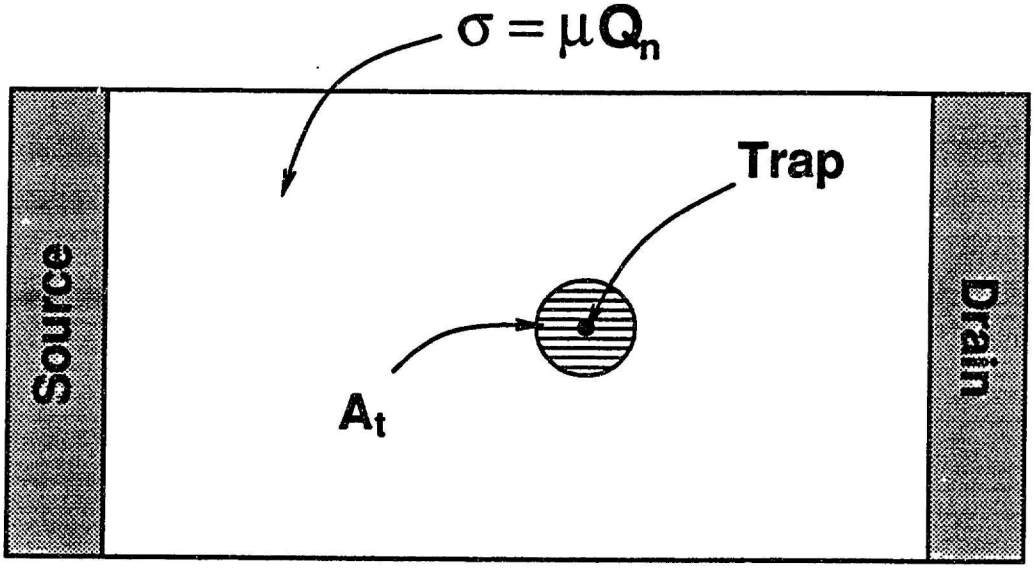


Figure 5.7: Plan view of MOSFET channel showing a single-electron trap. The cross-hatched region of area A_t is the region of the channel whose conductivity is affected by the trapping/de-trapping processes.

3. The area of the channel affected by scattering from the trapped charge is defined primarily by the screening radius for the trapped charge. At very low inversion layer densities, where the trapped charge is only weakly screened, the unscreened scattering cross section will determine the area of the channel affected by the mobility fluctuations.

4. The conductivity of the channel in regions far removed from the trapped charge is unaffected by trap occupancy fluctuations.

5. The induced mobility fluctuation is identically correlated to the charge number fluctuations, and so is represented as a proportionality factor S times the number fluctuation. The S-factor is a function of bias and temperature.

The fourth assumption of the model is a necessary consequence of the first three.

Simoen et al. recently presented a simplified version of this model for current fluctuations in the channel of a MOSFET by representing the region around the oxide trap as having a lower conductivity than the rest of the channel [70]. The

area around the trap A_t was described by a characteristic length L_t which included a combination of Coulombic scattering cross section and screening effects. For their model they obtained the relation

$$\frac{\Delta I_{ds}}{I_{ds}} = \frac{L_t^2}{WL}\frac{\Delta\sigma}{\sigma}, \tag{5.29}$$

where $\sigma = \mu_{eff}|Q_n|$ is the conductivity of the channel, $\Delta\sigma$ is the conductivity change of the area around the trap, and L_t^2 is proportional to A_t. A similar derivation which makes fewer simplifying assumptions about the current flow around the trap is given in Appendix E, producing a similar result as Equation (5.29) [compare with Equation (E.22)]. If we use the representation of $\Delta\sigma/\sigma$ from [70],

$$\Delta\sigma = \mu_{eff}\delta Q_n + |Q_n|\delta\mu, \tag{5.30}$$

the above Equation (5.29) reduces to the form

$$\frac{\Delta I_{ds}}{I_{ds}} = \frac{L_t^2}{WL}\left(\frac{\delta Q_n}{|Q_n|} + \frac{\delta\mu}{\mu_{eff}}\right). \tag{5.31}$$

Note that the quantity $\Delta\sigma/\sigma$ includes both the effects of the fluctuation of charge density, and the fluctuation in mobility, caused by a single trapping event. This is the final form of the model reported in [70]. We will use this expression as the starting point for our model.

In Equation (5.25), the single trapped electron is assumed to come from somewhere in the channel region. For this model, we are assuming that all the effects of the trapped charge are localized near the trap site, due to the effects of screening. The net result of this is that the mobile inversion charge density far from the trap site will be unchanged, while the mobile inversion charge density near the trap will change by an amount

$$\delta Q_n = \frac{q}{L_t^2}, \tag{5.32}$$

where we have defined L_t^2 as the channel area over which the conductivity is affected by trap occupancy fluctuations. Inserting this result into Equation (5.31) above yields the result

$$\frac{\Delta I_{ds}}{I_{ds}} = \frac{q}{WL|Q_n|} + \frac{L_t^2}{WL}\frac{\delta}{\mu_{eff}}, \tag{5.33}$$

which can be seen to be an ideal number fluctuation plus a mobility fluctuation term.

Numerous workers in the past have treated the effect of adding a single Coulombic scattering site in the channel by using Mattheissen's rule [22, 26, 76]. This model is used here as

$$\frac{1}{\mu} = \frac{1}{\mu_o} + S\delta Q_T, \tag{5.34}$$

where μ represents the instantaneous effective mobility of the channel, μ_o represents the "dc" effective mobility of the channel, S represents a scattering constant, and δQ_T is the change in the density of trapped carriers. The scattering constant S is a complex function of bias and temperature, being proportional to the scattering cross section of a fully screened Coulombic scatterer [66, 77, 78]. Note that the units of δQ_T are given in [charge/cm^2] to facilitate comparison of the S-factor here with previous work on Coulombic scattering in MOSFET channel regions. Hung et al. [26] used estimates of S based on the effect of mobility reduction with higher normal channel electric field (as reported by Sun and Plummer [36]). These estimates of S yielded values of the right order of magnitude to explain the RTS results. We can rewrite Equation (5.34) as

$$\mu = \frac{\mu_o}{1 + S\mu_o\delta Q_T}. \tag{5.35}$$

This expression can be written in a more useful form by assuming the induced mobility fluctuation term is small

$$\mu = \mu_o + \delta\mu \simeq \mu_o - \mu_o^2 S\delta Q_T. \tag{5.36}$$

Assumption 2 of our model indicates that the fluctuation of trapped charge equals the negative of the fluctuation of the local mobile inversion layer charge,

$$\delta Q_T = -\delta Q_n. \tag{5.37}$$

The expression for δQ_T written in this way is a statistical construct to indicate that the fluctuation in mobility is identically correlated to the fluctuation in the density of mobile channel carriers; the action of electron capture by the trap has the dual effect of reducing the local mobile carrier density as well as reducing the local effective mobility.

Inserting the quantity from Equation (5.32) into Equation (5.36) allows us to write

$$\delta\mu = \mu_o^2 S \frac{q}{L_t^2}. \tag{5.38}$$

Substituting this result into Equation (5.33), we obtain

$$\frac{\Delta I_{ds}}{I_{ds}} = \frac{q}{WL|Q_n|} + \mu_o S \frac{q}{WL}. \tag{5.39}$$

This is the final form of the model used here. Note that this equation is quite similar to the original rough calculation shown in Equation (5.27) in that the value of the area of effect L_t^2 of the single trapped charge does not appear in the final Equation (5.39).

Coulombic Scattering

The measured results in Figure 5.6 demonstrate that the induced mobility fluctuation is the dominant effect in determining the amplitude of the effect of single-electron-trap fluctuations on the drain current. Therefore, Coulombic scattering is discussed in more detail here to develop a qualitative model of the dependence of the scattering parameter S on bias and temperature.

There has been a considerable amount of work on the effects of Coulombic scattering on the channel mobility in MOSFETs. The review article by Ando et al. provides a good overview of this subject [79]. Many papers have characterized the mobility in a range of temperatures where Coulombic scattering dominates [66, 77, 78]. These papers have all reported that the mobility in these regions follows the relation

$$\frac{1}{\mu_n} \propto \sum_{\text{all } i}^{N_I} \sigma_{m,i} = N_I < \sigma_m >, \tag{5.40}$$

where N_I is the number of individual scatterers affecting the channel mobility, and $\sigma_{m,i}$ is the momentum scattering cross section for the ith scatterer.

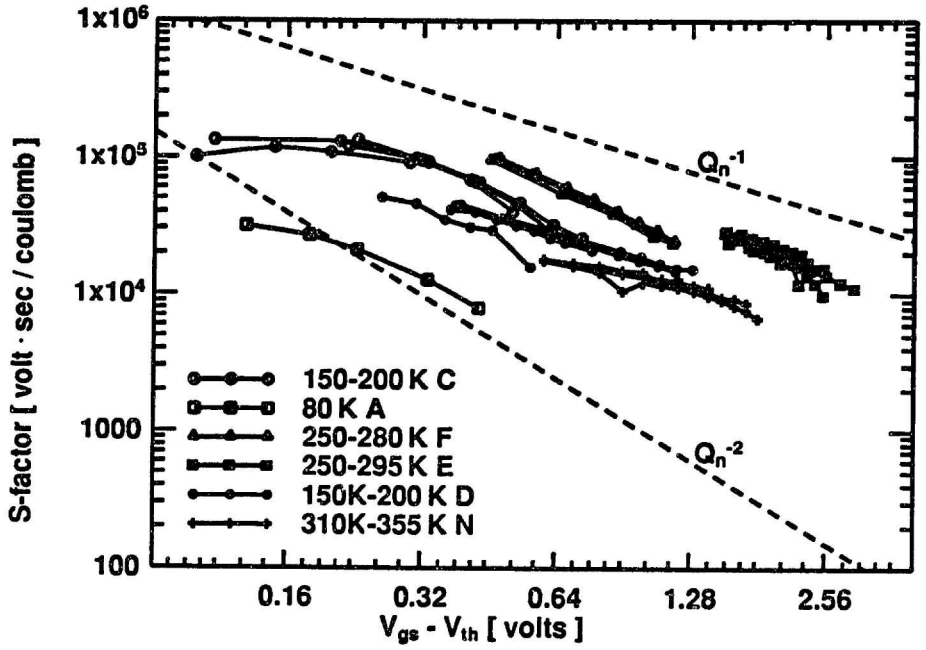
Theoretical calculations of the momentum scattering cross section for a single trap have been reported. However, these have mostly been for temperatures near absolute zero, where all of the inversion layer carriers lie in the lowest subband, and degenerate statistics apply [75, 78]. The calculations of momentum scattering cross section for a single Coulombic scatterer at temperatures of interest to this study have not been reported because of the difficulty of calculating this quantity at temperatures above absolute zero, in addition to the fact the Coulombic scattering is generally not a dominant scattering mechanism above liquid nitrogen temperature (77 K). However, it is possible to make a qualitative analysis of the behavior of the momentum scattering cross section based on the low temperature derivation, and reported measurement results (at temperatures of up to 150 K [66]).

At very low gate biases there are few inversion layer carriers, and the Coulombic scatterer is not screened. For this case the scattering cross section determined from the unscreened single charge potential is a constant independent of gate bias. As the bias is increased and the density of inversion layer carriers is increased, screening of the Coulombic scatterer by the inversion charge begins to be an important effect. Screening has the effect of reducing the momentum scattering cross section. From

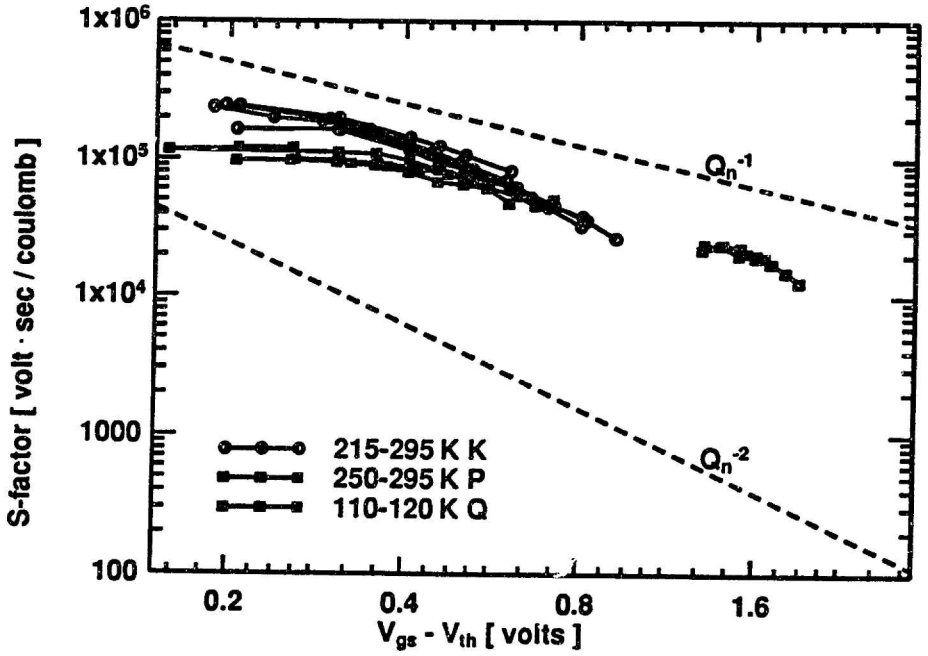
the analysis of Stern and Howard, the momentum scattering cross section is approximately inversely proportional to inversion charge density [75]. The measured results of Vinter [78], Mori and Ando [77], and Hartstein et al. (Figure 62 in [79]) all suggest that the momentum scattering cross section is approximately proportional to $Q_n^{-0.8}$. In addition to the reports described above, the measured results of Sah et al. suggest that the scattering coefficient S is proportional to absolute temperature [66]. It has been suggested that screening effects begin to be important in silicon inversion layers with as few as 10^{11} cm^{-2} mobile carriers in the inversion region [66, 75].

A plot of the measured single-trap scattering coefficient S is shown in Figure 5.8. Lines with slope Q_n^{-1} and Q_n^{-2} are included on the plots to aid in observing the bias dependence. Note the roll-off of the scattering coefficient observed at higher gate bias. This roll-off is believed to be due to the increased screening of the single scatterer, which reduces the effect of that scatterer on channel mobility. The flat region observed at low gate biases could be due to the fact that screening has not become dominant in this region. This cannot be asserted with certainty, since the flat region could also be due to moderate-inversion effects on trap amplitude, first suggested by Reimbold and discussed in more detail below [17]. All of the curves roll off as $Q_n^{-\nu}$, where ν has values between 1.0 and 1.5. There is a considerable amount of scatter in the data from trap to trap. Also, the temperature dependence of the scattering coefficient is too weak to allow any firm conclusions to be made.

Reimbold has modeled the effect of a single trap on the noise characteristics of MOSFETs by using a charge sharing description of charge fluctuations in the inversion layer [17]. In weak inversion, any trapped charge fluctuation is shared among the depletion-layer, interface-state, inversion, and gate charge fluctuations. The expression relating the change in mobile inversion charge Q_n with respect to trapped



a) Oxide devices.



b) ROXNOX devices.

Figure 5.8: Plot of the measured induced mobility fluctuation correlation coefficient S vs gate bias. Note that the variation of the S-factor is close to the variation expected from qualitative analysis.

charge Q_T is given as

$$\frac{\partial Q_n}{\partial Q_T} = \frac{-\beta|Q_n|}{C_{ox} + C_D + C_{IT} + \beta|Q_n|}, \tag{5.41}$$

where $\beta = q/kT$, and C_{ox}, C_D, and C_{IT} are the oxide, depletion-layer, and interface-state capacitances, respectively [17]. As the device is biased more heavily into strong inversion, the above expression approaches a value of -1, indicating that the trapped charge is taken totally from the mobile inversion layer charge. In general, it is assumed that

$$\frac{\partial Q_n}{\partial Q_T} \simeq -1, \tag{5.42}$$

but for the lowest gate biases characterized it is possible that this effect could explain the plateau observed in the S-factor at low biases.

Summary

From the preceding discussion, a general model to explain the magnitude of the drain current fluctuation due to single-electron traps has been developed. For all of the devices characterized in this study, the dominant effect is that the fluctuation of oxide trapped charges induces local fluctuations in the channel *mobility.*

Screening effects are very important in modeling the bias dependency of the amplitude of the single-electron-trap fluctuations, causing the single-electron trap amplitude to roll-off rapidly with higher gate bias.

There is a considerable amount of scatter in the measured trap amplitude data. The degree of scatter is consistent with the results reported by other workers [25, 70].

5.3 Inversion Layer Model

In order to obtain the full dependence of the inversion layer wave functions on temperature and bias conditions, a simulation was implemented in the C programming

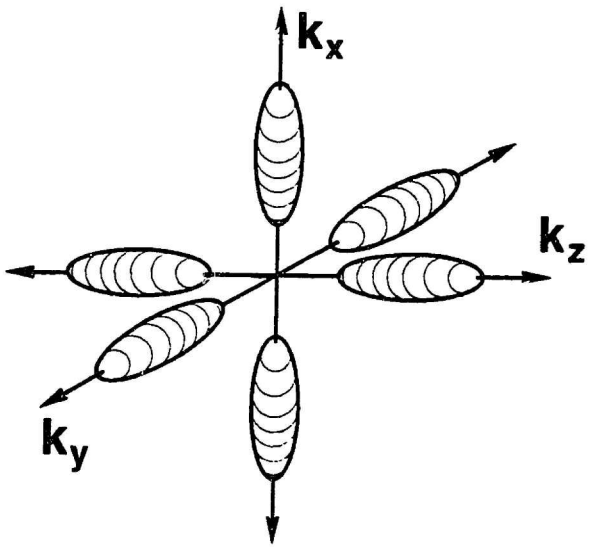


Figure 5.9: Fermi surfaces of a Brillouin zone for silicon

language on a DEC series 5000 workstation. The inversion layer of a MOSFET is modeled using the physics of quantized subbands [75]. This is a method to calculate a self-consistent solution to the wave function of carriers in the inversion layer using both Schrödinger's equation and Poisson's equation. This method was first reported by Stern and Howard [75] to model the mobility in inversion layers. Later workers expanded on the use of this method to estimate the behavior of mobility in the inversion layer [77, 80 - 83]. Weinberg used this method to estimate the inversion layer wave function penetration for Fowler-Nordheim tunneling [84]. The review article by Ando et al. details the work in this field [79].

5.3.1 Assumptions

The conduction band of silicon has six equivalent conduction band minima in the Brillouin zone, shown schematically in Figure 5.9 as Fermi surfaces in reciprocal space ($\vec{k}$-space). Each Fermi-ellipsoid can be modeled using two effective masses: a heavy mass for motion along the long axis of the ellipsoid, and a lighter mass for motion in each of the two directions perpendicular to the long axis. For (100)-oriented silicon,

two of the Fermi-ellipsoids are positioned such that the heavy mass describes motion in the direction normal to the silicon surface (the $\vec{k_z}$ direction in Figure 5.9). The other four Fermi-ellipsoids are positioned such that the light mass describes motion in the direction normal to the silicon surface. In depletion and inversion, the electric field created by the depletion and inversion charges forms a potential well to confine the free inversion carriers near the interface. The high values of electric field encountered near the silicon surface lifts the degeneracy of the conduction band, and the two heavy mass ellipsoids become preferentially occupied.

The solution to Schrödinger's equation for the inversion layer wave function consists of an envelope function in the direction normal to the interface (the $\hat{z}$-direction), times a plane wave function in the other two dimensions,

$$\psi(x,y,z) = \zeta_n(z)e^{j(k_x x + k_y y)}. \tag{5.43}$$

In general, the simulation solves the one-dimensional Schrödinger equation for the envelope function $\zeta_n(z)$, and coupling this with the solution of Poisson's equation, converges to a single self-consistent solution. Complete details of the simulation are given in Appendix F.

5.3.2 Simulation Results

It is interesting to note that the quantization effects are, as expected, negligible at low bias at room temperature. For this case, the carriers are nearly uniformly divided among the six conduction band minima as in the classical case. However, quantization effects begin to be pronounced for even moderate values of gate bias. For example, with an inversion layer density of 4×10^{12} carriers per square centimeter (corresponding to a gate bias of 2 V above threshold for a 100 Å oxide) at room temperature, the two heavy mass conduction band minima contain over 60% of the

inversion layer carriers, compared to only 33% classically! As the temperature is reduced below room temperature, quantization effects become even more pronounced. The results of this simulation demonstrate that estimating the carrier density using the Boltzmann approximation as

$$n = N_c e^{-\frac{E_c - E_f}{kT}} \tag{5.44}$$

is a poor approximation, which strongly affects the model parameters reported in previous work [25, 26, 85].

Simulations were performed using a substrate doping density of 10^{16} cm^{-3}. This is comparable to the substrate density of the measured devices, which were doped to a level around 7×10^{16} cm^{-3} [30]. The substrate density of 10^{16} cm^{-3} was chosen because of convergence problems encountered at high values of inversion charge for simulations of more heavily doped devices.

A plot of the Fermi level relative to nominal conduction band for various values of temperature is shown in Figure 5.10. These results are in qualitative agreement with the simulation results in [80]. Note that the variation in Fermi level with gate bias will cause a correction in the extraction of the depth of an oxide trap. We can rewrite Equation (5.8) without making any assumptions as

$$\frac{\partial ln\left(\frac{<\tau_c>}{<\tau_e>}\right)}{\partial V_{gs}} = \frac{1}{kT}\frac{\partial}{\partial V_{gs}}(E_t - E_f) = \frac{1}{kT}\frac{\partial}{\partial V_{gs}}(E_t - E_c) + \frac{1}{kT}\frac{\partial}{\partial V_{gs}}(E_c - E_f). \tag{5.45}$$

The first term of this expression can be written as

$$\frac{1}{kT}\frac{\partial}{\partial V_{gs}}(E_t - E_c) = \frac{q}{kT}\frac{\partial}{\partial V_{gs}}\left(-\frac{x_t}{t_{ox}}\frac{|Q_n + Q_B|}{C_{ox}}\right) \simeq -\frac{q}{kT}\frac{x_t}{t_{ox}}\frac{1}{C_{ox}}\frac{\partial |Q_n|}{\partial V_{gs}}. \tag{5.46}$$

The differentiation of inversion charge with respect to gate voltage is equal to

$$\frac{1}{C_{ox}}\frac{\partial |Q_n|}{\partial V_{gs}} = 1 - \frac{1}{q}\frac{\partial (E_f - E_c)}{\partial V_{gs}}. \tag{5.47}$$

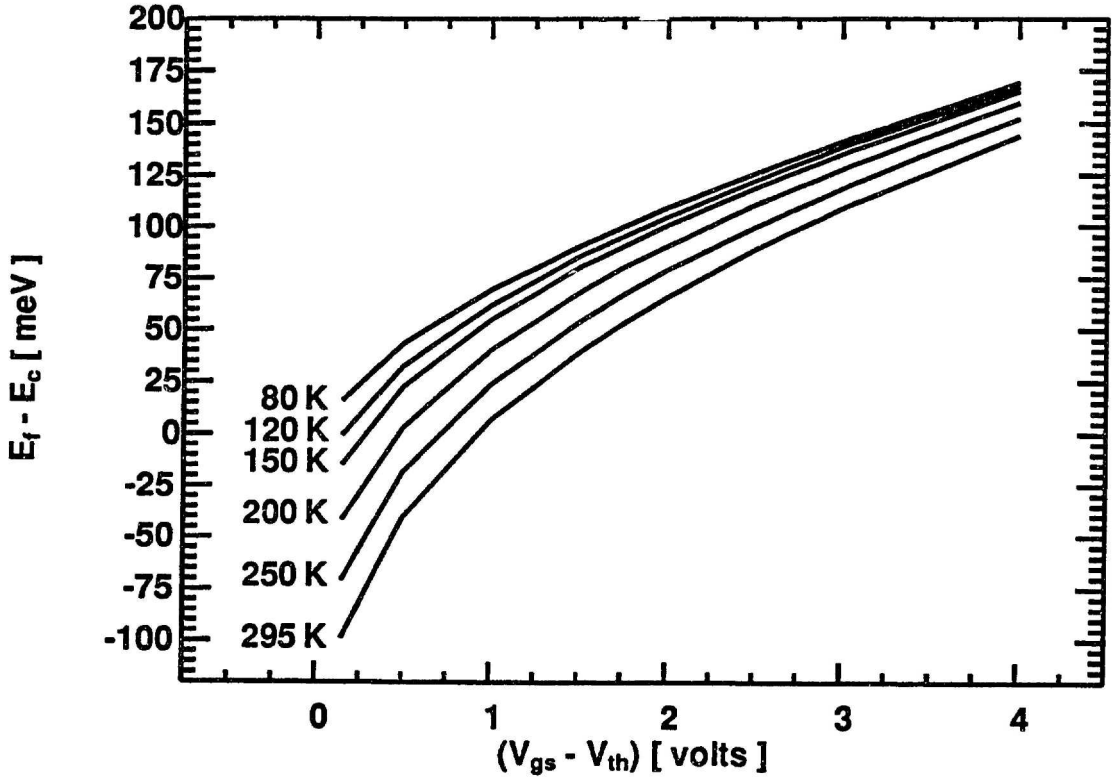


Figure 5.10: Variation of E_f with V_{gs} for various temperatures.

Equation (5.45) can now be simplified to be

$$\frac{\partial ln\left(\frac{<\tau_c>}{<\tau_e>}\right)}{\partial V_{gs}} = -\frac{q}{kT}\frac{x_t}{t_{ox}} + \frac{1}{kT}\frac{x_t}{t_{ox}}\frac{\partial(E_f - E_c)}{\partial V_{gs}} + \frac{1}{kT}\frac{\partial(E_c - E_f)}{\partial V_{gs}}$$
$$= -\frac{q}{kT}\frac{x_t}{t_{ox}} - \frac{1}{kT}\frac{(t_{ox}-x_t)}{t_{ox}}\frac{\partial(E_f - E_c)}{\partial V_{gs}}. \qquad (5.48)$$

Using the slope of the curves in Figure 5.10, we can see that ignoring the variation of the trap quasi-Fermi level with gate bias will cause an overestimation of trap depth into the oxide. The depths of the traps x_t from Table 5.1, recomputed using the results of the inversion layer simulation and Equation (5.48) above, are presented in Table 5.2.

A plot of the average carrier density in the inversion layer vs V_{gs} for various temperatures is shown in Figure 5.11. The variation of the average carrier density with gate bias in Figure 5.11 is seen to be much closer to being linear than to the exponential dependence expected using the Boltzmann approximation. The variation of carrier density with temperature is also seen to be relatively weak (for example, when compared to the variation in N_c with temperature).

5.4 Time Constant Model

From the characterization of single-electron traps a number of interesting properties have become evident. The distribution of capture and emission times makes it clear that the capture and emission processes can each be modeled as Poisson random processes. Each of the capture and emission times is well described by a single rate parameter $\lambda(V_{gs}, T)$ which is a function of bias and temperature. Another interesting property of the capture and emission times is the large activation energies for these processes.

There has been a great deal of work on $1/f$ noise which has treated traps in the oxide as, more or less, a particle in a box problem. The trap is treated as being at

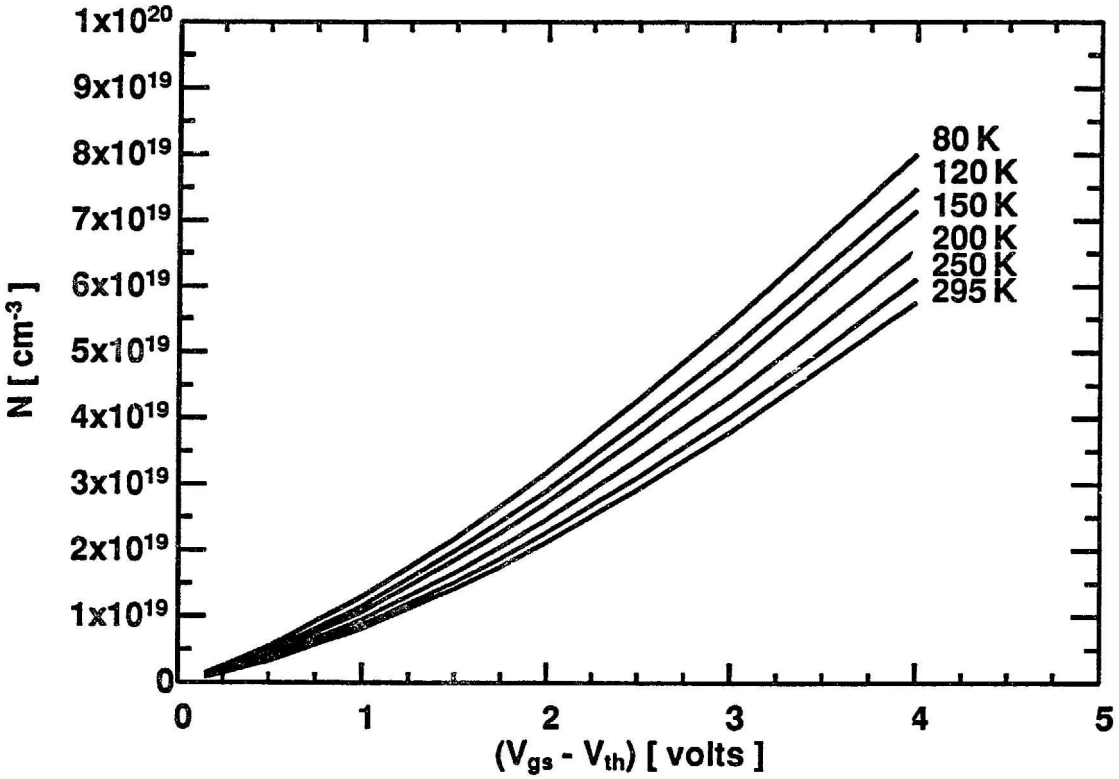


Figure 5.11: Variation in average carrier density n with V_{gs}.

I.D.	Device Type	Temp. (K)	V_{gs} Range (V)	x_t approx (Å)	x_t (Å)
A	OX	80-100	1.0-1.6	7.5	2.6
B	OX	120-185	1.1-1.7	9.8	4.9
C	OX	150-200	0.7-1.2	16.0	10.3
D	OX	150-200	0.9-1.8	14.3	7.9
E	OX	250-310	2.2-3.0	9.8	6.0
F	OX	250-340	1.0-1.7	16.9	8.5
G	ROXNOX	110-165	1.8-2.5	11.2	7.4
H	ROXNOX	150-200	1.9-2.4	23.5	20.2
I	ROXNOX	165-200	1.0-1.3	41.9	37.2
J	ROXNOX	200-250	0.5-0.8	34.8	26.8
K	ROXNOX	215-295	0.2-0.7	23.9	10.5
L	ROXNOX	310-355	0.6-1.2	28.5	21.0

Table 5.2: Trap depths computed using estimate of the variation of Fermi level with gate bias from the inversion-layer simulation. The column on the far right shows the recomputed trap depths.

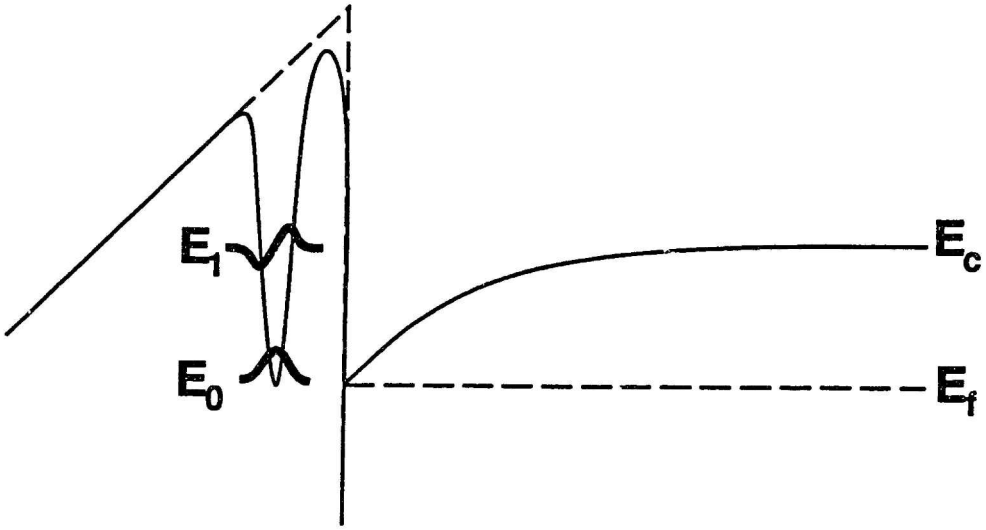


Figure 5.12: Oxide trap model applied in previous $1/f$ noise work. The two lowest energy states of the trap are shown schematically as E_0 and E_1.

a discrete energy, and communicates with the oxide by elastic tunneling (see Figure 5.12). We will call this model the "simple oxide trap" model. A detailed discussion of this model will be given here to demonstrate that it is inadequate to explain the measured behavior of single-trap kinetics. Following this discussion, a discussion of the multiphonon emission model will be presented. The multiphonon emission mechanism of capture and emission has been proposed in recent work to explain the observed trap kinetics [25].

5.4.1 General Rate Equations

The trapping and de-trapping processes can be described in a general way by writing the rate equations for the two processes. A few basic assumptions will be made. The first assumption is that the traps communicate only with the conduction band of the silicon. In an actual device it is expected that traps could capture from or emit carriers to the gate electrode, to the valence band of the silicon, or even to other traps, but it is expected that these processes will always be much slower than

the processes communicating with the conduction band. The second assumption is that Shockley-Read-Hall (SRH) statistics apply to the capture and emission processes [86]. SRH statistics are widely applied to the trapping and de-trapping processes in interface states.

For the moment, consider a distribution of traps with density N_t, all of which are located at a specific energy E_t relative to the vacuum level. An individual trap will have a capture rate given by the product $c_o n$, where n is the density of carriers in the conduction band, and an emission rate given by the product $e_o N_c$, where N_c is the effective density of empty conduction band states assuming $N_c \gg n$. The net change in the density of trapped electrons n_t can be written as

$$\frac{\partial n_t}{\partial t} = c_o(N_t - n_t)n - e_o N_c n_t. \tag{5.49}$$

Extending this rate equation for tunneling transitions is straightforward. The rate constants c_o and e_o are simply modified to include the effects of tunneling. In past work, this has been treated by reducing the capture and emission rates by multiplying them by a factor proportional to $e^{-\alpha x}$ to account for the probability for an electron to tunnel to a trap located at distance x from the interface using the Wentzel-Kramers-Brillouin (WKB) approximation [87].

Interface State Kinetics

Before proceeding, a discussion of how the above rate equation is applied to a traditional interface state will provide a convenient point of reference. This is especially useful for determining the temperature dependence of the capture and emission processes in the simple oxide trap model.

The following discussion is reproduced in part from Van der Ziel, p. 134 [86]. For application to interface states, the number of empty conduction band states is

considered to always be much greater than the number of full states. This is a good approximation well into strong inversion. For this case, we can redefine an effective emission rate e_o^* as the product of the emission rate times the effective density of conduction band states

$$e_o^* = e_o N_c. \tag{5.50}$$

The rate equation (5.49) can be rewritten as

$$\frac{\partial n_t}{\partial t} = c_o(N_t - n_t)n - e_o^* n_t. \tag{5.51}$$

This equation is applicable to interface state trapping and emission assuming the density of accessible empty states in the conduction band is always much greater than the number of free carriers.

For interface states, the emission rate is simply the rate of thermal emission for an individual trap, given by

$$e_o^* = \nu e^{-\frac{E_c - E_t}{kT}}, \tag{5.52}$$

where ν is a thermal release frequency on order of 10^{12} per second. The exponential term is a Boltzmann factor representing the probability for the trapped electron to gain an energy $E_c - E_t$ in one cycle. The capture rate is proportional to the average thermal velocity of free carriers $<v_{th}>$ and a capture cross section σ as

$$c_o n = n <v_{th}> \sigma. \tag{5.53}$$

Approximating n using Boltzmann statistics, and recognizing that in steady-state the left-hand side of Equation (5.49) will be equal to zero, we can write

$$c_o(N_t - n_t)N_c e^{-\frac{E_c - E_f}{kT}} = e_o^* n_t. \tag{5.54}$$

Rearranging terms and using the results of Equations (5.52) and (5.53) above, we can write

$$\frac{N_t - n_t}{n_t} = \frac{\nu}{N_c <v_{th}> \sigma} e^{-\frac{E_f - E_t}{kT}}. \tag{5.55}$$

We can define a trap quasi-Fermi level $\overline{\mu_t}$ to describe the density of filled traps n_t in terms of the total density of traps N_t, so that we can write

$$\frac{N_t - n_t}{n_t} = g e^{-\frac{\overline{\mu_t} - E_t}{kT}} = \frac{\nu}{N_c <v_{th}> \sigma} e^{-\frac{E_f - E_t}{kT}}, \tag{5.56}$$

where g is the interface trap degeneracy factor. In general, it is assumed that

$$\frac{\nu}{N_c <v_{th}> \sigma} = g \simeq 1.0, \tag{5.57}$$

which also indicates that the trap quasi-Fermi level will coincide exactly with the substrate quasi-Fermi level. For the case of interface states, it is clear that this assumption must hold, since the substrate material is assumed to be a homogeneous conducting material, and so will equilibrate with a single-electron quasi-Fermi level to describe the occupancy of all states. With this assumption, the trap quasi-Fermi level coincides with the electron quasi-Fermi level.

A common solution technique is to solve for or characterize only one of either the capture or emission rates and use the assumption that the trap quasi-Fermi level must be the same as the electron quasi-Fermi level in steady state. This fully constrains the problem. We have seen from the previous development that independently specifying the capture and emission rates will result in a distribution which will not necessarily have the same quasi-Fermi level as the free electrons in the substrate.

The emission characteristic for an interface state can be seen to have an activation energy of the same order as the difference in energy between the conduction band E_c and the trap energy level E_t [from Equation (5.52)]. For a device operated in strong inversion, traps which are full and empty for nearly equal portions of the time, i.e. those near the Fermi level, will have a very small temperature activation. The capture rate of an interface state will change with temperature as the thermal velocity, carrier density, and even the capture cross section [86] change with temperature. Using

estimates of the inversion layer density from the inversion layer simulation, we can expect the capture rate to change by no more than an order of magnitude over the temperature range of interest, i.e., certainly not as much as has been characterized for our single-electron traps (this point will be further discussed in Section 5.4.3). Thus we see that the behavior of a typical interface state is radically different from that of the measured single-electron traps.

A two step process for trapping carriers to traps in the oxide has been suggested, whereby transitions occur from the conduction band to an interface state and then to an oxide trap. Interface state mediated transitions do have high activation energy emission behavior. However, the interface states which have high activation energy emission behavior lie at very deep energies in the bandgap, and so would be nearly always full in strong inversion. Therefore, they would be too inefficient at coupling to oxide traps to account for the observed behavior.

General Oxide Trap

For a trap in the near-interface region of the oxide in direct communication with electrons in the conduction band, it is not immediately clear that the assumption of the trap quasi-Fermi level equaling the substrate silicon quasi-Fermi level [Equation (5.56)] must hold. A number of things are different about the case of an oxide trap. The trap is located in an insulating material, and so the concept of a constant, well-defined Fermi level does not apply. However, we must note that for traps for which the dominant communication is with the conduction band, the principle of detailed balance leads to the result that the trap quasi-Fermi level must be equal to the conduction band quasi-Fermi level in steady state. For example, if 99% of the states at a certain energy in the conduction band are empty it is inconceivable that a trap located in the oxide a few angstroms away and at the same energy could be full for

any amount of time very much different than 1% (it has been suggested that entropy considerations could alter the trap occupation from being strictly 1% for this case [88]).

Equipped with the above analysis, we can evaluate a number of possible permutations of the simple oxide trap model. A number of past workers have treated the problem of tunneling to oxide traps in a way similar to the interface state kinetics described above, with the added assumption that the capture cross section is proportional to $e^{-\alpha x}$ [3, 89]. This factor was included based on the familiar result that the probability for a wave to quantum-mechanically tunnel through a barrier is exponentially proportional to the width of the barrier [87]. An added assumption was that tunneling to the trap was elastic; hence, the traps involved in the noise process are located in the oxide at energies adjacent to the conduction band of the silicon. A difficulty with this model is that it fails to explain the high activation energies observed in the measured data. This makes it seem unlikely that the tunneling transitions are occurring purely elastically to traps adjacent to the silicon conduction band.

The high activation energies suggest the possibility of capture to a state much higher in energy than the conduction band, and emission from a state much lower in energy than the conduction band. It should be noted that we can not *a priori* rule out inelastic tunneling. Traps which lie at energies much below the conduction band can emit carriers to the conduction band without relying on a single-step inelastic tunneling transition, as described in Appendix G. A mechanism involving capture to a trap state above the ground state, followed by decay to the ground state, would fit the above requirements. However, it must be remembered that a trap well below the Fermi level will be full most of the time—any other configuration is energetically unfavorable.

The data on the activation energy of capture and emission suggest some sort of

process involving emission over a barrier. The activation energy of both the capture and emission processes are nearly the same when the trap is full and empty for equal fractions of the time. A model using a purely thermal mechanism for capture and emission is untenable for a number of reasons. The first is that even if the barrier to trapping was much lower than the barrier to the oxide conduction band, the tunneling probability through this barrier would still be much higher than the thermal emission probability. The distances that the traps are away from the interface strongly suggest that tunneling is always going to be a dominant mechanism. For example, all of the oxide traps characterized are less than 20 Å away from the interface.

The net result of this discussion is that the simple oxide trap model (Figure 5.12) provides no insight into the physical mechanisms involved in the capture and emission processes. This analysis also confirms the idea that interface states are not responsible for the trapping and emission behavior. These ideas are emphasized in this section, because of the large volume of work preceding this which adopts the simple oxide trap model to explain $1/f$ noise in MOSFETs [3, 22, 90 - 93].

5.4.2 Multiphonon Emission Model

A number of researchers have modeled the capture and emission processes using a multiphonon emission model. The review article by Kirton et al. presents an overview of the results in this area and of the multiphonon emission model applied to deep-submicron MOSFETs [25].

The multiphonon emission mechanism was first adopted to explain anomalous properties of glassy systems observed at low temperature, linked to a metastability of the atomic configuration of individual bonds [94]. More recently, this mechanism has been successfully applied to explain complex fluctuation noise observed in metal-insulator-metal (MIM) tunnel diodes [95, 96].

The basic idea behind the multiphonon emission model is the fact that a change in the arrangement of the atomic nuclei around a defect in a solid can change the electrical properties of that defect. A complete treatment of this type of problem has been given in the textbooks by Lanoo and Bourgoin [29, 85].

At any defect in a solid, the atomic nuclei around a trap site will relax to a position which minimizes the total energy of the trap. The total energy here is both the electronic energy of the electrons involved in the trap and the elastic energy from the restoring forces that hold the atom in the solid. A useful example is that of a strained bond in a solid which acts as a electron trap. When an electron is removed from the bond, the atomic nuclei move away from one another due to the Coulombic repulsion of the two nuclei and will settle at an equilibrium separation Q_o. When two electrons are in the bond, the Coulombic repulsion is overcome and the two nuclei relax to a new equilibrium separation Q_1. It is entirely possible that the total energy for each of the two situations are equivalent.

To describe the trap transitions involved in multiphonon emission, a configuration-coordinate diagram is used [29]. A configuration-coordinate diagram to describe the above situation is shown in Figure 5.13. This is a graph of the elastic energy of the trap system versus a linear distortion coordinate, representing the distortion of the nuclei which constitute the trap. Generally, a constant offset is added to each curve of trap elastic energy, to represent the electronic energy of the carriers involved in capture and emission. Note that the trap elastic energy of an actual physical system can not usually be described in terms of a single normal coordinate; Figure 5.13 is simply a heuristic description.

The dashed line in Figure 5.13 represents the elastic energy of the empty trap as it is distorted, with an electron at the Fermi level. This is the energy zero of the trap system. Moving the electron to the conduction band will add a constant offset to this

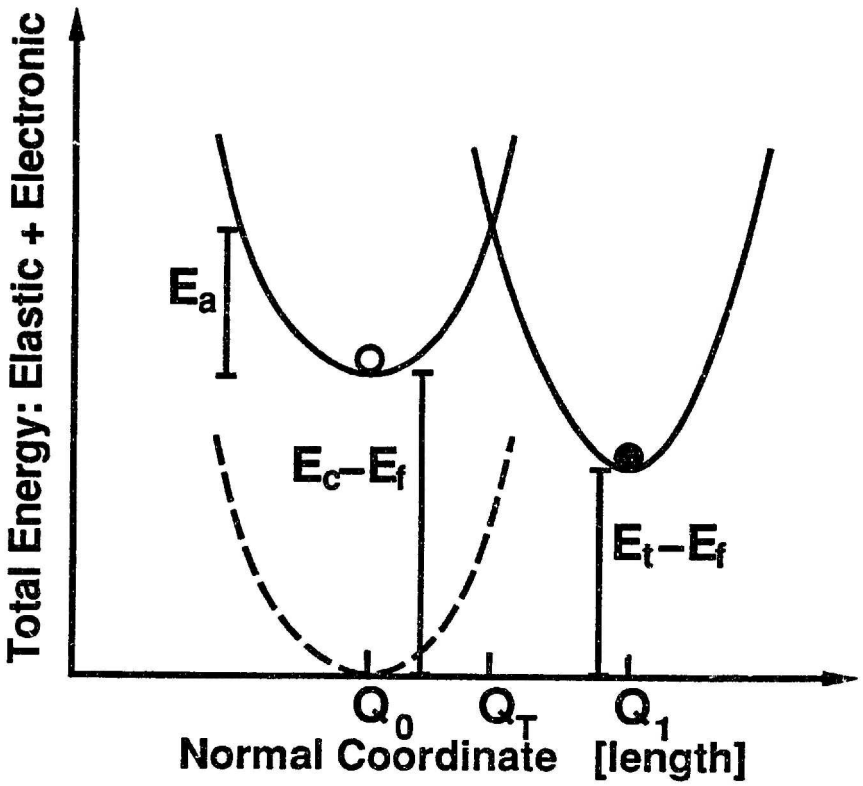


Figure 5.13: Configuration coordinate diagram. The dashed curve represents the total energy of the empty trap with an electron at the Fermi level. Solid curve with o is energy of empty trap with free electron in the conduction band. Solid curve with • is energy of full trap. E_a is the activation energy for the capture process.

curve. This is shown by the solid curve with the open circle, to represent the total elastic plus electronic energy of the system with an electron in the conduction band. Note that as the trap is distorted away from its equilibrium distortion Q_o, the elastic energy of the system is increased. As the trap is distorted, the electronic energy of the system remains constant: only the elastic energy of the system changes. The solid curve with the full circle represents the elastic plus electronic energy of the system with an electron in the trap. The stable distortion value for a full trap is at point Q_1. As the full trap is distorted away from its equilibrium value, the elastic energy of the system is likewise increased.

Note that the transition from a full trap in equilibrium to an empty trap in equilibrium not only involves movement of an electron, but also a significant movement of the atoms around the trap. It is for this reason that tunneling from the ground

state of the empty trap (at Q_o) to the ground state of the full trap (at Q_1) is unlikely, even if the difference in energy of these two states is equivalent.

If the atoms which make up the trap are moved to distortion coordinate Q_T, the total energy of the system (elastic plus electronic) is the same regardless of whether the trap is full or empty. Also, with the trap distorted to the point Q_T, the only tunneling/movement required for the trap to change states is for the electron to tunnel from the conduction band to the trap, or vice versa. Because of this, tunneling transitions between the trap being full and the trap being empty are overwhelmingly favored at the trap distortion coordinate Q_T. Note that the tunneling transitions do not necessarily conserve electronic energy, but they do conserve *total* elastic plus electronic energy.

It has been shown that the trap near its equilibrium distortion can be modeled as a simple harmonic oscillator [29]. When the trap is at nonzero temperature, the trap can be excited to higher vibrational energies by the absorption of phonons from the surrounding lattice. When the vibrations of the trap cause it to be distorted to coordinate Q_T, tunneling is favored. If an electron tunnels to or from the trap causing a change in state, the trap can relax into the alternate state and the excess energy is quickly dissipated as multiple phonons. The release of multiple phonons when relaxing from one lattice coordinate to another is characteristic of both the capture and emission processes and is the origin of the name, "multiphonon emission" mechanism.

The energy difference between the energy at the transition point Q_T and the energy at the equilibrium distortion of the empty trap defines an activation energy E_a for trap capture. A similar activation energy is obtained for the trap to emit. In effect, the activation energy acts as a potential energy barrier for trap capture and emission. Generally, it can be assumed for the trap capture rate λ_c and emission rate

λ_e that

$$\lambda_c \propto e^{-\frac{E_{a,c}}{kT}},$$

$$\lambda_e \propto e^{-\frac{E_{a,e}}{kT}}, \tag{5.58}$$

where $E_{a,c}$ and $E_{a,e}$ are the activation energies for capture and emission, respectively. Note that $E_{a,c}$ and $E_{a,e}$ differ by the difference in ground state energies of the empty trap and the full trap. This is only a few kT for traps which are full and empty for similar fractions of the time. Results of the analysis of defects in bulk silicon indicate that the multiphonon emission mechanism can explain trap activation energies of up to 1 eV [29].

5.4.3 Application to Deep Submicron MOSFETs

The multiphonon emission process is generally modeled as a capture cross section which is proportional to an activation factor [25],

$$\sigma = \sigma_o e^{-\frac{E_a}{kT}}. \tag{5.59}$$

For this model the activation energy for capture and emission is assumed to be the same, and any observed difference in the temperature activation of the data is due to the difference in energy of the trap quasi-Fermi level and the trap ground state energy. The capture time can be written as

$$\tau_c = (\lambda_c)^{-1} = (n < v_{th} > \sigma_o e^{-\frac{E_a}{kT}})^{-1}. \tag{5.60}$$

The emission rate can be written by applying the detailed balance condition of Equation (5.56),

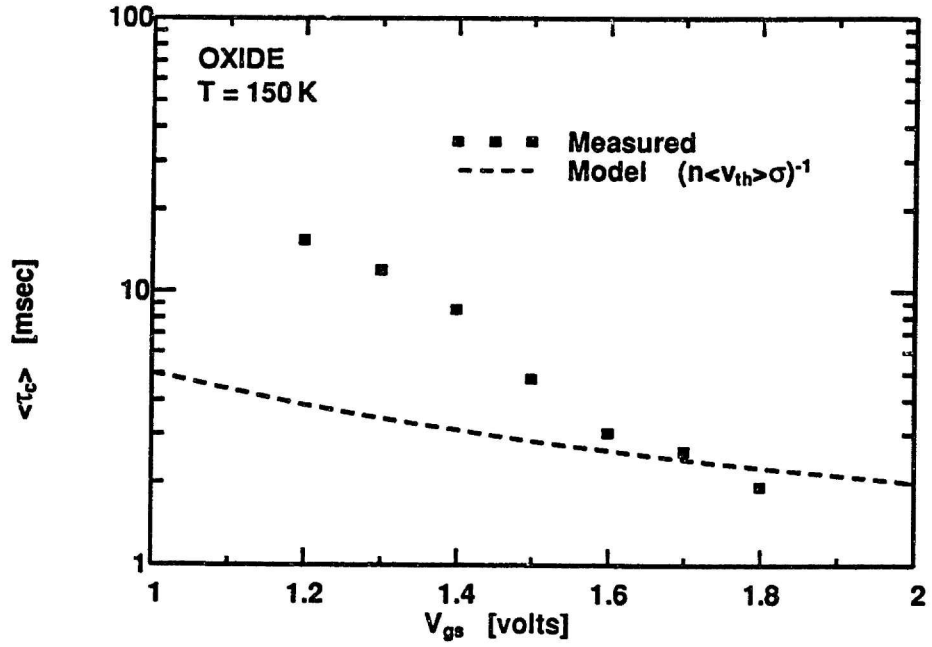
$$\tau_e = \frac{\tau_c}{g} e^{-\frac{E_t - E_f}{kT}}. \tag{5.61}$$

These equations predict the general trends of the data quite well.

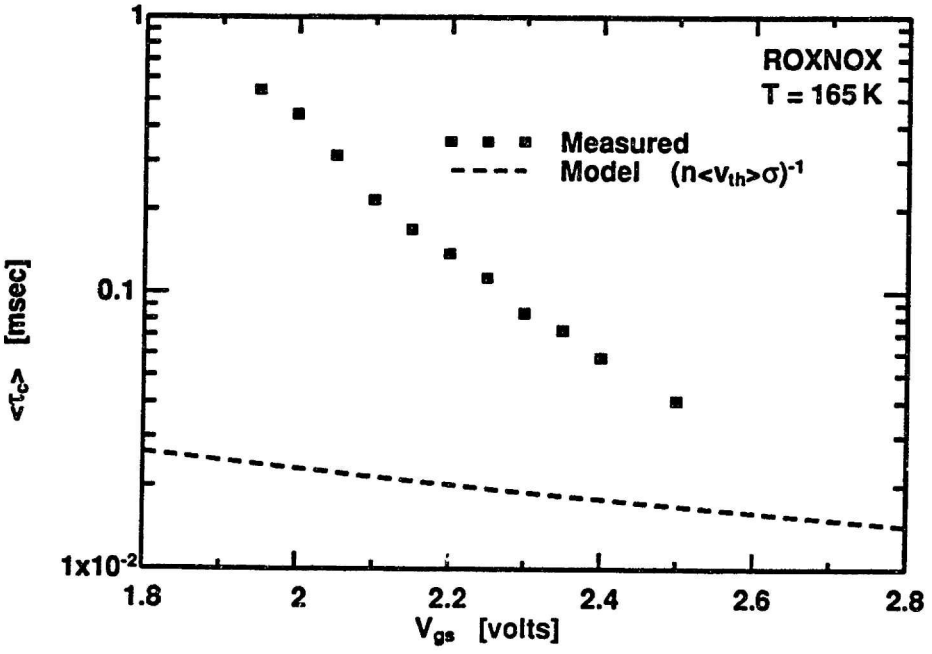
In previous work on analyzing the capture and emission times, a number of simplifying assumptions were made which obscure some interesting trends in the data. In the work of Kirton et al., the volume density of inversion layer carriers vs bias and temperature is extracted from drain current measurements using the Boltzmann approximation and simplified models of the inversion layer thickness and channel mobility [25]. The Boltzmann approximation is clearly not valid in strong inversion. The temperature dependence of channel mobility is also seen to not follow the assumed $T^{-1.5}$ dependence over the entire temperature range. These shortcomings detract somewhat from the reported temperature and bias dependencies reported in that work.

In Figure 5.14 is a plot of the capture time vs gate voltage for the model of Kirton et al. [25] using typical trap parameters of $\sigma_o = 10^{-17}$ cm^{-2}, $< v_{th} > = 10^7(T/300)^{\frac{1}{2}}$ cm/sec and using estimates of n from the quantum inversion layer simulations. A set of typical measured capture time data is shown for comparison.

In nearly all cases the change in capture time with gate bias is much larger than is predicted from Equation (5.60). This indicates that the capture cross section is an increasing function of bias. A graph of the dependence of capture cross section on V_{gs} for the data from Figure 5.14 is shown in Figure 5.15. This is a result which has not been previously reported. This view of a capture cross section which increases with higher gate bias is entirely consistent with our modeling of the capture cross section to incorporate the tunneling properties of the trap. As the gate bias is increased, we expect the penetration of the inversion layer wave function to increase consistent with an increase in tunneling probability.

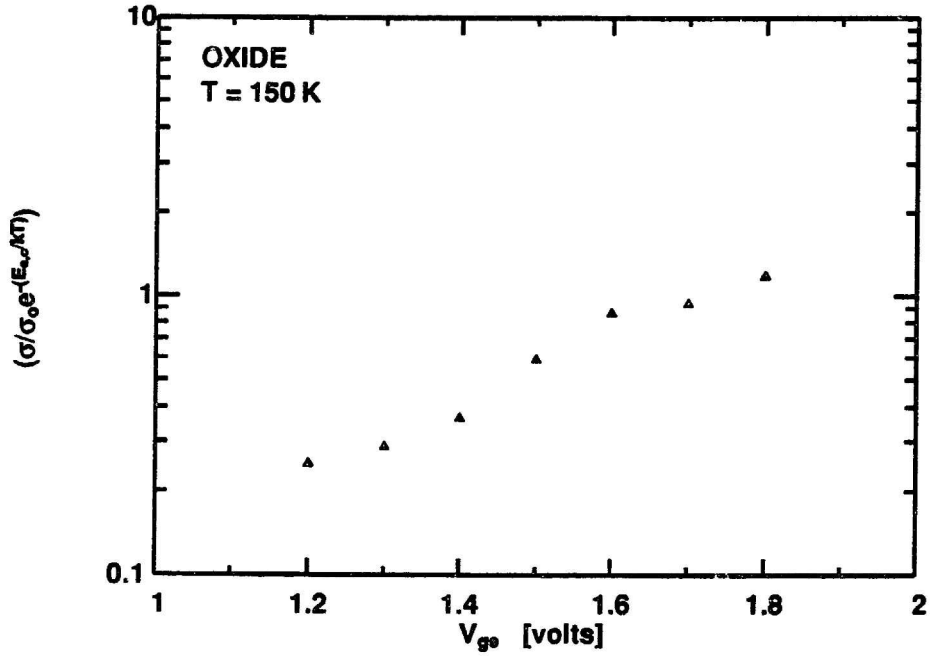


(a) Oxide dielectric

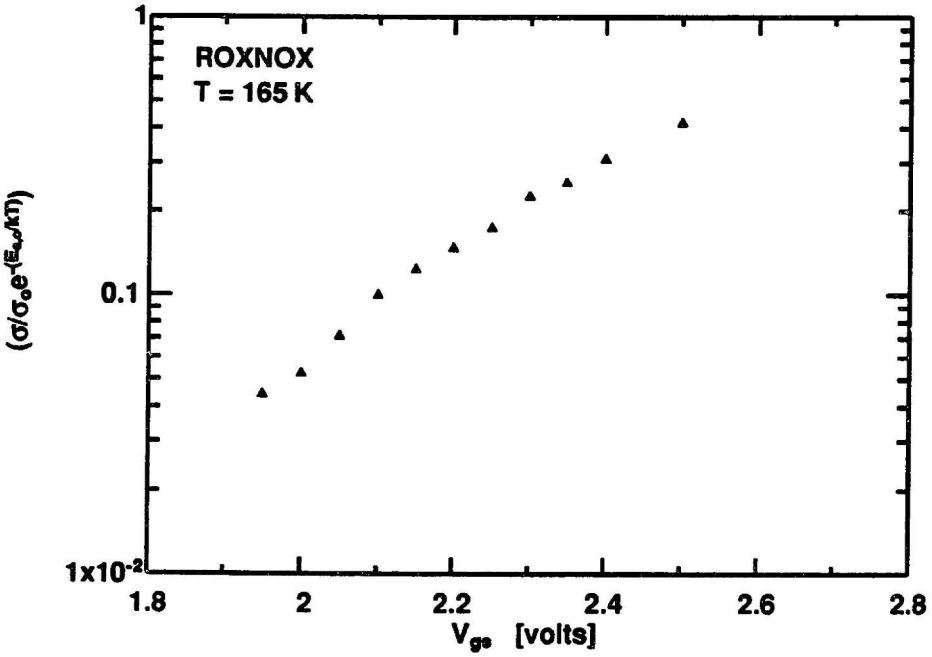


(b) ROXNOX dielectric

Figure 5.14: Plot of capture time vs V_{gs}. The model curve uses the parameters $\sigma_o = 1.0e^{-17}$ cm^2, $E_{a,c} = 285$ meV for oxide, and $E_{a,c} = 250$ meV for ROXNOX.



(a) Oxide dielectric



(b) ROXNOX dielectric

Figure 5.15: Variation of capture cross section with V_{gs}. The data were calculated using the parameters $\sigma_o = 1.0e^{-17}$ cm^2, $E_{a,c} = 285$ meV for oxide, and $E_{a,c} = 250$ meV for ROXNOX.

5.4.4 Summary of Single-Trap Model

The trap time constant τ has been shown to be a strong function of both bias and temperature. This is in contrast to the time constant being a constant dependent only on trap depth into the oxide as assumed by many previous studies of $1/f$ noise. The capture and emission time constants are modeled well by the relations

$$\tau_c = (n < v_{th} > \sigma)^{-1},$$

$$\tau_e = \frac{\tau_c}{g} e^{\frac{E_f - E_t}{kT}}, \tag{5.62}$$

where σ is given by

$$\sigma = \sigma_o e^{\xi V_{gs}} e^{-\frac{E_a}{kT}}, \tag{5.63}$$

in which the factor ξ is positive and varies from trap to trap.

The magnitude of drain current fluctuations caused by individual single-electron traps varies a great deal from the ideal value. In general the normalized trap amplitude factor η is a decreasing function of increasing gate bias.

It is important to comment on the variation of the trap time constant with gate bias. The trap time constant τ given by Equation 5.22 is equivalent to

$$\tau = (\frac{1}{\tau_c} + \frac{1}{\tau_e})^{-1}. \tag{5.64}$$

Using the above relations for capture and emission times, we can rewrite the trap time constant as

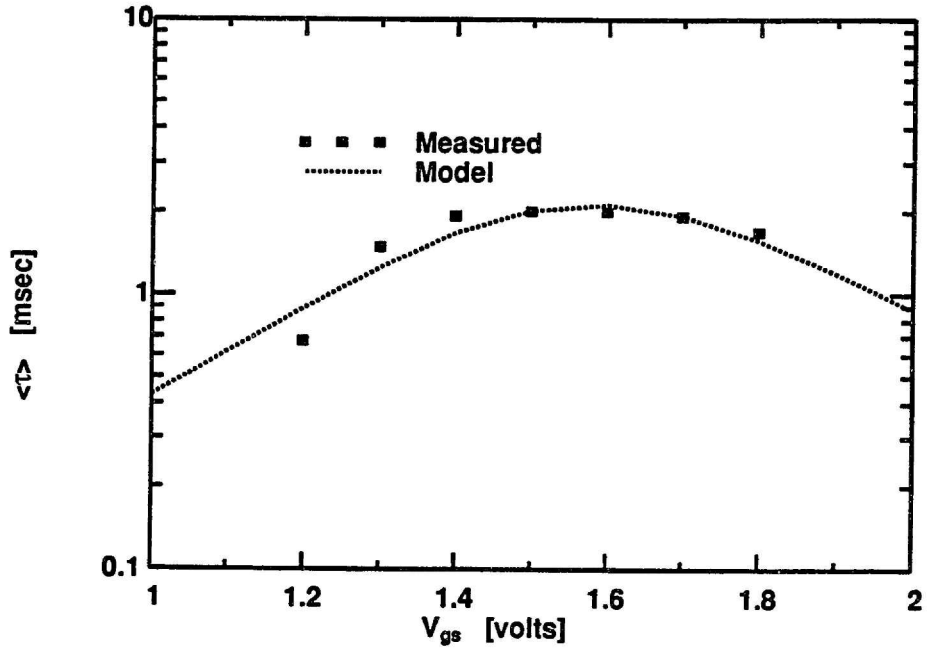
$$\tau = \frac{e^{\frac{E_a}{kT}}}{n < v_{th} > \sigma_o} \left[\frac{e^{-\xi V_{gs}}}{1 + g e^{\frac{E_t - E_f}{kT}}} \right]. \tag{5.65}$$

If we make use of the approximation of Equation (5.7) to estimate the variation of $E_f - E_t$ with gate bias, we can rewrite the above equation as

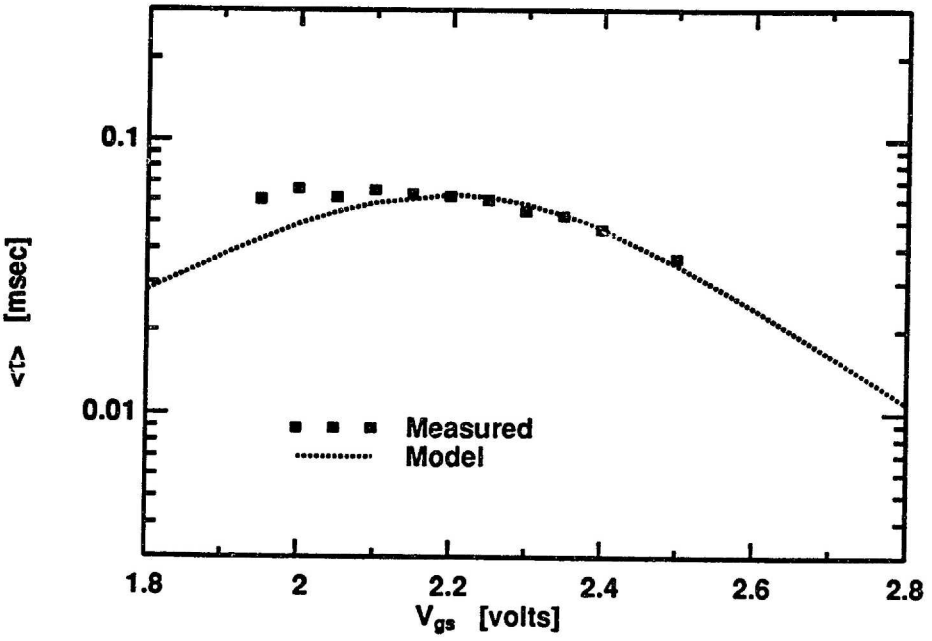
$$\tau = \frac{e^{\frac{E_a}{kT}}}{n < v_{th} > \sigma_o} \left[\frac{e^{-\xi V_{gs}}}{1 + g e^{-\frac{q V_{gs}}{kT} \frac{x_t}{t_{ox}}}} \right]. \tag{5.66}$$

A plot of the trap time constant $\tau = (\frac{1}{\tau_c} + \frac{1}{\tau_e})^{-1}$ *versus* gate bias V_{gs} is given in Figure 5.16. The model curve of Equation (5.66) is also plotted for comparison. This plot demonstrates the hyperbolic-cosine dependence on gate bias exhibited by the trap time constant. The trap time constant becomes very short for both very low and very high values of gate bias. This behavior is contrary to that assumed in the past to model $1/f$ noise in MOSFETs.

Note that the variation of trap time constant with bias is the lowest at gate biases where the trap is full and empty for nearly equal fractions of the time. This is also the region of gate bias where the noise power of a single trap will be the largest. These two effects will be used in Chapter 6 to allow the variation of trap time constant with gate bias to be neglected to first order when calculating the total noise power of an ensemble of single-electron traps.



(a) Oxide dielectric.



(b) ROXNOX dielectric.

Figure 5.16: Trap time constant τ_t *versus* V_{gs} compared to the model of Equation (5.66). $\xi = 2.80$ V^{-1} for oxide, and $\xi = 3.62$ V^{-1} for ROXNOX.

Chapter 6

1/f Noise

$1/f$ noise has been found in a wide variety of electron devices. An excellent review of the literature on this subject was presented by Van der Ziel [9]. Some very interesting measurements have demonstrated that $1/f$ noise is present in any material dominated by bulk conductance [97]. This has lead many to speculate that the $1/f$ noise process may be a product of some fundamental property of conduction in solids [98]. The issue of whether or not $1/f$ noise is a fundamental property of bulk conduction in solids will not be addressed in this thesis.

$1/f$ noise in semiconductors has been observed over a frequency range from microhertz to higher than megahertz, making it a very important effect for understanding the operation of circuits [99]. For example, for signal acquisition in the presence of white noise, the effect of the noise can be reduced without limit by simply averaging the system over successively longer time periods. This averaging will effectively decrease the variance of the signal. This is not true of $1/f$ noise, for which the variance of any given signal will theoretically increase without limit when the signal is averaged over successively longer periods of time [74]. Further ramifications of $1/f$ noise on circuit design will be discussed in Section 6.4.

$1/f$ noise has also been most commonly found in devices which are dominated by surface effects [9]. The devices used by Johnson were vacuum tubes with oxide

coated cathodes [7]. McWhorter observed $1/f$ noise in germanium filaments [10]. He attributed the $1/f$ noise in the filaments as arising from electrons tunneling to traps distributed throughout the native oxide on the germanium filament. This mechanism of $1/f$ noise generation is called the McWhorter noise model.

This chapter begins with an overview of the literature in this field. Next, the measurement results are presented, and a general model for $1/f$ noise is developed which applies the results of the single electron trap characterizations of Chapter 5. Some of the ramifications of the results of this work on circuit design will then be discussed. In the last section of this chapter, results of a study of the effect of radiation on the mobility of ROXNOX transistors are presented, as an example of using $1/f$ noise measurements to investigate the near interface region of the dielectric.

6.1 $1/f$ Noise in MOSFETs

The most important surface dominated device today is the MOSFET [11]. The noise performance of the MOSFET is dominated at low frequency by $1/f$ noise. For the devices used in this work, the corner frequencies, where the $1/f$ noise power spectral density (PSD) of the device equals the thermal noise PSD, are typically in the 100kHz range.

Past investigators have presented models for noise in MOSFETs which fall into two general categories. The first category is the McWhorter model [10], which states that noise in a MOSFET is due to fluctuations in the *number* of mobile carriers in the channel caused by tunneling of channel carriers to and from traps in the dielectric. This mechanism is also known as the number-fluctuation noise model. The traps which participate in the $1/f$ noise process for a device operated in inversion are located very near the interface (within 10 or 20 angstroms) and adjacent to the conduction (for NMOS devices) and valence (for PMOS devices) bands of the substrate silicon.

The second category is the Hooge model [100], which is an empirical model based on the observation of $1/f$ noise fluctuations in the bulk *mobility* of a wide number of metals and semiconductors. Similarly, the Hooge noise model is also known as the mobility fluctuation noise model.

Christensson et al. were the first to apply the McWhorter model to MOSFET devices in a rigorous way [3]. Their theory predicts many aspects of observed $1/f$ noise behavior, although, as mentioned in Chapter 1, it fails to adequately predict the temperature dependence of $1/f$ noise in NMOS devices. Their characterization results vs temperature are very thorough [4]. The Christensson et al. derivation will be discussed in Section 6.3.1.

A number of early workers sought to establish a link between the interface-state density in the MOSFET channel region, and the magnitude of $1/f$ noise in the device. Sah and Hielshcher were the first to show the strong correlation between interface-state density in the MOSFET channel region, and the magnitude of the noise. They demonstrated using NMOS transistors that the structure present in the D_{it} vs gate-bias characteristic measured using capacitor characterization methods is also present in the magnitude of $1/f$ noise vs gate-bias characteristic [101]. This work was extended by a number of workers. Abowitz et al. characterized MOSFET noise at both 300 K and 77 K and observed an increase in noise at lower temperature. They attributed this increase to the higher density of interface states at the Fermi level for the low temperature case, where the Fermi level is at a higher energy in the bandgap. The interface-state density in this case was determined from the non-ideality in the drain current characteristic near threshold [102]. Work by Hsu [90] and Fu and Sah [103] followed the same general approach. More recently, Maes et al. have reported a correlation between $1/f$ noise magnitude and D_{it} of both fresh and stressed devices. For their work, the interface-state density was characterized by charge-pumping mea-

surements [91, 104]. One problem with these papers is that the interface-state density for energies near the band-edge is difficult to accurately determine, calling into question the 1:1 correspondence reported by them. The weak-inversion methods used to extract interface-state density is inherently inaccurate near the band edges [19]. Likewise, high-frequency capacitance methods [20] and charge pumping measurements [21] are also inaccurate for extracting D_{it} information near the band edges. It is clear from these papers, however, that a strong correlation exists between interface-state density and $1/f$ noise.

The work of Van der Ziel is summarized very well in his review article [9]. Most of his work is on the application of the Hooge model of $1/f$ noise to MOSFETs, although his derivations are often in a general form so that conclusions about the number fluctuation mechanism in MOSFETs can be made. Van der Ziel reported on the proper method to normalize input referred noise to determine actual channel noise, as was discussed in Section 3.4.1 [40].

Reimbold characterized $1/f$ noise in n-channel MOSFETs from weak to strong inversion [17], giving his results in terms of relative noise PSD, i.e.

$$\text{relative noise PSD} \equiv \frac{S_{I_{ds}}(f)}{I_{ds}^2}. \tag{6.1}$$

Reimbold's extension of the Christensson et al. model to weak-inversion involved applying charge conservation to the fluctuating trap occupancy: in weak inversion a fluctuation in trapped charges will produce not only fluctuation in inversion charge, but also will produce fluctuations in depletion charge and interface-state charge as was mentioned in Section 5.2.4 [17]. He found that while the results predicted by the Christensson model fit well in the linear region, the noise value in subthreshold is a constant independent of V_{gs}. This behavior was observed later by Boukriss et al., who also characterized the effect of channel-hot-electron stress on $1/f$ noise

[105]. Recently, Kleinpenning provided an excellent explanation for the noise plateau observed in weak inversion by noting that the variance of the number of carriers in the channel cannot be greater than the number of carriers [106]. In effect, when the number of carriers in the inversion layer is smaller than the number of active traps in the channel, the noise departs from the expected results of the Christensson model. Furthermore, in the weak inversion region where the noise is limited by the low number of inversion layer carriers, the relative noise PSD will be proportional to $1/C_{ox}$, as is reported by a number of sources [107 - 109].

Mikoshiba measured input-referred noise vs gate bias in strong inversion for devices with various dielectric thicknesses. He found that the noise varied linearly with gate bias, and could be modeled well by two terms: one which varied as C_{ox}^{-2} and one which varied as C_{ox}^{-1}. He presented a theory of $1/f$ noise in MOSFETs which included both independent number fluctuations and independent mobility fluctuations [110]. A similar approach was taken by Grabowski [111].

A series of papers by the students of T.Y.Hsiang extended the assumption of a uniform trap density of the Christensson et al. model [92, 112 - 116]. Worth noting is the work to extend the Christensson et al. model to all drain bias in [92], and the extraction of PMOS oxide trap density from low-temperature measurements in [115]. Also of interest is the proposal made in [113], that the fluctuation of mobility caused by trapping and de-trapping of channel carriers could be a dominant effect. A thermally activated model of the individual traps was suggested in [116], and applied to PMOS devices to explain the exponent of the $1/f^{\gamma}$ relation vs temperature with only limited success.

The work of Jayaraman et al. treated the fluctuations caused by trapping and de-trapping of carriers as causing both a fluctuation in the number of carriers in the channel and a *correlated* fluctuation in the mobility of the channel [67]. This paper

used the assumption that the induced fluctuations in the channel mobility were a relatively small effect compared to the effect of number fluctuations at low bias [67]. Using his model, he was able to extract trap density vs space, to determine that ROXNOX dielectrics introduce more oxide traps nearer to the interface than in the bulk of the dielectric [89].

The model of Christensson et al. has been widely applied to modeling the $1/f$ noise power in the channel of a MOSFET. This model states that all of the noise power is due to the fluctuation of the *number* of channel carriers, and uses the relation [3],

$$\frac{S_{I_{ds}}(f)}{I_{ds}} = \frac{S_Q(f)}{|Q_n|^2} = \frac{q^2 4kTN(E_f)}{2WL|Q_n|^2 \alpha f}, \tag{6.2}$$

where $N(E_f)$ is the density of oxide traps at the same energy as the semiconductor Fermi level, and α is a tunneling parameter relating to the penetration of the inversion layer wave function into the oxide. This expression has a number of properties which have been widely verified by a number of researchers. Normalized $1/f$ noise has been shown to be proportional to C_{ox}^{-2} [22, 110], and also to be inversely proportional to channel area [117]. A limitation of the models which apply the Christensson et al. model of $1/f$ noise is that they have all used a simplistic model for trap time constant in the dielectric. As we have seen from the previous chapter, the dependence of time constant on device bias and trap energy are very complex quantities. Further discussion will be deferred until after the presentation of the results of the conventional device $1/f$ noise measurements.

6.2 $1/f$ Noise Results

In the literature there is a lack of $1/f$ noise measurements over a range of temperatures open to unambiguous interpretation. A prime example of this is the recent paper by Wong and Cheng, which gives good measurement results of the slope expo-

nent γ of $1/f$ noise power over a wide range of temperatures [118], but little data on other aspects of $1/f$ noise performance. At best, only very qualitative trends can be determined with this type of data. The results presented here are significant in that $1/f$ noise characteristics for modern MOSFET devices characterized over a range of temperatures have not been presented in the literature.

The noise results presented here are given in terms of normalized drain current noise multiplied by gate overdrive squared, i.e.

$$\text{Normalized Noise Power} = (V_{gs} - V_{th})^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2}. \tag{6.3}$$

This yields a quantity which varies relatively slowly with bias. As was pointed out in Section 3.4.1, the normalized noise power is equivalent to gate-referred noise power for devices where the mobility degradation with high normal gate electric field is negligible; at low values of gate overdrive, the normalized noise power is approximately equal to the gate-referred noise. Also note that the measured noise power spectral density is presented at a frequency of 100 Hz for easy comparison with existing data in the literature. The noise values were obtained by logarithmic interpolation between the two measured data points nearest to 100 Hz. For all the data presented here, the nearest points to 100 Hz are at 87Hz and 144Hz.

6.2.1 Area Dependence

The noise power of MOSFET devices has been widely reported to vary as the inverse of the channel area. An interesting paper by Van der Ziel demonstrated using a general model of noise in the MOSFET channel region that the $1/f$ noise PSD is inversely proportional to device area independent of the noise mechanism involved, i.e.

$$\frac{S_{I_{ds}}(f)}{I_{ds}^2} \propto \frac{1}{WL}. \tag{6.4}$$

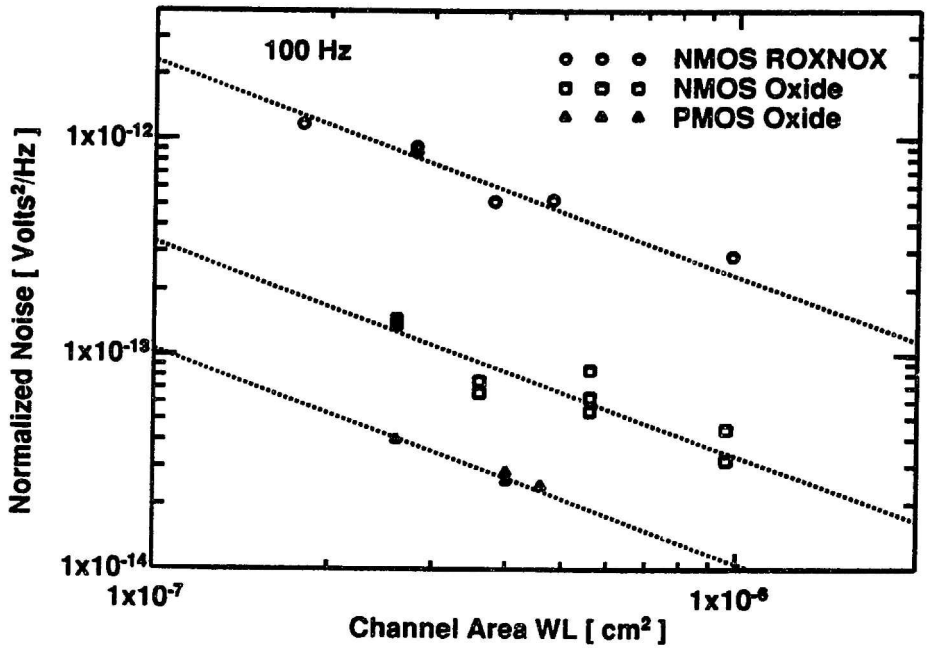


Figure 6.1: The normalized noise power of MOSFET devices plotted *versus* channel area. $|V_{gs} - V_{th}| = 1.0$ V. $|V_{ds}| = 200$ mV. T = 295 K. T_{ox} Oxide (ROXNOX) = 9 nm (11 nm).

This relationship was found to always hold for devices with long channel lengths, and to be true for noise mechanisms whose magnitude is independent of position in the channel (which is the case for the number fluctuation model) [117]. From the results presented in Chapter 5, the source of $1/f$ noise can be assumed to be uniformly distributed local sources in the channel region.

A plot of the dependence of normalized drain current noise PSD *versus* channel area is shown in Figure 6.1. Note that for all of the devices shown, the $1/f$ noise power is as expected inversely proportional to channel area.

The data reported in the subsequent sections of this thesis are for devices which have channel widths of 20 μm, and channel lengths which vary from 0.8 μm to 4.8 μm. All of the data in subsequent plots is normalized to an effective channel length of 2 μm, to aid in comparison. Most of the data plots are logarithmic plots to allow

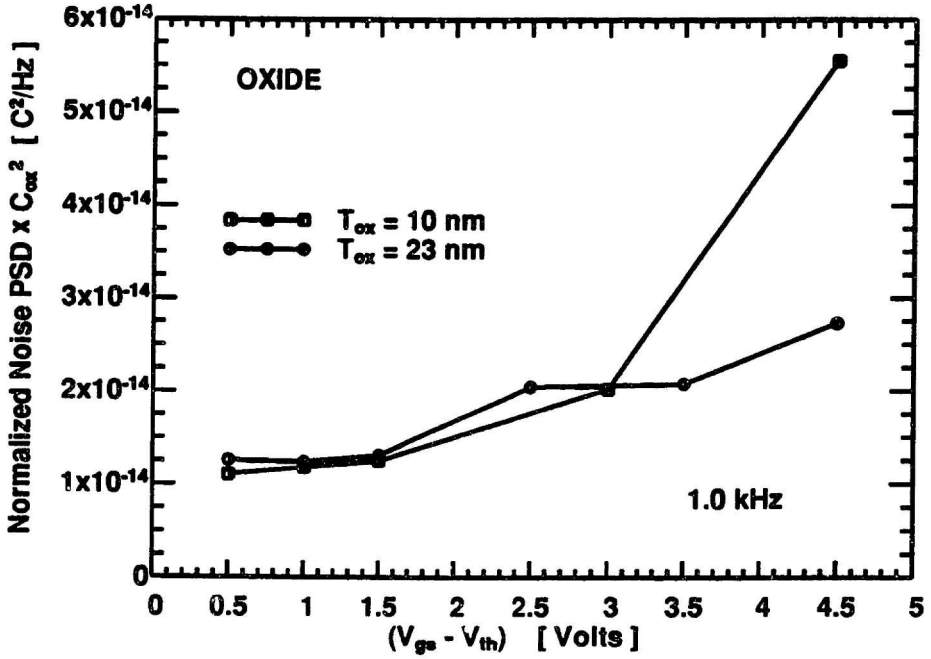


Figure 6.2: NMOS normalized noise PSD *versus* gate bias for device with two different oxide thicknesses. Drawn W/L = 20 μm/3 μm.

the proportionality of the noise PSD to gate bias and temperature to be readily determined.

6.2.2 Oxide Thickness Dependence

The $1/f$ noise PSD of MOSFET transistors has been widely shown to be proportional to C_{ox}^{-2} [3, 22, 92, 107]. This dependence has also been observed in the devices used in this work.

A plot of normalized drain current noise PSD multiplied by C_{ox}^2 for two devices with different oxide thicknesses is shown in Figure 6.2. At low values of gate bias (below $V_{gs} - V_{th} = 3$ volts for the $T_{ox} = 10$ nm devices) the two curves coincide, demonstrating the C_{ox}^{-2} dependence of the $1/f$ noise PSD. The 10 nm oxide deviates from the C_{ox}^{-2} dependence at high bias. This is believed to be due to the effect of the higher electric field in the thin oxide bending the oxide bands to bring a higher

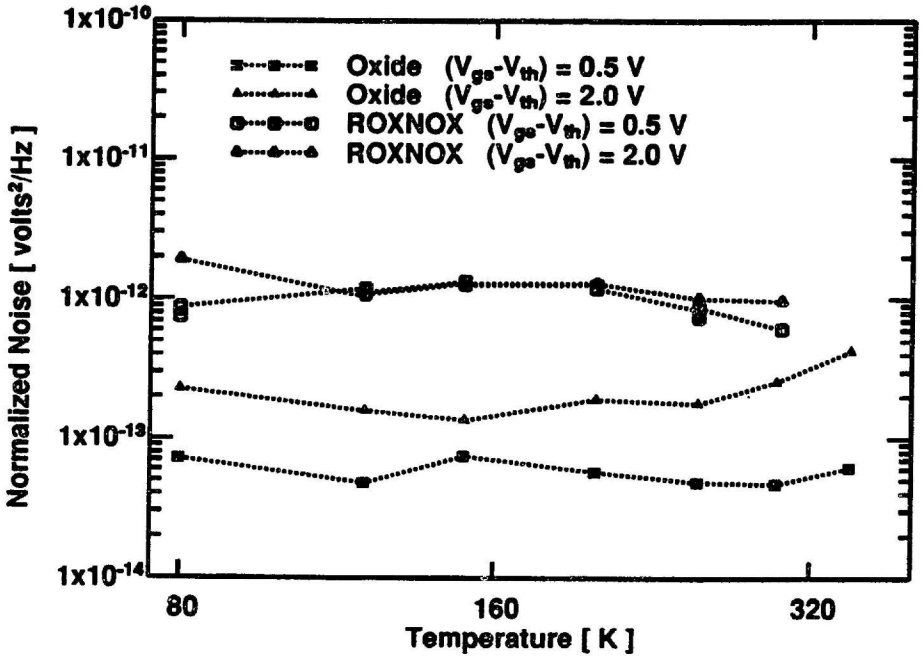


Figure 6.3: NMOS normalized noise PSD *versus* temperature. $|V_{ds}|$ = 200 mV. T_{ox} Oxide (ROXNOX) = 9 nm (11 nm). Data normalized to W/L = 20 μm/2 μm. f = 100 Hz.

density of oxide traps adjacent to the Fermi level. This effect will be discussed in the context of device scaling in Section 6.4.

6.2.3 NMOS Devices

A plot of normalized drain current noise PSD at a frequency of 100 Hz *versus* temperature is shown in Figure 6.3 for typical devices with both conventional oxide and optimized 850°C ROXNOX dielectrics. The normalized noise PSD is shown at two values of gate bias for comparison. A number of features are evident from this plot. The first is that the ROXNOX dielectrics have a higher value of noise over most of the temperature range, as was shown in the noise curves of Figure 4.9a. The other feature is that the noise PSD characteristics of both conventional oxide and ROXNOX devices are relatively bias independent. This result is similar to the result reported by Christenson et al. [4] as was shown in Figure 1.2, but is contrary to the

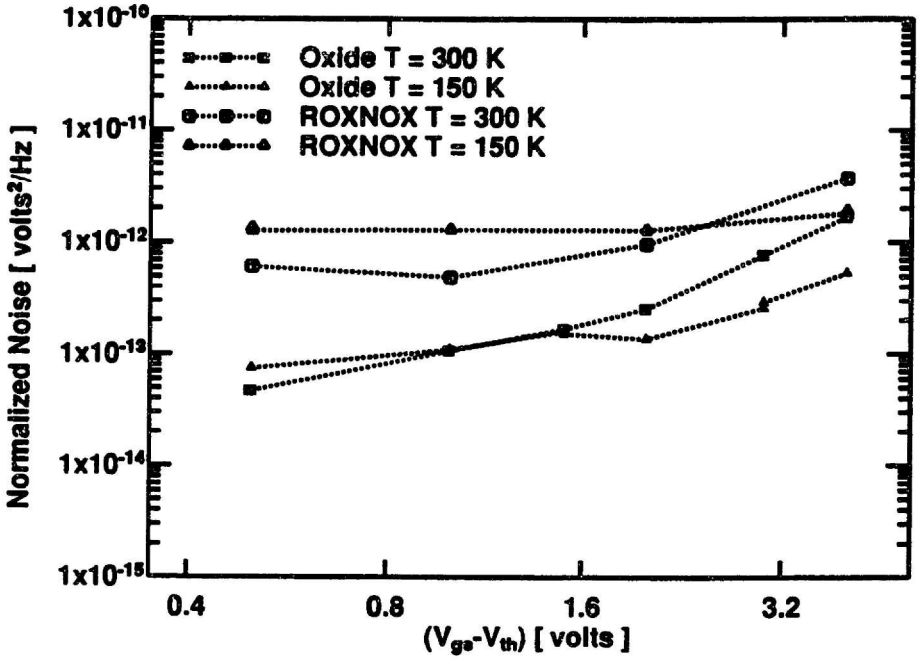


Figure 6.4: NMOS normalized noise PSD *versus* gate bias. $|V_{ds}| = 200$ mV. T_{ox} Oxide (ROXNOX) = 9 nm (11 nm). Data normalized to W/L = 20 μm/2 μm. f = 100 Hz.

model proposed by Christensson et al. in [3].

A plot of normalized noise PSD at a frequency of 100 Hz *versus* gate bias for two different temperatures is shown in Figure 6.4. At low values of gate overdrive, both the conventional oxide and ROXNOX devices exhibit a relatively constant normalized noise power spectral density, as has been observed by other researchers [22, 113]. At higher values of gate bias, the normalized noise PSD increases quite a bit, especially for the oxide device. It should be noted that at the highest values of gate overdrive shown in these plots ($V_{gs} - V_{th} = 4$ volts), the fields in the oxide are sufficiently high (≈ 4 MVolt/cm) to cause damage in the dielectric [12], which could cause an error in the noise measured at those biases.

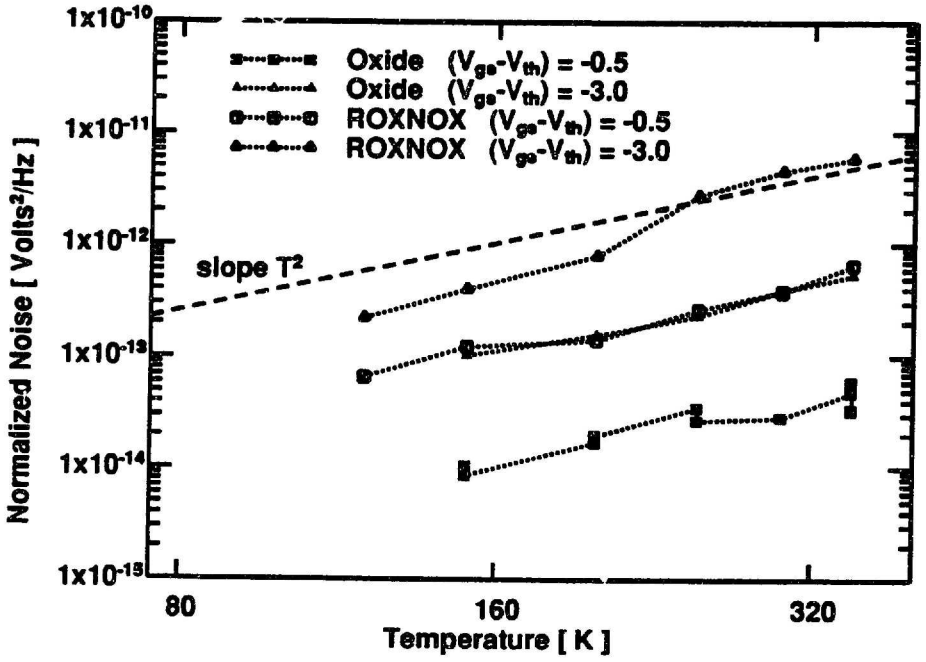


Figure 6.5: PMOS normalized noise PSD *versus* temperature. $|V_{ds}|$ = 200 mV. T_{ox} Oxide (ROXNOX) = 9 nm (11 nm). Data normalized to W/L = 20 μm/2 μm. f = 100 Hz.

6.2.4 PMOS Devices

A plot of normalized noise PSD at 100 Hz for typical PMOS devices with both conventional oxide and ROXNOX dielectrics plotted *versus* temperature are shown in Figure 6.5. Note that the PMOS noise characteristics are markedly different than those observed for the NMOS devices. At all biases characterized, the PMOS devices exhibited approximately a square law dependence on temperature over the temperature range characterized. A dashed line indicating an ideal temperature-squared dependence is superimposed on the plot for a point of reference. This was a previously unexpected result.

A plot of normalized noise PSD at 100 Hz *versus* gate overdrive for both conventional oxide and ROXNOX dielectrics is shown in Figure 6.6. The ROXNOX devices have significantly higher noise than the conventional oxide devices as expected from

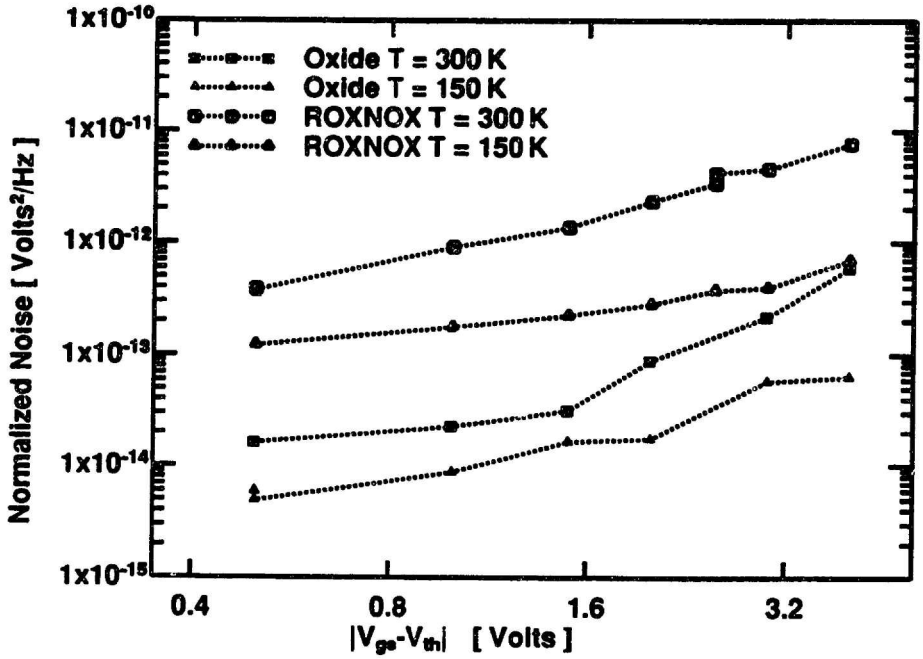


Figure 6.6: PMOS normalized noise PSD *versus* gate bias. $|V_{ds}| = 200$ mV. T_{ox} Oxide (ROXNOX) = 9 nm (11 nm). Data normalized to W/L = 20 μm/2 μm. f = 100 Hz.

the noise results shown in Figure 4.9b. The PMOS devices also have a noise PSD which monotonically increases with higher gate overdrive; unlike the NMOS devices.

6.3 $1/f$ Noise Model

There are a number of points which will be presented in this section. Initially, a general model of the contribution of individual traps to noise in a MOSFET is presented. Results of the modeling of single-electron traps from Chapter 5 are then applied to this model to obtain expected device noise performance. Reasonable assumptions about the distribution of trap activation energies, and the distribution of traps in the oxide bandgap, allow good agreement to be made between the model and the measured results. Inferring the behavior of PMOS single-trap characteristics from NMOS results allows good agreement between the model and PMOS $1/f$ noise performance to be obtained.

6.3.1 General Model

A model description involving the summing up of traps distributed throughout the channel region is presented for the case where the location of all the traps are known. This model is then generalized to a $1/f$ model which uses a probability distribution of traps in the oxide.

If the exact location and nature of every trap in the gate dielectric of a given MOSFET is known, the calculation of the $1/f$ noise of that device proceeds in a straightforward manner. Assume that for a given MOSFET there is an ensemble of N_{tot} traps in the oxide, each of which we will denote as N_i. Each trap N_i is located at energy E_i in the oxide bandgap, has activation energy $E_{a,i}$, and is located at position (x_i, y_i, z_i) using the coordinate system of Figure 6.7. Furthermore, we can assume that the effect of each single-electron trap on drain current is independent of every other trap. Each single-electron trap causes a drain current fluctuation of amplitude given by A_i and has a trap time constant given by τ_i. Note that the bias and temperature dependencies of the trap parameters have not been explicitly expressed for simplicity. We can compute the noise spectrum of the ensemble of traps as simply the sum over all the traps in the device,

$$S_{I_{ds}}(\omega) = \sum_{i=0}^{N_{tot}} A_i^2 f(E_i)[1 - f(E_i)] \frac{4\tau_i}{1 + (\tau_i \omega)^2}, \tag{6.5}$$

where $f(E_i)$ is the Fermi function for trap occupancy given by

$$f(E_i) = \frac{1}{1 + e^{\frac{E_i - E_f}{kT}}}. \tag{6.6}$$

The above Equation (6.5) is completely general. Equation (6.5) is of little practical use in its current form, since in general the exact location of every trap in the dielectric of any given conventional MOSFET is not known. An exception to this is found in the characterization of deep-submicron MOSFETs, as we have demonstrated in the

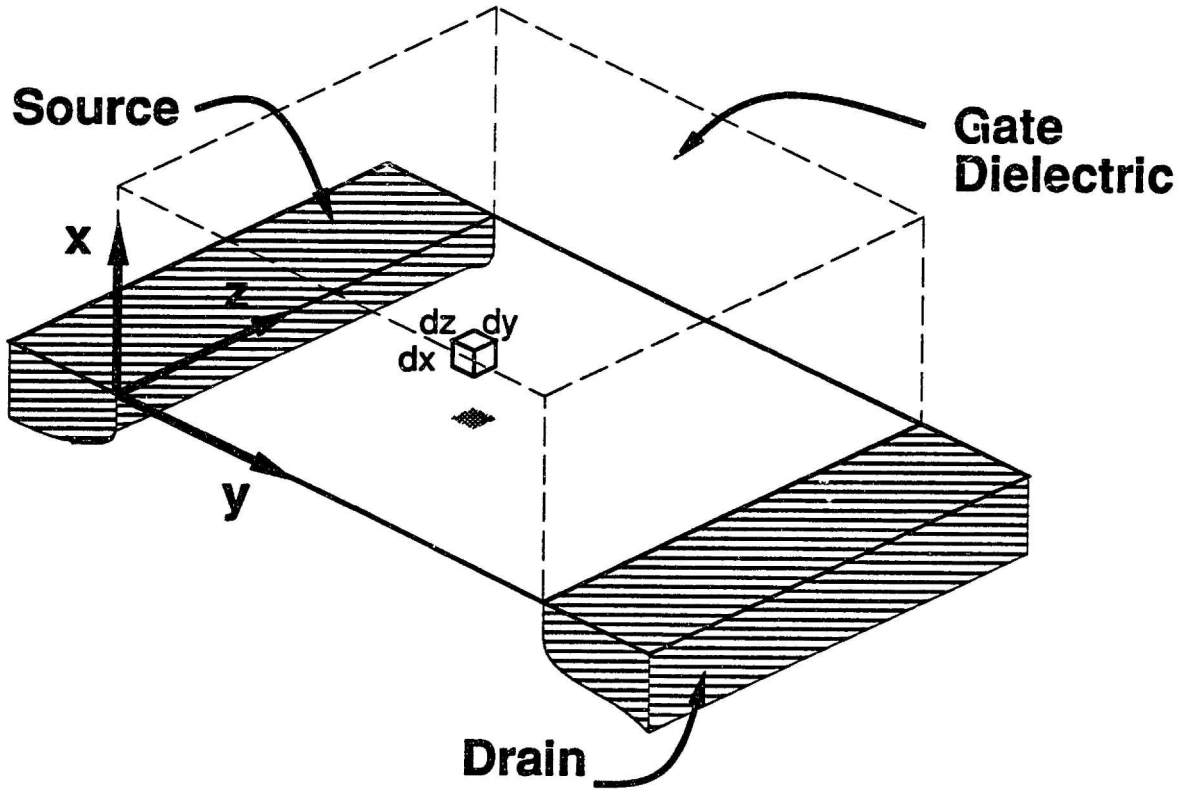


Figure 6.7: Schematic diagram of MOS transistor defining the coordinate axes and the differential volume element in the bulk of the SiO_2. The gate electrode is not shown for clarity.

previous chapter. Equation (6.5) would be applicable to a simulation of MOSFET noise which uses Monte-Carlo techniques to generate trap location and energy values.

Rather than assume particular values of trap energy and location, most workers assume a mean trap density distributed over energy and location. We can generate an integral equation for this case by writing the trap density function $N(x, y, z, E, E_a)$ for the case discussed above, where the location of every trap is known, as

$$N(x, y, z, E, E_a) = \sum_{i=0}^{N_{tot}} \delta(x - x_i)\delta(y - y_i)\delta(z - z_i)\delta(E - E_i)\delta(E_a - E_{a,i}). \quad (6.7)$$

Here, the function $\delta(x)$ is the unit impulse function [119]. For the moment, let us make the assumption that each trap will have the same amplitude given by A. This allows us to write the above Equation (6.5) as a single integral equation

$$S_{I_{d_s}}(\omega) = \int_0^L dy \int_0^W dz \int_0^{T_{ox}} dx \int_{-\infty}^{+\infty} dE \int_0^{+\infty} dE_a \; N(x, y, z, E, E_a) A^2 f(E)[1 - f(E)] S_\tau(\omega), \quad (6.8)$$

where we have made use of the definition

$$S_\tau(\omega) = \frac{4\tau}{1 + (\tau\omega)^2}. \quad (6.9)$$

Inserting the density of Equation (6.7) into the integral Equation (6.8) will recover the original summation of Equation (6.5).

Equation (6.8) is similar to the equation used by Christensson et al. [3] to model the noise performance of MOSFETs. The derivation of Christensson et al. is presented here because of the large number of papers which adopt their result to model $1/f$ noise in MOSFETs [3, 22, 90 - 93]. The noise integral of Christensson et al. is given by

$$S_Q(\omega) = \int_0^L dy \int_0^W dz \int_0^{T_{ox}} dx \int_{-\infty}^{+\infty} dE \; N(x, y, z, E) \left(\frac{q}{WL}\right)^2 f(E)[1 - f(E)] S_\tau(\omega), \quad (6.10)$$

which is an expression that does not model the effect of trap activation energy on time constant. Note that Christensson et al. use the ideal number fluctuation model, where the trap amplitude is uniformly given by

$$A = \frac{q}{WL}. \tag{6.11}$$

To evaluate the noise integral, Christensson et al. assume a uniform trap density over location in the oxide and over trap energy. Furthermore, the trap time constant was assumed to be described by the equation

$$\tau = \frac{1}{c_o N_{inv}} e^{2\alpha x}, \tag{6.12}$$

where N_{inv} is the volume density of carriers in the inversion layer, c_o is a capture rate constant, and α is a parameter for tunneling to a trap at some distance into the oxide [3]. Evaluation of this integral equation, and normalization by the inversion charge density squared, yields the dimensionless result shown in Equation (6.2).

6.3.2 Application of RTS Results

The results of characterization of single-electron-trap amplitude and time constant demonstrate that the simple models of single-electron-trap behavior assumed in the Christensson et al. derivation are significantly different than the observed single-electron-trap behavior. Combining the results of the $1/f$ noise characterization with the results of the study of single-electron traps in deep-submicron MOSFETs allows the behavior of $1/f$ noise in MOSFETs to be more accurately modeled. In this section, assumptions about the trap amplitude and time constant, based on the RTS characterization results reported in Chapter 5, will be developed for application to the $1/f$ noise model. The final form of the $1/f$ noise model is derived in the next Section 6.3.3.

Trap Amplitude

One of the important effects observed from the single-electron trap study is the nonideal behavior of single-trap amplitude, believed to be primarily due to the effect of the trapped charge on the channel mobility. Using Equation (5.39), we can write the trap amplitude A as

$$A = |\delta I_{ds}| = I_{ds}\left(\frac{1}{|Q_n|} + \mu_o S\right)\frac{q}{WL}. \tag{6.13}$$

This value of A is used in the integral of Equation (6.8) to evaluate the total MOSFET noise PSD. Using this relation and Equation (5.21), we can write the drain current power spectral density for a single-electron trap as

$$S_{RTS}(\omega) = A^2 f(E)[1-f(E)]S_\tau(\omega) = I_{ds}^2 f(E)[1-f(E)]\left(\frac{1}{|Q_n|} + \mu_o S\right)^2\left(\frac{q}{WL}\right)^2 S_\tau(\omega), \tag{6.14}$$

where $S_\tau(\omega)$ is defined in Equation (6.9).

Note that the above Equation (6.14) takes advantage of the fact that the induced mobility fluctuations are identically correlated to the inversion layer number fluctuations. Thus we can express the induced mobility fluctuations as a correlation factor multiplied by the magnitude of the ideal number fluctuation term. This is contrary to the approach taken by Mikoshiba [110] and Grabowski [111], where the mobility fluctuations are treated as being independent of the carrier number fluctuations.

The approach used with this model is a different approach than that used with the model of Jayaraman, where the induced mobility fluctuations were assumed to be much smaller than the number fluctuation contribution at low bias [22]. The model of Jayaraman can be readily inferred from the above Equation (6.14), by expanding the quantity A in a quadratic form

$$S(\omega) \propto (V_{gs}-V_{th})^2\left(\frac{1}{|Q_n|} + \mu_o S\right)^2 = \frac{1}{C_{ox}^2} + \frac{(V_{gs}-V_{th})\mu_o S}{C_{ox}} + (V_{gs}-V_{th})^2\mu_o^2 S^2, \tag{6.15}$$

where the multiplication by $(V_{gs} - V_{th})^2$ is from the noise normalization used here, and in [22]. The assumption of the model used by Jayaraman is that the value of the scattering parameter is essentially bias independent, and that the value of $\mu_o S$ is of the same magnitude as $1/C_{ox}$; thus at low bias levels, the last term in the above equation can be neglected, and the second term contributes only a small amount [22]. The characterization of single-electron traps has demonstrated that the scattering parameter S is strongly bias dependent, and that the quantity $\mu_o S$ is relatively much larger than $1/C_{ox}$ for NMOS devices.

As was mentioned before, from the analysis of Stern and Howard, the scattering coefficient is approximately inversely proportional to $|Q_n|$. Although there is a great deal of spread in the measured scattering parameter values shown in Figure 5.8, the value of S can be estimated to be

$$S \sim \frac{S_o}{|Q_n|} \frac{\text{V} \cdot \text{sec}}{\text{coulomb}}, \tag{6.16}$$

where S_o is of the order of 10^{-2}. This relation fits both the oxide and ROXNOX single-electron-trap data fairly well.

Trap Time Constant

Another important feature observed in the single-electron trap characterization was the existence of a strong activation energy for the trap time constant. Furthermore, this activation energy is apparently independent of the distance the trap is away from the interface. This is in contrast to the model proposed in Christensson et al. where the primary variation in trap time constant was due to the variation of trap depth in the dielectric [3]. It is obvious that tunneling to traps in the oxide becomes increasingly unfavorable as the distance of the traps is made greater, and so there must be some dependence of trap time constant on depth. From the characterizations

made in Chapter 5, however, the depth of the trap into the oxide is not the primary explicit determining factor of the trap time constant.

In applying the results of the single-electron-trap characterization to the conventional device model, the individual traps in the dielectric will be treated as a form of interface state; the traps lie in the oxide distributed over some region x_o near the interface, but the depth distribution has no explicit effect on trap time constant. The activation energies for capture and emission distinguish these oxide traps from actual interface states. Note that the trap depth *is* important to determine the effective density of traps adjacent to the Fermi level. This assumption is expressed by approximating the integral over x in Equation (6.8) as

$$\int_0^{T_{ox}} dx\, N(x,y,z,E,E_a) \simeq x_o N(x,y,z,E,E_a)|_{x=0}\,. \tag{6.17}$$

This assumption will be validated at the end of Section 6.3.6. For the purposes of the model derivation, it is assumed that the trap energy level E, trap activation energy E_a, and trap depth x are all independent quantities. Although no explicit dependence of these quantities on one another is observed in the single-electron-trap characterization data, it is expected that the physical structure of the trap will cause these three parameters to be related. It is fully expected that the depth of the trap into the oxide can be treated as a hidden variable, for example.

Another interesting feature observed from the deep-submicron device characterization is that the trap time constant is a much stronger function of bias than is expected from the model used by Christensson et al. [3]. This effect of bias will enter in the calculation of the effective trap density in the oxide at the same energy as the silicon Fermi level $N_t(E_f)$, which will be discussed shortly. For a device with a uniform density of traps in the oxide bandgap and a uniform distribution of activation energies, the trap bias dependence causes a negligible difference in the estimation of

$1/f$ noise behavior of MOSFET devices.

Other Assumptions

It is difficult to determine the distribution of traps in energy, activation energy, and space without measuring a huge number of devices to obtain a statistically significant sample. However, a rough approximation of the trap density can be obtained from the device areas listed in Table 5.1. The channel area of the devices for which single-electron traps are observed is about 0.4 μm^2 for oxides, and about 0.2 μm^2 for ROXNOX dielectrics. Considering that only traps within a few kT of the Fermi energy are active at any time, the density of traps at the interface for oxide devices can be estimated to be on the order of $2–5 \cdot 10^9$ cm^{-2} eV^{-1}, which is close to the right order of magnitude for devices with high quality interfaces. The traps contributing to $1/f$ noise will be assumed to be uniformly distributed throughout the channel. This assumption allows us to write the integral of over L and over W in Equation (6.8) as simply a multiplication by LW.

The distribution of traps in energy in the oxide bandgap and the distribution of trap activation energies are not readily apparent from the RTS data. The characterization results of Chapter 5 demonstrated that the activation energies of single-electron traps can range in value from less than 100 meV to greater than 600 meV. Because of this broad range of measured activation energies, the trap density will be assumed to be uniformly distributed over activation energy from $E_a = 0$ to $E_a = +\infty$. A more physical upper bound for the distribution over activation energy would be on the order of 1 eV [29]. Assuming an upper bound of 1 eV produces a negligible difference in the final noise calculation from assuming that the traps are uniformly distributed up to $E_a = +\infty$. In this derivation, the traps are initially assumed to be uniformly distributed in energy in the oxide bandgap. It is well known that the density of con-

ventional interface states is highest near the band edges of silicon [11]. Because of this, the possibility of a trap energy density which increases with energy away from midgap will also be discussed.

The device area is assumed to be large enough such that the discreteness of the trap capture and emission is not significant. The limits of validity of this assumption will be examined in more detail in Section 6.4.1. Each single trap in the channel will be assumed to have an identical amplitude to every other trap, given by Equation (6.13) above.

It was mentioned previously that the characterization of deep-submicron devices could have introduced an unintentional bias toward the largest amplitude single-electron traps. It is entirely possible that traps which have a much smaller induced mobility fluctuation effect, and therefore a much lower single-trap amplitude, are also active in the device channel. It should be noted that the contribution of a single trap's fluctuations to total device noise power is proportional to that trap's amplitude squared. The net result of this is that even if there is some distribution of trap amplitudes at values lower than the values measured here, it will still be the largest traps which dominate device noise performance. The rough estimate of trap density above, giving reasonable interface trap densities for the large-amplitude traps characterized, reinforces the idea that the single-electron-trap characteristics reported in Chapter 5 will dominate the $1/f$ noise performance.

The same qualitative behavior observed for NMOS single-electron traps will be assumed for the PMOS single-hole traps for application to PMOS devices. We do not have single-trap characterization results for PMOS devices, so we have no *a priori* information about the relative magnitude of the hole scattering cross section of traps near the PMOS channel region. From the measured $1/f$ noise data, we will be able to infer information about the hole scattering cross section.

6.3.3 Full Model

Using the assumption of a uniform trap area density we can define the trap density $N_o(E, E_a)$ as

$$x_o WL\, N_o(E, E_a) = \int_0^{T_{ox}} dx \int_0^L dy \int_0^W dz\, N(x, y, z, E, E_a). \tag{6.18}$$

Using this result and the assumptions from the previous section we can rewrite Equation (6.8) as

$$S_{I_{ds}}(\omega) = x_o WL \int_{-\infty}^{+\infty} dE \int_0^{+\infty} dE_a\, N_o(E, E_a) f(E)[1 - f(E)]\, I_{ds}^2 \left(\frac{1}{|Q_n|} + \mu_o S\right)^2 \left(\frac{q}{WL}\right)^2 S_\tau(\omega). \tag{6.19}$$

In the evaluation of the integral over activation energy, we can simplify things by writing the trap time constant as

$$\tau = \tau_o e^{\frac{E_a}{kT}}, \tag{6.20}$$

where τ_o is a complex function of gate-bias, and temperature (see Equation (5.66)). We can replace the integration over E_a in Equation (6.19) with an integration over trap time constant τ, noting that for $E_a = 0$ eV the trap time constant is very small ($\tau = \tau_o$), and for the limit of $E_a \to \infty$, the trap time constant approaches $+\infty$. To complete the change of integration variable, we make the substitution

$$dE_a = \frac{kT}{\tau} d\tau. \tag{6.21}$$

Inserting this into Equation (6.19) and using the expression for $S_\tau(\omega)$ of Equation (6.9) we can evaluate the integral over the trap time constant as

$$\begin{aligned} S_{I_{ds}}(\omega) &= \frac{q^2 x_o I_{ds}^2}{WL} \left(\frac{1}{|Q_n|} + \mu_o S\right)^2 \int_{-\infty}^{+\infty} dE\, N_t(E) f(E)[1 - f(E)] \int_{\tau_o}^{+\infty} d\tau\, \frac{kT}{\tau} \frac{4\tau}{1 + (\omega\tau)^2} \\ &\simeq \frac{q^2 x_o I_{ds}^2 kT}{WLf} \left(\frac{1}{|Q_n|} + \mu_o S\right)^2 \int_{-\infty}^{+\infty} dE\, N_t(E) f(E)[1 - f(E)]. \end{aligned} \tag{6.22}$$

In the evaluation of this integral we have made use of the fact that $\tau_o\omega \ll 1$ for all frequencies ω in the band of interest. This is a valid assumption considering that for all the single-electron traps characterized in this work τ_o was much less than nanoseconds. It is worth pointing out that the details of the variation of trap time constant with bias do not affect this integration for the case discussed here of a trap population with uniformly distributed activation energies.

If the trap density $N_t(E)$ varies only slowly with bias, the integral over trap energy in the midgap E can be approximated as [4, 22]

$$\int_{-\infty}^{+\infty} dE\ N_t(E)f(E)[1-f(E)] \simeq 4kTN_t(E_f). \tag{6.23}$$

Using this approximation, and dividing both sides by drain current squared, we can write Equation (6.22) as

$$\frac{S_{I_{ds}}(f)}{I_{ds}^2} = \frac{4q^2x_o(kT)^2N_t(E_f)}{WLf}\left(\frac{1}{|Q_n|}+\mu_oS\right)^2. \tag{6.24}$$

Note that this expression is proportional to $1/f$, and inversely proportional to channel area as expected from the measurement results, and from the literature.

If the induced mobility fluctuation term is neglected, i.e. if

$$\mu_oS \ll \frac{1}{|Q_n|}, \tag{6.25}$$

the noise PSD is inversely proportional to inversion charge squared as expected from the results in the literature, and from the measured results. Note also for this case that the noise PSD is proportional to temperature squared! This is a result which agrees with the measured results obtained on the PMOS devices.

If we use the estimate of the induced mobility fluctuation coefficient S from Equation (6.16), Equation (6.24) can be simplified to be

$$\frac{S_{I_{ds}}(f)}{I_{ds}^2} = \frac{4q^2x_o(kT)^2N_t(E_f)}{WL|Q_n|^2f}(1+\mu_oS_o)^2. \tag{6.26}$$

Multiplying both sides of this equation by $(V_{gs} - V_{th})^2$ yields normalized drain current noise PSD and allows direct comparison to the measured data as

$$\text{Normalized Noise PSD} = (V_{gs} - V_{th})^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2} = \frac{4q^2 x_o (kT)^2 N_t(E_f)}{WLC_{ox}^2 f} (1 + \mu_o S_o)^2 . \tag{6.27}$$

This expression is proportional to C_{ox}^{-2}, as is observed in the measured data. This is the final form of the model for $1/f$ noise in N-MOSFETs.

The above model Equation (6.27) also applies to PMOS devices as can be readily determined from the symmetry of the derivation. Even though single-hole-trap devices were not characterized for this thesis, it is safe to assume that the behavior of single-hole traps is similar to that of single-electron traps. Discussion of the magnitude of the scattering coefficient for holes is deferred to the next section.

6.3.4 Model Discussion

For the model of Equation (6.27) the only bias and temperature dependent quantities are the channel mobility μ_o, and the trap density $N_t(E_f)$ which varies with the variation of E_f with gate bias.

Temperature Variation

Two cases are considered in this section; 1) the condition that $\mu_o S_o \gg 1$, and 2) the condition that $\mu_o S_o \ll 1$. For case 1), the above Equation (6.27) can be written as

$$\text{Normalized Noise PSD} = \frac{4q^2 x_o (kT)^2 N_t(E_f)}{WLC_{ox}^2 f} (\mu_o S_o)^2 . \tag{6.28}$$

Assuming a uniform trap density $N_t(E_f)$, the only terms which vary with temperature are μ_o and T. Using the $T^{-1.5}$ dependence of mobility on temperature shown in

Figure 4.7, the variation of normalized noise PSD can be written as

$$\text{Normalized Noise PSD} \sim T^2(T^{-1.5})^2 = \frac{1}{T}. \tag{6.29}$$

This dependence will apply as long as $\mu_o S_o \gg 1$. At high gate biases, the mobility is reduced significantly due to the effects of high normal field. Also, at high temperatures, the mobility is reduced due to increased phonon scattering. For these reasons, it is expected that the dependency of Equation (6.29) will apply particularly at low bias and low temperatures. This term can be thought of as the noise power caused by induced mobility fluctuations.

Stated another way, the term $\mu_o S_o$ can be said to represent induced mobility fluctuations. When the mobility is highest the effect of adding a single scatterer to the channel will be largest; hence the induced mobility fluctuations will have a dominant effect. At high bias and high temperature, the mobility is reduced due to the increased effect of other scattering mechanisms in the channel, and the effect of a single Coulombic scatterer is relatively much less.

For the second case, where $\mu_o S_o \ll 1$, the model Equation 6.27 can be written as

$$\text{Normalized Noise PSD} \simeq (V_{gs} - V_{th})^2 \frac{S_{I_{ds}}(f)}{I_{ds}^2} = \frac{4q^2 x_o (kT)^2 N_t(E_f)}{WLC_{ox}^2 f}. \tag{6.30}$$

The variation of this term with temperature can be readily seen to be T^2. This dependence will apply as long as $\mu_o S_o \ll 1$. This condition holds when the effective mobility μ_o is low, or when the scattering parameter S_o is low. This term can be thought of as the noise power caused by only number fluctuations.

At values of $\mu_o S_o \sim 1$, both number and induced mobility fluctuations are important, and the full Equation (6.27) must be used.

Gate Bias Variation

In Equation (6.27) these is no explicit gate bias variation, as expected from the use of the normalized noise PSD. The normalized noise PSD will vary due to the variation of μ_o and of $N_t(E_f)$ with gate bias. As the gate bias changes, the position of the Fermi level relative to the energy of the traps in the oxide is shifted. This causes any non-uniformity in trap density to be reflected as a variation of noise power with gate bias. Further discussion of the variation of noise with gate bias is deferred to the trap extraction discussion in Section 6.3.6.

6.3.5 Comparison to Measured Results

From the data plots of PMOS noise *versus* temperature shown in Figure 6.5, the T^2 dependence of normalized noise PSD is evident. This data is discussed first, because of the simple structure present in the data. This data indicates that the primary mechanism responsible for noise in PMOS devices is the number fluctuation mechanism. Induced mobility fluctuations are negligible in this device.

The behavior of NMOS devices is expected to be dominated by induced mobility fluctuations at low bias from the behavior of single electron trap devices reported in Chapter 5.

The data of Christensson et al. plotted in Figure 1.2 is replotted in Figure 6.8 with the model curves described by Equation (6.27). Note that the NMOS device noise is dominated by induced mobility fluctuations, indicated by the inverse temperature dependence. The PMOS device behavior is slightly more complicated than this. At low temperatures ($\sim$ 100 K) the device mobility is high enough such that induced mobility fluctuations are significant. At higher temperatures, the mobility is reduced so that number fluctuations dominate the device noise characteristic. Note the excellent fit of the noise model to this data. This result indicates that the trap

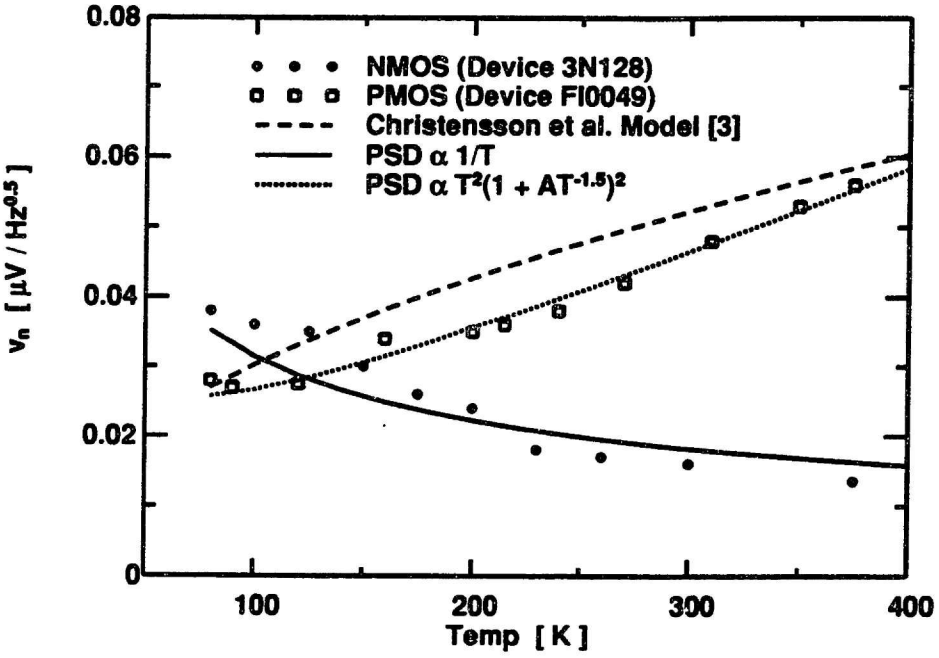


Figure 6.8: Comparison of data from Christensson et al. [4] with the model of Equation (6.27).

scattering cross section for PMOS devices is a least an order of magnitude lower than the measured scattering cross section of NMOS single traps.

The oxide noise data of Figure 6.3 is replotted in Figure 6.9, along with an additional noise curve for the gate bias of $(V_{gs} - V_{th}) = 4.0$ volts. The dashed lines indicate the model curves for comparison. The scattering parameter for this case was assumed to be $S = 0.02/|Q_n|$, which is consistent with the single electron trap characterization results. At low bias and low temperature, the noise characteristic is dominated by induced mobility fluctuations, as evidenced by the relatively flat variation of noise power with temperature. At high bias and temperature, the effect of induced mobility fluctuations are minimized and the noise power characteristic approaches the T^2 dependence expected for the ideal number fluctuation noise mechanism.

Note that the high doping level in the modern devices reported here will likely cause the mobility to be lower than for the devices characterized by Christensson et

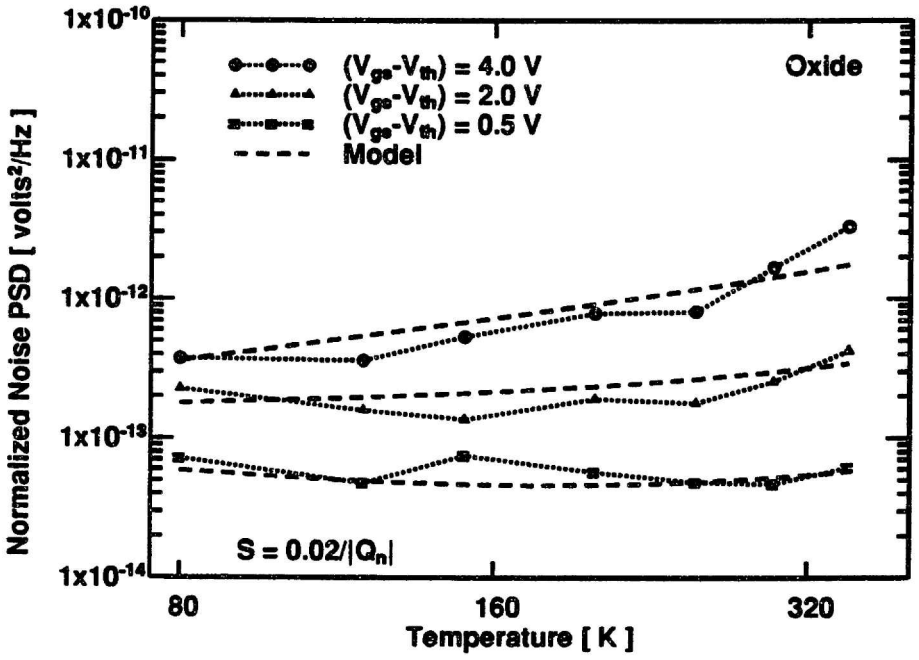


Figure 6.9: Comparison of NMOS noise data with the model of Equation (6.27). The dashed lines indicate the model curves.

al. This would explain why the induced mobility fluctuations appear to be slightly more important for the NMOS and PMOS devices reported by them [4].

6.3.6 Trap Distribution

Using the known bias dependencies of the quantities in Equation 6.27, the effective interface-trap density can be solved for by using the measured data from Figures 6.4 and 6.6 as

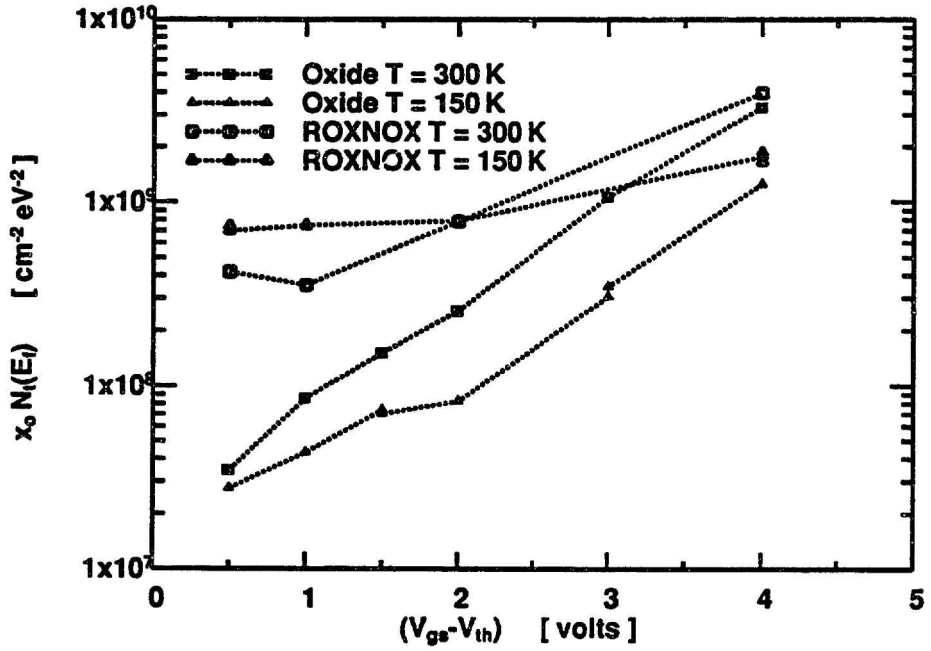
$$x_o N_t(E_f) = \frac{WLC_{ox}^2 f}{q^2(kT)^2(1+\mu_o S_o)^2}(\text{Noise PSD}), \tag{6.31}$$

where S_o is assumed to be 0.02 for the NMOS devices and zero for the PMOS devices, and $f = 100$ Hz to correspond to the data in the figures. The scattering coefficient for PMOS is assumed to be negligible because the variation of PMOS noise power with temperature indicates that the noise characteristic is dominated by number fluctuations only.

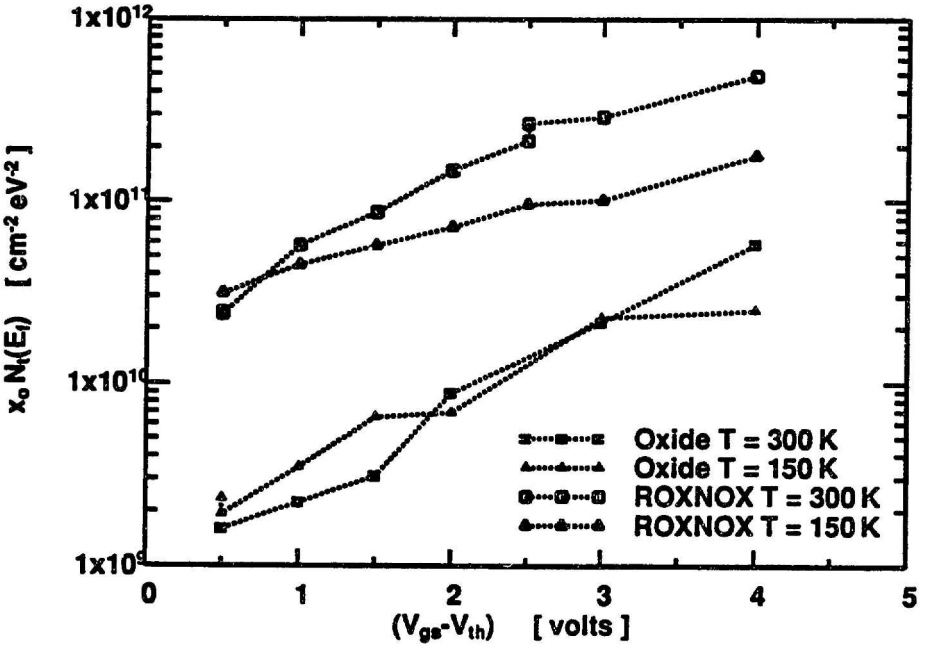
The extracted effective interface-trap density $x_o N_t(E_f)$ is plotted *versus* gate bias in Figure 6.10. Note that for both the NMOS devices and PMOS devices in the figure, there is a strong increase in effective trap density with gate bias. Also note that the extracted value of PMOS trap density is much higher than for the NMOS devices. The extracted NMOS density is extremely sensitive to the estimate of the scattering parameter S_o which could introduce a significant error in the extraction of NMOS trap density. The effective trap density also has a small temperature dependence; note that the extracted trap density in Figure 6.10 is not an *actual* trap density, but a weighted average computed from Equation (6.8). The extraction of trap density will be discussed further below.

The net result of the data in this plot is that there is a higher density of hole traps adjacent to the valence band than there are electron traps adjacent to the conduction band. This result reinforces the significance of the dominant mobility fluctuations in NMOS devices for affecting MOSFET noise performance. For the case of NMOS devices, fewer oxide traps account for more noise power than found in the PMOS devices.

The large activation energies observed for the trap time-constants for the single-electron-trap devices have important consequences for MOSFET noise behavior. Because the time constant for a given trap is reduced so much as the temperature is lowered, the traps which determine the noise performance of the device at one temperature are moved out of the band of interest at another frequency. An example of this behavior is shown in Figure 6.11. Traps important for the $1/f$ noise measurements reported in this thesis at any given temperature will have activation energies which lie between the two dotted lines in the figure. Note that a completely different set of traps are active at low temperatures than the traps which are active at high temperatures.



(a) NMOS Devices.



(b) PMOS Devices.

Figure 6.10: Effective trap density *versus* gate bias extracted from $1/f$ noise measurements. Scattering coefficient assumed to be $S = 0.02/|Q_n|$ for NMOS, and $S = 0$ for PMOS.

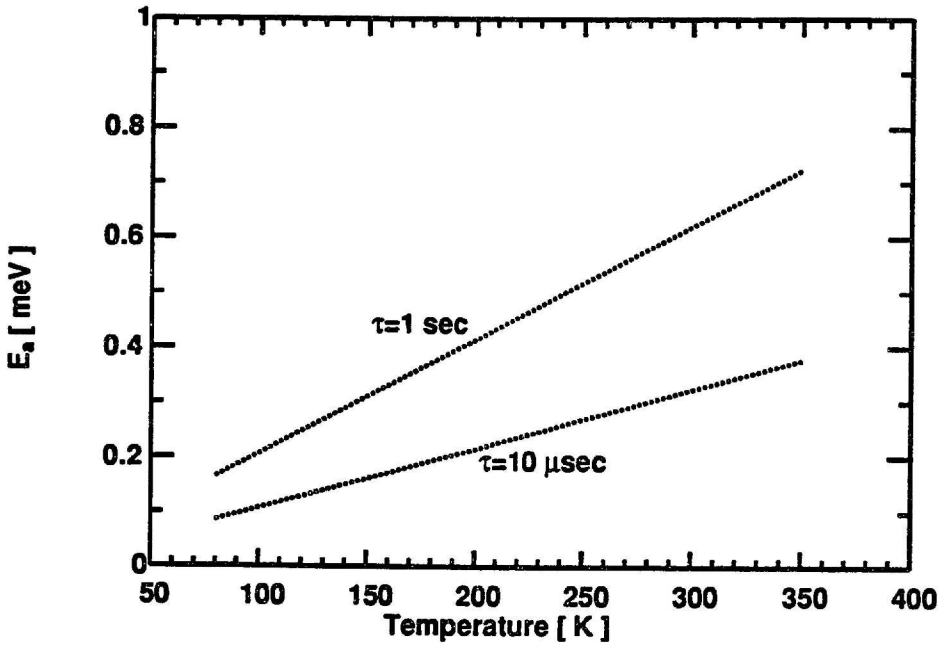


Figure 6.11: Trap activation energy *versus* temperature for a constant time constant. Traps which are active in the band of interest at any given temperature will have activation energies which lie between the two lines.

If the trap density is a decreasing function of activation energy, the $1/f$ noise of the device is expected to be relatively higher at lower temperature.

We can quantify the assumptions used to model the variation of $N_t(E_f)$ with gate bias and temperature made in the evaluation of the noise integral in Equations (6.17)-(6.23). The assumptions made can be expressed as

$$N_t(E_f)\frac{4x_o(kT)^2}{f} \simeq WL\int_0^{T_\infty} dx \int_{-\infty}^{+\infty} dE \int_0^{+\infty} dE_a\, N(x,y,z,E,E_a)\, f(E)[1-f(E)]\, S_\tau(\omega). \tag{6.32}$$

For the case of a trap density $N(x,y,z,E,E_a)$ assumed to be uniform over space, uniform over activation energy, and slowly varying in energy in the bandgap, this is equivalent to writing

$$N_t(E_f) \simeq N(x,y,z,E,E_a)|_{E=E_f}\,. \tag{6.33}$$

The effective trap density $N_t(E_f)$ for the general case of a trap density non-uniform

in depth x into the oxide and non-uniform in activation energy can be found directly from Equation (6.32)

$$N_t(E_f) = \frac{WLf}{4x_o(kT)^2} \int_0^{T_{ox}} dx \int_{-\infty}^{+\infty} dE \int_0^{+\infty} dE_a \, N(x, y, z, E, E_a) \, f(E)[1 - f(E)] \, S_r(\omega). \tag{6.34}$$

Given a trap density which varies with activation energy and with trap energy in the bandgap, the effective trap density for use in the $1/f$ noise can be determined from this expression. Evaluating this expression for an assumed oxide trap distribution will account for all of the variation of noise power with bias and temperature due to trap non-uniformity.

In general, the noise power spectral density is not strictly proportional to $1/f$, but is proportional to $1/f^{\gamma}$ where γ is in the range of 0.75-1.25 for the devices characterized in this work. If the trap densities used in evaluating the noise integral of Equation (6.19) are not uniform in activation energy and in energy in the oxide bandgap, a frequency dependence of other than $1/f$ will be obtained. For example, if the density of oxide traps decreases with higher activation energies, there will be less traps with long time constants active at any given temperature than traps with short time constants. This will have the effect of causing a relatively higher amount of noise power than expected from a strict $1/f$ dependence to exist at high frequencies, resulting in a slope factor γ of less than 1. Non-uniformity of trap density in energy in the bandgap can exaggerate the effect of non-uniformity in trap activation energy (and trap depth in the oxide, discussed below) on the frequency dependence of the noise PSD. Because of the large number of assumptions required to obtain concrete information from the $1/f$ noise slope factor γ, this topic will not be explored in this thesis. It should be noted that any assumptions about the effect of a non-uniform trap density on γ can be rigorously tested by performing the integrations of Equation

(6.34) above with no simplifying assumptions.

As long as the trap distribution in activation energy and in energy in the bandgap does not vary rapidly with energy, the temperature dependencies found in Equation (6.27) are expected to be valid. The primary effect of non-uniform trap density is a perturbation of the frequency dependence from being strictly $1/f$.

In the preceding development, the effect of trap depth in the oxide x on trap time constant was assumed to be negligible. What effect on the $1/f$ noise PSD would be observed if the trap time constants possessed a significant variation with x? Let us assume that the trap time constant is given by

$$\tau = \tau_o e^{\alpha x} e^{\frac{E_a}{kT}}, \tag{6.35}$$

where τ_o is $\sim 10^{-12}$ to correspond to the traps characterized in Chapter 5, and α is a tunneling parameter of value $\sim 2\times10^8$ cm^{-1} [22]. Using this expression for the time constant and using the substitution $dE_a = (kT/\tau)d\tau$ to effect a change of variables, we can write the integral over E_a and x as

$$\int_0^{T_{ox}} dx \int_0^{+\infty} dE_a \, N_t \frac{4\tau}{1+(\tau\omega)^2} = \int_0^{T_{ox}} dx \int_{\tau_o e^{\alpha x}}^{+\infty} d\tau \, N_t \frac{4kT}{1+(\tau\omega)^2}. \tag{6.36}$$

Here, N_t is a generalized effective trap density. The integral over τ and over x can be evaluated to be

$$\begin{aligned}\int_0^{T_{ox}} dx \int_{\tau_o e^{\alpha x}}^{+\infty} d\tau \, N_t \frac{4kT}{1+(\tau\omega)^2}, &= N_t \int_0^{T_{ox}} dx \frac{4kT}{\omega}\left[\frac{\pi}{2} - \tan^{-1}(\omega\tau_o e^{\alpha x})\right] \\ &\simeq \frac{kTN_t}{\alpha f}\ln\left(\frac{1}{\omega\tau_o}\right).\end{aligned} \tag{6.37}$$

This expression is valid as long as

$$T_{ox} > \frac{1}{\alpha}\ln\left(\frac{1}{\omega\tau_o}\right), \tag{6.38}$$

which is satisfied in the frequency band of interest. Because of the small value of τ_o, the logarithm term in the Equation (6.37) is only slowly varying with frequency,

causing a small perturbation to the $1/f$ frequency dependence. The noise integral can be approximated in the band of interest as

$$\int_0^{T_{ox}} dx \int_0^{+\infty} dE_a \, N_t \frac{4\tau}{1+(\tau\omega)^2} \simeq \frac{25}{\alpha} \frac{kTN_t}{f^{1.05}}. \tag{6.39}$$

An alternative approximation can be made by neglecting the small variation of the logarithm term with frequency to obtain an expression equivalent to Equation (6.17),

$$\int_0^{T_{ox}} dx \int_0^{+\infty} dE_a \, N_t \frac{4\tau}{1+(\tau\omega)^2} \simeq 15\text{Å} \int_0^{+\infty} dE_a \, N_t \frac{4\tau}{1+(\tau\omega)^2}\bigg|_{x=0}. \tag{6.40}$$

This result confirms that the approximation of Equation (6.17) is valid even if the trap time constant has a significant variation with depth.

6.3.7 Model Summary

The behavior of the trap scattering coefficient S has been characterized and the results reported in the previous chapter. Although previous workers have reported a relatively strong temperature dependence of the scattering coefficient [66], this dependence is absent in our single-electron-trap device measurements. The scattering coefficient S is also inversely proportional to the inversion charge area density $|Q_n|$, due to the effect of screening on reducing the momentum scattering cross section of a Coulombic scatterer.

Noise in the drain current of a MOSFET is determined by a combination of the effects of the number of channel carriers fluctuating and the mobility of the carriers fluctuating. The mobility fluctuation is induced by the fluctuating occupancy of traps which act as Coulombic scattering sites when they capture a carrier.

The induced mobility fluctuations dominate device noise behavior when the channel mobility is very high. This is due to the fact that when the mobility is high, there are few scattering mechanisms active in the channel, so that the addition of a single

Coulombic scatterer (the trapped charge) will have a relatively large effect on channel current. Conversely, when the channel mobility is low, induced mobility fluctuations will have a relatively small effect.

The NMOS devices characterized in this work have noise behavior which, at the lowest gate biases, is determined by a combination of the induced mobility fluctuations and pure number fluctuations. This results in a noise *versus* temperature characteristic which exhibits only a weak temperature dependence. At high biases and the highest temperatures, the NMOS noise behavior approaches that expected for noise due to ideal number fluctuations.

The noise behavior of the PMOS devices characterized in this work is well described as a pure number fluctuation. This is due to the combination of the channel hole mobility being lower than the electron mobility, and the scattering coefficient S for hole traps being lower than the scattering coefficient for electron traps. Characterization of deep-submicron PMOS devices to measure single hole traps is required to confirm this result.

Trap density at energies further away from the midgap point are higher for both PMOS and NMOS devices. This accounts for the bias dependence of $1/f$ noise in both types of devices. It should be noted that some of the bias dependence of normalized noise power in previous work can be attributed to the use of input-referred noise to report noise data, rather than normalized noise power [110]. The density of hole traps adjacent to the valence band is higher than the density of electron traps adjacent to the conduction band, underscoring the effect of induced mobility fluctuations; for NMOS devices at most bias and temperature, a lower density of traps yields higher noise than observed in the PMOS devices.

6.4 Circuit Ramifications

This section begins by dealing with the limits of the $1/f$ noise model encountered with device scaling. Any model development in the study of $1/f$ noise should keep in mind the overall purpose; understanding and improving the operation of circuits. The final part of this section will deal with how what we have learned in this study will apply to circuit design.

6.4.1 Device Scaling

The final noise model can be written in a simplified form as

$$\text{Noise PSD} = \frac{KN_t(E_f)}{WLC_{ox}^2 f}, \tag{6.41}$$

where K is a bias and temperature dependent parameter which can be determined from Equation (6.27). Writing the model in this form will allow us to examine the behavior of the normalized noise PSD as the device dimensions are scaled.

Two regions of bias will be considered; the case where the quiescent gate-bias voltage is low (where the resultant oxide electric field is less than <2 MV/cm) and the case where the quiescent gate-bias voltage is high (resultant oxide electric field >3 MV/cm). It is important to understand both regions of behavior, because the devices have slightly different behavior in each region. The low bias region is expected to be most applicable to devices used as current sources or in a small-signal amplifying role. The high bias case is expected to be applicable to devices operated as switches where the gate-to-source voltage V_{gs} can swing from rail to rail.

In the low bias operating region, as the oxide thickness is changed by a factor κ, the $1/f$ noise PSD of the device is changed by a factor of κ^2.

In the high bias operating region there is a significant bending of the oxide bands. Scaling the oxide thickness by a factor of κ under constant applied V_{gs} will cause

the oxide electric field to change by a factor of κ^{-1}. As the oxide electric field is increased, the density of traps adjacent to the Fermi level is also increased, as can be seen from the plots of Figure 6.10. This increase in trap density will partially offset the increase in C_{ox} as the oxide thickness is scaled, causing decreasing returns to scale. This behavior is evident from Figure 6.2, where the 10 nm oxide has a higher value of $1/f$ noise at high bias than is expected from the effect of the increase in C_{ox}.

Scaling the channel area to lower values will in all cases result in a higher expected value of $1/f$ noise for a given MOS device. Note that the value of $1/f$ noise given by Equation (6.41) is the mean noise value expected for a device of given geometry, the actual noise value of any given device can vary from this value. A derivation of the estimated variation in the number of traps active in a given bandwidth in the device channel is given in Appendix D. This derivation demonstrates that the fractional variation from the mean value of the device noise of a given device can be written as

$$\frac{S(f)}{<S(f)>} = \pm\frac{1}{WLN_t(E_f)}, \tag{6.42}$$

where the quantity $<S(f)>$ is the mean noise value at a given frequency. Using the measured value of $N_t(E_f)$ of $10^9 - 10^{10}$ cm^{-2}eV^{-1} for device geometry on the order of 10×10 μm, the fractional variation of noise from the mean for a given device is found to be on the order of a ±5%. Scaling the device geometries to be on the order of 3×3 μm, the fractional variation is on the order of ±20%. This could be an important consideration limiting parametric yield for circuit designs which use small geometry devices.

As the device area is scaled down, there comes a point where a discrete limit is reached, and the noise characteristic ceases to be described by a $1/f$ characteristic. (This is the same discrete limit exploited in characterizing single-electron traps in Chapter 5). The discreteness becomes very apparent when the device has on average

only one trap active in any decade of frequency; coupling this with the trap densities extracted in Figure 6.10 allows the discrete limit to be evaluated for the devices used in this work.

The discrete limit for NMOS oxide devices at low bias can be estimated to occur for device sizes on the order of 1×1 μm. For NMOS ROXNOX devices, the discrete limit at low bias occurs for device sizes on the order of 0.5×0.5 μm. These limits, calculated from the results of the $1/f$ noise characteristics of large geometry devices, are fully consistent with the characterization results on single-electron traps in predicting the device sizes for which single-electron traps will be observable. For PMOS oxide devices at low bias, the discrete limit should occur for device sizes on the order of 0.25×0.25 μm. This small size requirement could explain why reports of random-telegraph-signals for PMOS devices have not as yet been reported in the literature.

6.4.2 Low Noise Circuit Design

Noise can be a significant factor in designing analog and mixed signal circuits [109]. Even though circuit techniques have been developed to reduce the effects of $1/f$ noise (such as correlated double-sampling [120]), it is clear that the ultimate low noise performance will be limited by the noise floor of the devices in the circuit. In many applications, $1/f$ noise cannot be avoided, so ways to minimize its effects are desirable.

The scaling rules discussed above are useful for designing a circuit for low noise. A key parameter which can nearly always be varied by the circuit designer is channel area; maximizing the channel area will yield the minimum noise. Lowering the operating temperature of the circuit in question may in some cases be a viable option for noise reduction. The noise of resistive elements in the circuit will also be lowered in this case. The only real advantage in devices dominated by $1/f$ noise gained by

lowering the temperature is found in PMOS devices, or in NMOS devices operated at high gate bias. For these devices, the noise value decreases as the square of the absolute temperature. For NMOS devices at low bias, the noise value is relatively insensitive to temperature variations.

Analyzing the noise performance as a trade-off of device real-estate for device operating temperature, the same noise performance for PMOS devices can be obtained either by reducing the temperature by a factor κ, by increasing the device area by a factor of κ^2, or by decreasing the oxide thickness under constant current scaling rules by a factor of κ. In most applications, it is expected that maximizing the device area will yield the greatest returns to scale. These dependencies for both NMOS and PMOS devices are reviewed in Table 6.1.

	PMOS		NMOS	
	low V_{gs}	high V_{gs}	low V_{gs}	high V_{gs}
T_{ox}	κ^2	$< \kappa^2$	κ^2	$< \kappa^2$
Area	κ^{-1}	κ^{-1}	κ^{-1}	κ^{-1}
Temperature	κ^2	$< \kappa^2$	$\sim$	$> \kappa$
V_{gs}	$\sim \kappa$	$\sim \kappa$	$\sim$	$\sim \kappa$

Table 6.1: Dependencies of $1/f$ noise on scaling given device parameter/operating-condition by factor of κ for devices biased at low gate bias and at high gate bias.

Another interesting result of the work in this thesis is the large variation in $1/f$ noise with gate bias observed for both NMOS and PMOS devices. This indicates that for the best noise performance, a low value of gate bias is warranted. Bias in the moderate inversion region is known to be desirable as far as obtaining the largest ratio of g_m to drain current [18]; the results of this thesis indicate that bias in this region will also yield the best noise performance.

It is important to evaluate the relative importance of $1/f$ noise compared to other

noise sources in the circuit before measures are taken to minimize the effects of $1/f$ noise. For small-signal applications, it is clear that for any frequency much less than 100 kHz that $1/f$ noise will be important. For mixed-signal systems, where both analog and digital signals exist in the same circuit, the dominant noise source is quite often quantization noise from an A/D or D/A conversion, or from a sampled-data operation [120]. As an example, for the NMOS oxide devices plotted in Figure 6.3 in the presence of quantization noise for a 16 bit converter with a 20 kHz bandwidth and a full scale range of 1 volt, the $1/f$ noise of the device is comparable to the quantization noise from the conversion at a frequency of 800 Hz. The $1/f$ noise corner frequency for this system is therefore said to be at $\sim$800 Hz; at frequencies below this point $1/f$ noise is dominant, and at frequencies higher than this quantization noise will be dominant. For a 20 bit conversion, the quantization noise PSD is comparable to the $1/f$ noise PSD at a frequency of $\sim$200 kHz, indicating that the quantization noise from the conversion will probably not be the dominant noise source in any range of frequencies. For a 12 bit conversion, the $1/f$ noise corner frequency occurs at only 3 Hz, indicating that $1/f$ noise is not a dominant concern over most of the band of interest. Note that this corner frequency is extracted by directly comparing noise sources at the input; the gain and location of the noise sources in the particular system in question must be taken into account for accurate estimation of the $1/f$ noise corner frequency. This comparison is for 20 μm/2 μm NMOS devices at low bias; PMOS devices will have a corner frequency about an order of magnitude lower than these values. Devices with a larger channel area will also have a lower $1/f$ noise corner frequency.

6.5 Radiation Improved Mobility

An interesting phenomenon occurs involving the mobility of a ROXNOX MOSFET when subjected to radiation stress. Devices which are irradiated and then annealed have been shown to exhibit an *increase* in mobility [23]. This phenomenon was investigated by characterizing the $1/f$ noise of the devices used in the irradiation and anneal study, in collaboration with Gregg Dunn at Lincoln Laboratories [23].

According to the noise model presented in this thesis, noise in a MOSFET is caused by fluctuation of the number of inversion layer carriers as they are trapped and de-trapped to and from traps located near the Si/SiO_2 interface. These fluctuations can also induce fluctuations in the channel mobility of the remaining carriers in the channel. In the linear region, the normalized $1/f$ noise PSD can be described by

$$\text{Normalized Noise PSD} \propto \frac{N_t(E_f)}{WLC_{ox}^2 f} \quad (\frac{V^2}{\text{Hz}}) \tag{6.43}$$

where W and L are the device width and length, respectively, C_{ox} is the device capacitance per unit area, and $N_t(E_f)$ is the oxide trap density at an energy in the oxide adjacent to the Fermi level. It is because the normalized drain current noise PSD is proportional to the density of near-interface oxide traps that $1/f$ noise measurements are a useful tool for characterizing the dielectric.

$1/f$ Noise was measured for both NMOS and PMOS transistors. The open circuit drain voltage noise was measured directly by an HP3585 spectrum analyzer, with no pre-amplification. The devices were measured in the linear region, to insure that the inversion carrier quasi-Fermi level did not vary greatly along the length of the device. The noise was averaged over a number of devices on each die to determine the relative variance in the noise measurement. The typical relative variance in the observed noise for a given die varied from 5% to 20%. The $1/f$ noise measurement results presented in [23] were presented in terms of normalized drain *voltage* noise.

Since the results are given in the linear region at low gate bias, the normalized drain voltage noise can be shown to be approximately equal to the normalized noise PSD of Equation (6.3),

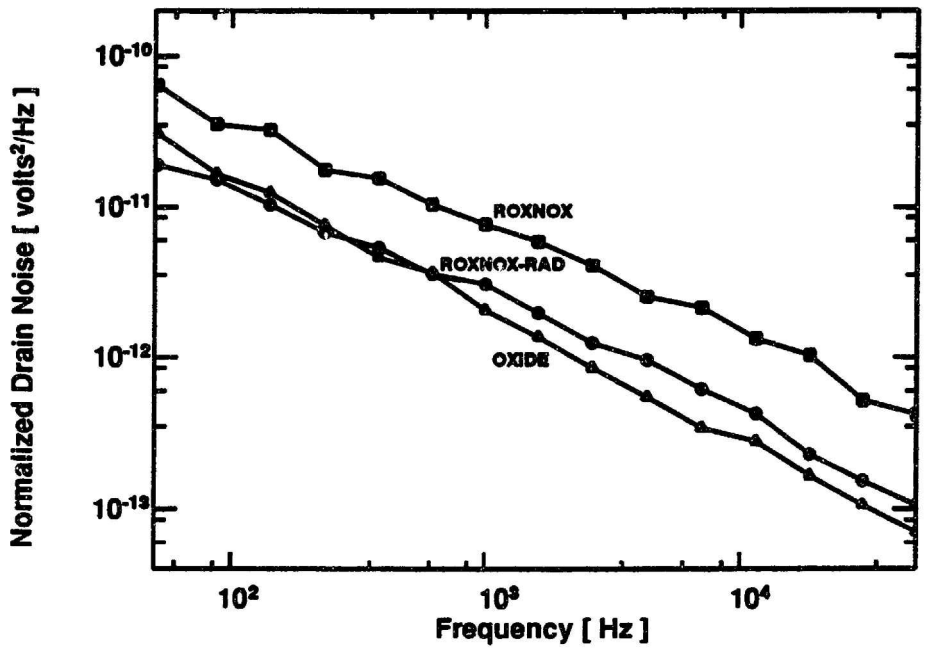
$$\frac{(V_{gs}-Vt)^2}{V_{ds}^2}S_{VD}(f)=\frac{g_{ds}^2}{g_{ds}^2}\frac{(V_{gs}-Vt)^2}{V_{ds}^2}S_{VD}(f)\simeq(V_{gs}-Vt)^2\frac{S_{I_{ds}}(f)}{I_{ds}^2} \qquad (6.44)$$

The results of the NMOS noise measurements are shown in Figure 6.12a. Note that the normalized drain noise of the ROXNOX dielectric is somewhat higher than in the control oxide. The higher value of noise PSD for the ROXNOX devices indicates the presence of a greater number of electron traps in the dielectric near the interface, and with energy levels near the conduction band of the silicon. Following irradiation, the $1/f$ noise level of the ROXNOX device (ROXNOX-RAD in Figure 6.12a), drops to a level comparable to that of the control oxide, indicating the reduction in the number of electron traps in the ROXNOX dielectric near the interface.

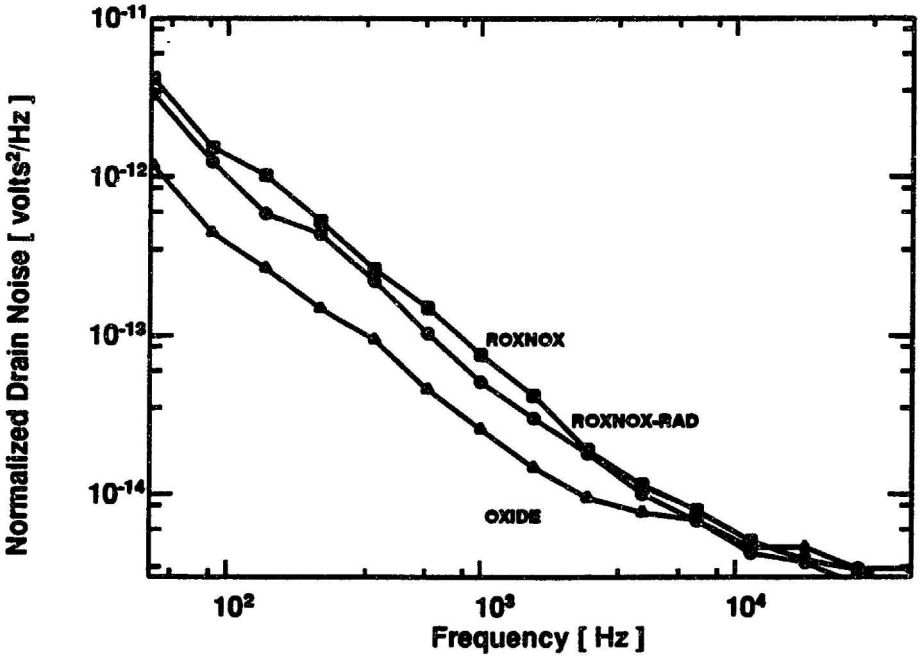
The $1/f$ noise results for the PMOS devices are shown in Figure 6.12b. Note that the noise value for the ROXNOX devices is higher than for the oxide devices. The $1/f$ noise level of the PMOS ROXNOX device is also reduced following irradiation, although to a much smaller extent than for the NMOS device. The flattening of the noise curves at high frequency is an artifact due to the noise at those frequencies being comparable to the noise of the measurement system.

Noise measurements were performed on individual devices which were then irradiated and the noise on those same devices re-measured to verify that the decrease in noise with irradiation was a real effect, and not only due to natural variations in trap-density or device geometry between different devices.

The reduction of $1/f$ noise in both PMOS and NMOS ROXNOX dielectrics following irradiation supports the hypothesis of an amphoteric trap which is present in ROXNOX dielectrics and which is removed by irradiation [23]. Because of the



(a) NMOS Devices.



(b) PMOS Devices.

Figure 6.12: Normalized drain noise of ROXNOX devices before (ROXNOX) and after (ROXNOX-RAD) irradiation. Control oxide noise value shown for comparison. $|V_{ds}| = 200$ mV, $|V_{gs} - V_{th}| = 1.0$ V. $T_{ox} = 37$ nm. $W_{eff} = 23.3$ μm. $L_{eff} = 1.3$-1.4 μm. The flattening of the PMOS curves at high frequency is an artifact due to the device noise becoming comparable to the measurement system noise.

difference in the reduction of $1/f$ noise between PMOS and NMOS devices with irradiation, it is required that either a) this amphoteric trap has a greater drain current fluctuation amplitude when trapping electrons than when trapping holes, or b) there are in general more traps in PMOS devices than NMOS devices, so that removing the amphoteric traps will have less of an effect on the noise behavior of PMOS devices. In any case, the presence of additional hole traps in the ROXNOX dielectric which are unaffected by irradiation is necessary to completely explain the $1/f$ noise results. Evidence has been presented in this thesis to support both of the above hypotheses.

6.6 Summary

Results of characterization of N- and P-MOSFET $1/f$ noise behavior has been reported. At low bias, the NMOS devices exhibit a relatively weak dependence of normalized drain noise on temperature, while the PMOS devices exhibit a temperature-squared dependence of noise PSD. The device noise varies as the inverse of channel area, and as the inverse of C_{ox}^2. Both N- and P-MOS devices exhibit an increase in noise PSD with higher gate bias.

These results are well modeled using a model of drain current fluctuations which are due to both the number of carriers in the channel fluctuating, and these fluctuations inducing fluctuations in the mobility of the remaining carriers in the channel. The near-interface traps responsible for these fluctuations are modeled based on the results of the single-electron-trap characterization reported in Chapter 5. The near-interface traps possess capture cross sections which are temperature activated, i.e.

$$\sigma = \sigma_o e^{-\frac{E_a}{kT}}. \tag{6.45}$$

An ensemble of these near-interface traps, uniformly distributed in activation energy E_a, yields the $1/f$ frequency dependence of the noise power spectral density. NMOS

devices in general are strongly affected by the induced mobility fluctuations, while PMOS devices are largely unaffected by induced mobility fluctuations. The variation of MOSFET noise PSD with gate bias is accounted for by the variation of trap density in the dielectric bandgap. This model not only explains the characterization data from this work, but also explains the characterization data from Christensson et al. [4].

A study of an effect of radiation on channel mobility suggests the possibility that amphoteric traps in the dielectric are responsible for a portion of both the NMOS and the PMOS device $1/f$ noise behavior. The trap density for the PMOS devices has been shown to be higher than the density for NMOS devices. Also, the magnitude of an individual trap fluctuation for NMOS devices is much greater than the assumed ideal single trap fluctuations responsible for $1/f$ noise in PMOS devices. These two facts help explain how the annealing of a single type of amphoteric trap in the MOSFET channel can cause a large improvement in NMOS noise power, but only a small improvement in PMOS noise power.

Chapter 7

Conclusions

7.1 Summary

This thesis investigated the $1/f$ noise performance of MOSFET devices with thin gate dielectrics. This investigation was performed to contribute $1/f$ noise data on modern MOSFET devices over a range of bias and temperature. The main thrust of this thesis was in testing the limits and validity of the assumptions commonly used in studying $1/f$ noise.

The $1/f$ noise study was performed on both conventional oxide and reoxidized nitrided oxide devices. To accomplish this, an 850°C ROXNOX process was developed based on the optimization of capacitor performance under Fowler-Nordheim stress. ROXNOX capacitors were verified to have greatly reduced interface-state generation and similar electron trapping under Fowler-Nordheim stress compared to conventional oxide. ROXNOX NMOS transistors were verified to have reduced g_m degradation under channel-hot-electron stress compared to conventional oxide. The drawback of the optimized 850°C ROXNOX dielectric, and ROXNOX processes in general, is that they possesses higher initial fixed charge and higher initial interface-state density causing the ROXNOX dielectric to have slightly shifted threshold voltages and lower mobility compared to conventional oxide.

Single-electron traps were characterized for deep-submicron NMOS devices with conventional oxide and ROXNOX gate dielectrics operated in the linear region for a range of gate bias and temperature. The trap parameters characterized were the trap time constant, the trap occupancy, and the magnitude of the current fluctuations caused by the capture and emission from a single trap.

The trap time constant was found to have a strongly temperature activated behavior. The trap activation energies were generally hundreds of meV, with traps active at room temperature having activation energies of 0.5 eV and greater. The trap capture and emission processes are modeled well by a multiphonon emission process, which accounts for the large range of activation energies. The trap occupancy varied exponentially with gate bias, and is modeled well by a single energy level, with the trap located at some depth into the oxide. The traps for the oxide dielectric were found to lie within 10 Å of the interface, while the traps for the ROXNOX dielectric were found to lie within 30 Å of the interface. These results are all in good agreement with recent reports of RTS behavior in oxide devices [25, 27].

A quantum-mechanical inversion-layer simulation was developed to accurately model the variation of Fermi level and inversion carrier density with gate bias and temperature. Application of this simulation to the results of the single-electron trap characterization demonstrated that the capture cross section of the single-electron traps were strongly bias dependent. This is a result which has not been previously reported.

The magnitude of the current fluctuation caused by the capture and emission from a single-electron trap was found to be a strong function of bias. The amplitude of the fluctuation was found to be much larger than that expected due to the removal of a single carrier from the channel region, indicating the importance of induced mobility fluctuations on the trap amplitude behavior. The trap amplitude is modeled

well as a term due to the removal of a carrier from the channel plus a term due to the change in mobility caused by the addition of a charged scattering site in the dielectric. The scattering coefficient S is shown to be a strongly decreasing function of gate bias, as expected from the effect of screening. This is a result which improves the model presented in [26], where the bias and temperature dependence of the scattering parameter was not developed.

The variation of trap time constant is found to have no explicit dependence on the distance the trap is located away from the interface. While it is expected that the distance the trap is located away from the interface will have some effect on the trap time constant, the variation in trap time constant with trap activation energy and capture cross section is relatively more important. This is a previously unexpected result.

The $1/f$ noise behavior of NMOS and PMOS transistors was characterized for devices in the linear region over a range of gate bias and temperature. The behavior of PMOS noise power spectral density is found to be proportional to temperature squared. This dependence of PMOS noise on temperature has not been previously reported in the literature. The behavior of NMOS noise PSD is found to be relatively temperature independent at low values of gate bias, and to be an increasing function of temperature at high bias.

The results of the single-electron-trap characterization were applied to modeling the ensemble of traps responsible for $1/f$ noise in the MOSFET channel region. The behavior of hole traps was assumed to be qualitatively the same as the characterized behavior of the NMOS traps. Using these results, a model was developed which predicts a temperature-squared dependence of noise PSD for device noise dominated by number fluctuations. For devices dominated by traps exhibiting induced mobility fluctuations, the noise PSD is found to be inversely proportional to temperature. In

previous work, the $1/f$ frequency dependence of the noise PSD is assumed to arise from traps uniformly distributed in depth x in the oxide which have capture cross sections

$$\sigma \propto e^{-\alpha x}. \tag{7.1}$$

In this work, the $1/f$ frequency dependence of noise PSD is found to arise from traps uniformly distributed in activation energy E_a which have capture cross sections

$$\sigma \propto e^{-\frac{E_a}{kT}}. \tag{7.2}$$

This proportionality is based on the measured capture cross sections obtained from the results of single-electron-trap characterization. This model agrees well not only with the measured results reported in this thesis, but also with the measured results reported by Christensson et al. [4].

Trap densities were extracted for both NMOS and PMOS devices using the $1/f$ noise model. Even though the noise PSD for PMOS devices is in general lower than that observed in NMOS devices, the number of near interface traps in PMOS devices is calculated to be higher than that observed in NMOS devices. This counter-intuitive result underscores the importance of induced mobility fluctuations on the noise performance of NMOS devices.

The utility of $1/f$ noise measurements for characterizing the near-interface oxide region is demonstrated in a study of a novel effect of irradiation and anneal observed in ROXNOX dielectrics. The mobility of NMOS devices with the ROXNOX dielectric is found to *increase* following irradiation and anneal. This increase is linked to the decrease in near-interface oxide trap density through the use of $1/f$ noise characterization. Both NMOS and PMOS trap density is reduced by the irradiation and anneal treatment. The relative decrease in NMOS trap density is inferred to be larger than the decrease in the PMOS trap density, which could be due to the fact that the initial

trap density in the PMOS devices is higher than in the NMOS devices.

7.2 Future Work

A number of suggestions for further study are presented here.

The large activation energies for trap capture and emission times should allow further significant study of single-electron trap behavior over a range of temperatures to be performed without an expensive low-temperature measurement system. This is because simply varying the temperature between 0°C and 100°C will cause trap time-constant variations of many orders of magnitude. While statistical data on the relative density of single-electron traps would be useful in verifying the noise model, gathering such data over a large number of devices and over a range of gate biases is prohibitively time consuming.

The $1/f$ noise model can be extended to weak inversion by applying the results of Reimbold described in Chapter 5 [17]. In the weak to moderate inversion region, electrons captured in a trap will not come solely from the mobile inversion layer carriers, but will be shared among the depletion, interface-state, and gate charges. This will cause each trap site to have a relatively smaller effect on the total $1/f$ noise power of the device. It is expected, however, that for NMOS devices, the effect of a trapped charge on inducing mobility fluctuations in the channel will be the dominant effect in the weak inversion region. A recent paper by Kleinpenning pointed out an interesting property of the $1/f$ noise of weakly inverted devices; the fact that the variance of the number of carriers in the channel cannot be greater than the number of carriers in the channel [106]. This property is expressed by the equation

$$\text{Noise PSD} \propto \frac{\frac{N_t(E_f)\,Q_n}{N_t(E_f)+Q_n}}{Q_n^2}. \tag{7.3}$$

For a strongly inverted surface, the density of inversion layer carriers is much greater

than the density of traps and the $1/f$ noise PSD is proportional to $N_t(E_f)/Q_n^2$, as is found in this work. In weak inversion, the density of inversion layer carriers can be less than the density of trap sites, and the noise will be proportional to [106]

$$\text{Noise PSD} \propto \frac{1}{Q_n}. \tag{7.4}$$

This property could be verified by noting that the noise performance of oxide and ROXNOX devices should approach the same value in weak inversion.

The extension of the model to the saturation region of operation should proceed in a relatively straightforward manner. It is expected that the noise of the device will be dominated by the strongly inverted drift-region of the channel near the source-end of the device. It should be assumed that the carriers flowing in the pinched-off region are not subject to $1/f$ noise fluctuations; the density of carriers is very small in this region indicating that few carriers will be trapped. Furthermore, the high average velocity of carriers in this region will cause the coulombic scattering cross section of any trapped charges to be very small [86]. The strongly inverted region near the source has a carrier density which is non-uniform, unlike the case for the measurements performed here in the linear region. It is expected that this non-linearity can be accounted for by using a differential-element model of the strongly inverted drift-region of the channel.

The modeling of PMOS $1/f$ noise PSD is based on the assumption that the induced mobility fluctuations for single-hole traps is much less, and the near-interface hole-trap density is much greater, for the PMOS devices than for the corresponding electron traps for NMOS devices. The testing of this assumption would verify many aspects of this work. The estimate of hole-trap density from the previous chapter indicates that PMOS devices of size less than 0.25×0.25 μm are required in order to obtain single-hole-trap characteristics. The fabrication of PMOS devices of this size should be possible with current technology.

The study of the radiation and anneal of ROXNOX MOSFET transistors indicates the possibility that a amphoteric trap is responsible for a portion of both NMOS and PMOS noise performance. A device designed to characterize single-electron amphoteric traps could provide insight into this problem. Such a device is shown schematically in Figure 7.1. This device has both n^+ and p^+ source/drain regions in contact

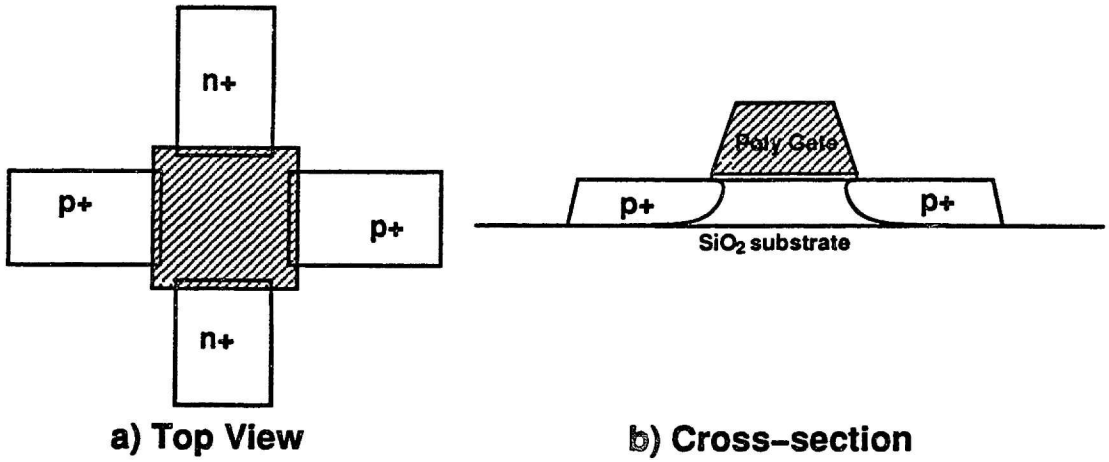


Figure 7.1: Schematic diagram of deep-submicron SOI device with dual source/drain regions. This device will allow single (amphoteric) traps to be measured in both inversion and accumulation. The device as shown uses mesa isolation.

with the channel. It is suggested that this device be formed in a silicon-on-insulator (SOI) technology to minimize the parasitic leakage path when the device is biased in accumulation. This device would allow not only the characterization of a single trap site in both accumulation and inversion, but also would allow the relative density of hole traps compared to electron traps in the MOSFET channel region to be ascertained.

The study of the effect of irradiation and anneal of ROXNOX dielectrics has demonstrated that the near-interface trap density can be altered. It may be worthwhile to characterize single electron traps present in ROXNOX devices, and then see if these traps can be removed by irradiation and anneal.

Bibliography

[1] W. Yang, R. Jayaraman, and C. G. Sodini, "Optimization of Low-Pressure Nitridation/Reoxidation of SiO_2 for Scaled MOS Devices," *IEEE Trans. Electron Devices*, vol. 35, pp. 935–944, 1988.

[2] G. J. Dunn, R. Jayaraman, W. Yang, and C. G. Sodini, "Radiation Effects in Low-Pressure Reoxidized Nitrided Oxide Gate Dielectrics," *Appl. Phys. Lett.*, vol. 52, no. 20, p. 1713, 1988.

[3] S. Christensson, I. Lundström, and C. Svensson, "Low Frequency Noise in MOS Transistors–I Theory," *Solid-State Electron.*, vol. 11, p. 797, 1968.

[4] S. Christensson and I. Lundström, "Low Frequency Noise in MOS Transistors–II Experiments," *Solid-State Electron.*, vol. 11, p. 813, 1968.

[5] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, pp. 40–42. Philadelphia, PA: Holt, Rinehart, and Winston, 1976.

[6] W. M. Siebert, *Circuits, Signals, and Systems*, ch. 19, p. 617. Cambridge, MA: M.I.T. Press, 1986.

[7] J. B. Johnson, "The Schottky Effect in Low Frequency Circuits," *Phys. Rev.*, vol. 26, p. 71, July 1925.

[8] A. van der Ziel, *Noise: Sources, Characterization, Measurement*, ch. 6, pp. 89–105. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1970.

[9] A. van der Ziel, "Unified Presentation of 1/f Noise in Electronic Devices: Fundamental 1/f Noise Sources," *IEEE Proc.*, vol. 76, no. 3, p. 233, Mar. 1988.

[10] A. L. McWhorter, "1/f Noise and Germanium Surface Properties," in *Semiconductor Surface Physics*, p. 207. Philadelphia: University of Pennsylvania Press, 1957.

[11] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*. New York: John Wiley and Sons, 1982.

[12] E. Takeda and N. Suzuki, "An Empirical Model for Device Degradation Due to Hot-Carrier Injection," *IEEE Electron Device Lett.*, vol. EDL-4, no. 4, p. 111, 1983.

[13] C. Hu, S. C. Tam, F.-C. Hsu, P.-K. Ko, T.-Y. Chan, and K. W. Terrill, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement," *IEEE Trans. Electron Devices*, vol. ED-32, no. 2, p. 375, 1985.

[14] J. E. Chung, M. C. Jeng, J. E. Moon, P. K. Ko, and C. Hu, "Performance and Reliability Design Issues for Deep Submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 38, no. 3, p. 545, 1991.

[15] M. A. Schmidt, F. L. Terry, B. P. Mathur, and S. D. Senturia, "Inversion Layer Mobility of MOSFET's with Nitrided Oxide Gate Dielectrics," *IEEE Trans. Electron Devices*, vol. 35, p. 1627, 1988.

[16] H. S. Momose, T. Morimoto, K. Yamabe, and H. Iwai, "Relationship Between Mobility and Residual-Mechanical-Stress as Measured by Raman Spectroscopy for Nitrided-Oxide-Gate MOSFETS," in *IEDM Tech. Dig.*, p. 65, 1990.

[17] G. Reimbold, "Modified 1/f Trapping Noise Theory and Experiments in MOS Transistors Biased from Weak to Strong Inversion–Influence of Interface States," *IEEE Trans. Electron Devices*, vol. ED-31, no. 9, p. 1190, Sept. 1984.

[18] Y. Tsividis, "Moderate Inversion in MOS Devices," *Solid-State Electron.*, vol. 25, no. 11, p. 1099, Nov. 1982.

[19] P. J. McWhorter and P. S. Winokur, "Simple Technique for Seperating the Effects of Interface Traps and Trapped-Oxide Charge in Metal-Oxide-Semiconductor Transistors," *Appl. Phys. Lett.*, vol. 48, no. 2, p. 133, Jan. 1986.

[20] E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, pp. 331–333. New York: John Wiley and Sons, 1982.

[21] G. Groeseneken, H. E. Maes, N. Beltrán, and R. F. D. Keersmaecker, "A Reliable Approach to Charge-Pumping Measurements in MOS Transistors," *IEEE Trans. Electron Devices*, vol. ED-31, no. 1, p. 42, 1984.

[22] R. Jayaraman, *Reliability and 1/f Noise Properties of MOSFETs With Nitrided Oxide Gate Dielectrics.* PhD thesis, Mass. Inst. of Tech., Dept. of Electrical Engineering and Computer Science, 1988.

[23] G. J. Dunn, B. J. Gross, and C. G. Sodini, "Radiation-Induced Increase in the Inversion Layer Mobility of Reoxidized Nitrided Oxide MOSFET's," *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 677–684, Mar. 1992.

[24] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, "Discrete Resistance Switching in Submicron Silicon Inversion Layers: Individual Interface Traps and Low Frequency ($1/f$?) Noise," *Phys. Rev. Lett.*, vol. 52, no. 3, p. 228, 1984.

[25] M. Kirton and M. Uren, "Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States, and Low-Frequency (1/f) Noise," *Adv. Phys.*, vol. 38, no. 4, p. 368, 1989.

[26] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "Random Telegraph Noise of Deep-Submicrometer MOSFETs," *IEEE Electron Device Lett.*, vol. 11, no. 2, p. 90, Apr. 1990.

[27] Z. Shi, J.-P. Miéville, J. Barrier, and M. Dutoit, "Low Frequency Noise and Quantum Transport in 0.1 μm N-MOSFET's," *IEDM Tech. Dig.*, p. 363, 1991.

[28] W. Yang, "Low Pressure Nitrided Oxide in MOS Capacitors," Master's thesis, Mass. Inst. of Tech., Dept. of Electrical Engineering and Computer Science, 1987.

[29] J. Bourgoin and M. Lanoo, *Point Defects in Semiconductors II.* Berlin: Springer-Verlag, 1983.

[30] P. K. Tedrow and C. G. Sodini, "MIT Twin-Well CMOS Process," VLSI Memo Version 1.2, Massachusetts Institute of Technology, 1988.

[31] J. Chung, M.-C. Jeng, J. Moon, A. Wu, T. Chan, P. Ko, and C. Hu, "Deep-Submicrometer MOS Device Fabrication Using a Photoresist-Ashing Technique," *IEEE Electron Device Lett.*, vol. 9, no. 4, p. 186, Apr. 1988.

[32] *HP TECAP Software User's Manual.* Santa Clara, CA: Hewlett-Packard Co., 1985.

[33] M. L. Naiman, C. T. Kirk, B. L. Emerson, J. B. Taitel, and S. D. Senturia, "The Constitution of Nitrided Oxides and Reoxidized Nitrided Oxides on Silicon," *J. Appl. Phys.*, vol. 58, no. 2, p. 779, 1985.

[34] T. Hickmott and R. Isaac, "Barrier Heights at the Polycrystalline Silicon-SiO_2 Interface," *J. Appl. Phys.*, vol. 52, no. 5, p. 3464, 1981.

[35] E. Kooi, J. van Lierrop, and J. Appels, "Formation of Silicon Nitride at the Si-SiO_2 Interface During Local Oxidation of Silicon During Heat-Treatment of Oxidized Silicon in NH_3 Gas," *J. Electrochem. Soc.*, vol. 123, p. 1729, 1976.

[36] S. C. Sun and J. D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces," *IEEE Trans. Electron Devices*, vol. ED-27, pp. 1497–1508, 1980.

[37] J. G. J. Chern, P. Chang, R. F. Motta, and N. Godinho, "A New Method to Determine MOSFET Cahnnel Length," *IEEE Electron Device Lett.*, vol. EDL-1, no. 9, p. 170, Sept. 1980.

[38] F. Reif, *Fundamentals of Statistical and Thermal Physics*, pp. 178–184. New York, NY: McGraw-Hill, 1965.

[39] J. Roberge, *Operational Amplifiers: Theory and Practice*, pp. 171–196. New York, NY: John Wiley and Sons, 1975.

[40] A. van der Ziel and H. S. Park and S. T. Liu, "A Discrepancy in the Elementary Theory of MOSFET Modeling," *Appl. Phys. Lett.*, vol. 35, no. 12, p. 942, Dec. 1979.

[41] C. Sodini, P.-K. Ko, and J. Moll, "The Effect of High Fields on MOS Device and Circuit Performance," *IEEE Trans. Electron Devices*, vol. ED-31, no. 10, p. 1386, Oct. 1984.

[42] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York, NY: McGraw-Hill, Inc., 1984.

[43] T. Ito, T. Nozaki, and H. Ishikawa, "Direct Thermal Nitridation of Silicon Dioxide Films in Anhydrous Ammonia Gas," *J. Electrochem. Soc.*, vol. 127, no. 9, p. 2053, 1980.

[44] M. L. Naiman, F. L. Terry, J. A. Burns, J. I. Raffel, and R. Aucoin, "Properties of Thin Oxynitride Gate Dielectrics Produced by Thermal Nitridation of Silicon Dioxide," in *IEDM Tech. Dig.*, pp. 562–565, 1980.

[45] S. K. Lai, J. Lee, and V. Dham, "Electrical Properties of Nitrided-Oxide Systems for Use in Gate Dielectrics and EEPROM," in *IEDM Tech. Dig.*, pp. 190–193, 1983.

[46] F. L. Terry, R. J. Aucoin, M. L. Naiman, and S. D. Senturia, "Radiation Effects in Nitrided Oxides," *IEEE Electron Device Lett.*, vol. EDL-4, pp. 191–193, 1983.

[47] T. Ito, T. Nakamura, and H. Ishikawa, "Advantages of Thermal Nitride and Nitroxide Gate Films in VLSI Process," *IEEE Trans. Electron Devices*, vol. ED-29, no. 4, p. 498, 1982.

[48] C.-T. Chen, F.-C. Tseng, C.-Y. Cheng, and M.-K. Lee, "Study of Electrical Characteristics on Thermally Nitrided SiO_2 (Nitroxide) Films," *J. Electrochem. Soc.*, p. 875, 1984.

[49] S.-T. Chang, N. M. Johnson, and S. A. Lyon, "Capture and Tunnel Emission of Electrons by Deep Levels in Ultrathin Nitrided Oxides on Silicon," *Appl. Phys. Lett.*, vol. 44, no. 3, p. 316, 1984.

[50] T. T. L. Chang, H. S. Jones, C. S. Jenq, W. S. Johnson, J. Lee, S. K. Lai, and V. K. Dham, "Oxidized-Nitridized Oxide (ONO) for High Performance EEPROMS," in *IEDM Tech. Dig.*, p. 810, 1982.

[51] M. M. Moslehi and K. C. Saraswat, "Thermal Nitridation of Si and SiO_2 for VLSI," *IEEE J. Solid-State Circuits*, vol. 20, no. 1, p. 26, 1985.

[52] A. T. Wu, V. Murali, J. Nulman, B. Triplett, D. B. Fraser, and M. Garner, "Gate Bias Polarity Dependence of Charge Trapping and Time-Dependent Dielectric Breakdown in Nitrided and Reoxidized Nitrided Oxides," *IEEE Electron Device Lett.*, vol. 10, no. 10, p. 443, 1989.

[53] H. S. Momose, S. Kitagawa, K. Yamabe, and H. Iwai, "Hot Carrier Related Phenomena for N- and P-MOSFETs With Nitrided Gate Oxide by RTP," in *IEDM Tech. Dig.*, pp. 267–270, 1989.

[54] T. Hori and H. Iwasaki, "The Impact of Ultrathin Nitrided Oxide Gate-Dielectrics on MOS Device Performance Improvement," in *IEDM Tech. Dig.*, pp. 459–462, 1989.

[55] C. Wong, Y.-C. Sun, Y. Taur, C. Oh, R. Angelucci, and B. Davari, "Doping of n^+ and p^+ Polysilicon in a Dual-Gate Process," in *IEDM Tech. Dig.*, p. 238, 1988.

[56] C.-Y. Lu, J. Sung, H. Kirsch, S. Hillenius, T. Smith, and L. Manchanda, "Anomalous C-V Characteristics of Implanted Poly MOS Structures in n^+/p^+ Dual-Gate CMOS Technology," *IEEE Electron Device Lett.*, vol. 10, no. 5, p. 192, 1989.

[57] J. Pfiester, F. Baker, T. Mele, H.-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. Electron Devices*, vol. 37, no. 8, p. 1842, 1990.

[58] T. Morimoto, H. Momose, Y. Ozawa, K. Yamabe, and H. Iwai, "Effects of Boron Penetration and Resultant Limitations in Ultra Thin Pure-Oxide and Nitrided-Oxide Gate-Films," in *IEDM Tech. Dig.*, p. 429, 1990.

[59] G. Lo and D. Kwong, "The Use of Ultrathin Reoxidized Nitrided Gate Oxide for Suppression of Boron Penetration in BF_2^+-Implanted Polysilicon Gated p-MOSFETs," *IEEE Electron Device Lett.*, vol. 12, no. 4, p. 175, 1991.

[60] H. Fang, K. S. Krisch, B. J. Gross, C. G. Sodini, J. E. Chung, and D. A. Antoniadis, "Low-Temperature Furnace-Grown Reoxidized Nitrided Oxide Gate Dielectrics as a Barrier to Boron Penetration," *IEEE Electron Device Lett.*, vol. 13, no. 4, p. 217, Apr. 1992.

[61] Z. H. Liu, P. Nee, P. K. Ko, C. Hu, C. G. Sodini, B. J. Gross, T.-P. Ma, and Y. C. Cheng, "Field and Temperature Acceleration of Time-Dependent Dielectric Breakdown for Reoxidized-Nitrided and Fluorinated Oxides," *IEEE Electron Device Lett.*, vol. 13, no. 1, p. 41, 1992.

[62] Z. H. Liu, E. Rosenbaum, P. K. Ko, C. Hu, , Y. C. Cheng, C. G. Sodini, B. J. Gross, and T.-P. Ma, "A Comparative Study of the Effects of Dynamic Stressing on High-Field Endurance and Stability of Reoxidized-Nitrided, Flourinated and Conventional Oxides," in *IEDM Tech. Dig.*, pp. 723–726, 1991.

[63] C. Sodini, T. Ekstedt, and J. Moll, "Charge Accumulation and Mobility in Thin Dielectric MOS Transistors," *Solid-State Electron.*, vol. 25, no. 9, p. 833, 1982.

[64] A. T. Wu, T. Y. Chan, V. Murali, S. W. Lee, J. Nulmand, and M. Garner, "Nitridation Induced Surface Donor Layer in Silicon and It's Impact on the Characteristics of n- and p-Channel MOSFETs," in *IEDM Tech. Dig.*, p. 271, 1989.

[65] G. Lo, W. Ting, and D. Kwong, "Study of Inversion Layer Mobility in Metal-Oxide-Semiconductor Field-Effect Transistors with Reoxidized Nitrided Oxides," *Appl. Phys. Lett.*, vol. 56, no. 25, p. 2548, 1990.

[66] 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," *Surf. Sci.*, vol. 32, p. 561, 1972.

[67] R. Jayaraman and C. G. Sodini, "A 1/f Noise Technique to Extract the Oxide Trap Density Near the Conduction Band Edge of Silicon," *IEEE Trans. Electron Devices*, vol. 36, pp. 1773–1782, 1989.

[68] T. Y. Chan, P. K. Ko, and C. Hu, "Simple Method to Characterize Substrate Current in MOSFET's," *IEEE Electron Device Lett.*, vol. EDL-5, pp. 505–507, 1984.

[69] P. Fang, K. Hung, P. Ko, and C. Hu, "Hot-Electron-Induced Traps Studied Through the Random Telegraph Noise," *IEEE Electron Device Lett.*, vol. 12, no. 6, p. 273, June 1991.

[70] E. Simoen, B. Dierickx, C. L. Claeys, and G. J. Declerck, "Explaining the Amplitude of RTS Noise in Submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 39, no. 2, p. 422, Feb. 1992.

[71] B. J. Gross, K. S. Krisch, and C. G. Sodini, "An Optimized 850°C Low Pressure Furnace Reoxidized Nitrided Oxide Process," *IEEE Trans. Electron Devices*, vol. 38, no. 8, Aug. 1991.

[72] A. W. Drake, *Fundamentals of Applied Probability Theory.* New York: McGraw-Hill, 1967.

[73] C. T. Sah, "Theory of Low-Frequency Generation Noise in Junction-Gate Field-Effect Transistors," *IEEE Proc.*, vol. PR-52, no. 7, p. 795, July 1964.

[74] B. Mandelbrot, "Some Noise with 1/f Spectrum, a Bridge Between Direct Current and White Noise," *IEEE Trans. Inform. Theory*, vol. IT-13, no. 2, p. 289, Apr. 1967.

[75] F. Stern and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit," *Phys. Rev.*, vol. 163, no. 3, p. 816, Nov. 1967.

[76] J. R. Brews, "Theory of the Carrier-Density Fluctuations in an IGFET Near Threshold," *J. Appl. Phys.*, vol. 46, no. 5, p. 2181, May 1975.

[77] S. Mori and T. Ando, "Intersubband Scattering Effect on the Mobility of a Si (100) Inversion Layer at Low Temperatures," *Phys. Rev. B*, vol. 19, no. 12, p. 6433, June 1979.

[78] B. Vinter, "Impurity Scattering in Inversion Layers in Density Functional Formalisms," *Surf. Sci.*, vol. 98, p. 197, 1980.

[79] T. Ando, A. B. Fowler, and F. Stern, "Electronic Properties of Two-Dimensional Systems," *Rev. Mod. Phys.*, vol. 54, no. 2, p. 437, 1982.

[80] F. Stern, "Self-Consistent Results for n-Type Si Inversion Layers," *Phys. Rev. B*, vol. 5, no. 12, p. 4891, June 1972.

[81] T. H. Ning and C. T. Sah, "Effects of Inhomogeneities of Surface-Oxide Charges on the Electron Energy Levels in a Semiconductor Surface-Inversion Layer," *Phys. Rev. B*, vol. 9, no. 2, p. 527, Jan. 1974.

[82] B. Vinter, "Self-Consistent Calculation of Impurity Scattering in Inversion Layers," *Solid-State Com.*, vol. 28, no. 10, p. 861, 1978.

[83] S. Padmanabhan and A. Rothwarf, "Quantum Inversion Layer Mobility: Numerical Results," *IEEE Trans. Electron Devices*, vol. 36, no. 11, p. 2557, Nov. 1989.

[84] Z. A. Weinberg, "Tunneling of Electrons from Si into Thermally Grown SiO_2," *Solid-State Electron.*, vol. 20, p. 11, 1977.

[85] M. Lanoo and J. Bourgoin, *Point Defects in Semiconductors I.* Berlin: Springer-Verlag, 1981.

[86] A. van der Ziel, *Solid State Physical Electronics, 3rd Ed.*, pp. 55–57. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1976.

[87] R. L. Liboff, *Introductory Quantum Mechanics*, p. 232. Oakland, CA: Holden-Day, 1980.

[88] O. Engström and M. S. Shivaraman, "Statistics for the Interpretation of Deep Level Transient Spectroscopy on Insulator-Semiconductor Interfaces," *J. Appl. Phys.*, vol. 58, no. 10, p. 3929, Nov. 1985.

[89] R. Jayaraman and C. G. Sodini, "1/f Noise Interpretation of the Effect of Gate Oxide Nitridation and Reoxidation on Dielectric Traps," *IEEE Trans. Electron Devices*, vol. 37, no. 1, p. 305, Jan. 1990.

[90] S. T. Hsu, "Surface State Related 1/f Noise in MOS Transistors," *Solid-State Electron.*, vol. 13, p. 1451, 1970.

[91] H. E. Maes and S. H. Usmani, "1/f Noise in Thin Oxide p-Channel Metal-Nitride-Oxide-Silicon Transistors," *J. Appl. Phys.*, vol. 54, no. 4, p. 1937, Apr. 1983.

[92] Z. Çelik and T. Y. Hsiang, "Study of 1/f Noise in N-MOSFETs: Linear Region," *IEEE Trans. Electron Devices*, vol. ED-32, no. 12, p. 2797, Dec. 1985.

[93] C. Y. H. Tsai and J. Gong, "$1/f$ Noise in the Linear Region of LDD MOSFET's," *IEEE Trans. Electron Devices*, vol. 35, no. 12, p. 2373, Dec. 1988.

[94] P. W. Anderson, B. I. Halperin, and C. M. Varma, "Anomalous Low-temperature Thermal Properties of Glasses and Spin Glasses," *Phil. Mag.*, vol. 25, p. 1, 1972.

[95] C. T. Rogers and R. A. Buhrman, "Composition of $1/f$ Noise in Metal-Insulator-Metal Tunnel Junctions," *Phys. Rev. Lett.*, vol. 53, no. 13, p. 1272, Sept. 1984.

[96] C. T. Rogers and R. A. Buhrman, "Nature of Single-Localized-Electron States Derived from Tunneling Measurments," *Phys. Rev. Lett.*, vol. 55, no. 8, p. 859, Aug. 1985.

[97] F. N. Hooge and L. K. J. Vandamme, "Lattice Scattering Causes 1/f Noise," *Phys. Lett.*, vol. 66A, no. 4, p. 315, May 1978.

[98] P. H. Handel, "Quantum Approach to 1/f Noise," *Phys. Rev. B*, vol. 22, no. 2, p. 745, Aug. 1981.

[99] M. A. Caloyannides, "Microcycle Spectral Estimates of 1/f Noise in Semiconductors," *J. Appl. Phys.*, vol. 45, no. 1, p. 307, Jan. 1974.

[100] F. N. Hooge, "1/f Noise," *Physica*, vol. 83B, p. 14, 1976.

[101] C. T. Sah and F. H. Hielscher, "Evidence of the Surface Origin of the $1/f$ Noise," *Phys. Rev. Lett.*, vol. 17, no. 18, p. 956, Oct. 1966.

[102] G. Abowitz, E. Arnold, and E. A. Leventhal, "Surface States and 1/f Noise in MOS Transistors," *IEEE Trans. Electron Devices*, vol. ED-14, no. 11, p. 775, Nov. 1967.

[103] H.-S. Fu and C.-T. Sah, "Theory and Experiments on Surface 1/f Noise," *IEEE Trans. Electron Devices*, vol. ED-19, no. 2, p. 273, Feb. 1972.

[104] H. E. Maes, S. H. Usmani, and G. Groeseneken, "Correlation Between 1/f Noise and Interface State Density at the Fermi Level in Field Effect Transistors," *J. Appl. Phys.*, vol. 57, no. 10, p. 4811, May 1985.

[105] B. Boukriss, H. Haddara, S. Cristoloveanu, and A. Chovet, "Modeling of the 1/f Noise Overshoot in Short-Channel MOSFET's Locally Degraded by Hot-Carrier Injection," *IEEE Electron Device Lett.*, vol. 10, no. 10, p. 433, Oct. 1989.

[106] T. G. M. Kleinpenning, "On 1/f Trapping Noise in MOST's," *IEEE Trans. Electron Devices*, vol. 37, no. 9, p. 2084, Sept. 1990.

[107] M. B. Das and J. M. Moore, "Measurements and Interpretation of Low-Frequency Noise in FET's," *IEEE Trans. Electron Devices*, vol. ED-21, no. 4, p. 247, Apr. 1974.

[108] L. K. J. Vandamme and H. M. M. D. Werd, "1/f Noise Model for MOSTs Biased in Nonohmic Region," *Solid-State Electron.*, vol. 23, p. 325, 1980.

[109] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, pp. 664–667. New York, NY: John Wiley and Sons, 1984.

[110] H. Mikoshiba, "1/f Noise in Silicon-Gate MOS Transistors," *IEEE Trans. Electron Devices*, vol. ED-29, no. 6, p. 965, June 1982.

[111] F. Grabowski, "Influence of the Interface and of the Channel Volume on 1/f Noise of MOS Transistors Biased in the Linear Region at Strong Inversion," *Solid-State Electron.*, vol. 31, no. 1, p. 115, Jan. 1988.

[112] Z. Çelik Butler and T. Y. Hsiang, "Spectral Dependence of $1/f^{\gamma}$ Noise on Gate Bias in N-MOSFETs," *Solid-State Electron.*, vol. 30, no. 4, p. 419, Apr. 1987.

[113] C. Surya and T. Y. Hsiang, "Surface Mobility Fluctuations in Metal-Oxide-Semiconductor Field-Effect Transistors," *Phys. Rev. B*, vol. 35, no. 12, p. 6343, Apr. 1987.

[114] Z. Çelik Butler and T. Y. Hsiang, "Spatial Correlation Measurements of 1/f Noise in Semiconductors," *Solid-State Electron.*, vol. 31, no. 2, p. 241, Feb. 1988.

[115] Z. Çelik Butler and T. Y. Hsiang, "Determination of Si-SiO_2 Interface Trap Density by 1/f Noise Measurements," *IEEE Trans. Electron Devices*, vol. 35, no. 10, p. 1651, Oct. 1988.

[116] C. Surya and T. Y. Hsiang, "A Thermal Activation Model for $1/f^{\gamma}$ Noise in Si-MOSFETs," *Solid-State Electron.*, vol. 31, no. 5, p. 959, May 1988.

[117] A. van der Ziel, "Dependence of Flicker Noise in MOSFETs on Geometry," *Solid-State Electron.*, vol. 20, p. 267, 1977.

[118] H. Wong and Y. Cheng, "Study of the Electronic Trap Distribution at the SiO_2-Si Interface Utilizing the Low-Frequency Noise Measurement," *IEEE Trans. Electron Devices*, vol. 37, no. 7, p. 1743, 1990.

[119] W. M. Siebert, *Circuits, Signals, and Systems*, ch. 11, p. 314. Cambridge, MA: M.I.T. Press, 1986.

[120] R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, pp. 502–513. New York: John Wiley & Sons, 1986.

[121] L. Shen and J. Kong, *Applied Electromagnetism*, pp. 326–341. Monterey, CA: Brooks/Cole Engineering Division, 1983.

[122] G. Strang, *Introduction to Applied Mathematics*, pp. 412–414. Wellesley, MA: Wellesley-Cambridge Press, 1986.

[123] F. Reif, *Fundamentals of Statistical and Thermal Physics*, pp. 202–206. New York, NY: McGraw-Hill, 1965.

[124] R. Eisburg and R. Resnick, *Quantum Physics, 2nd Ed.*, pp. K1 – K5. New York, NY: John Wiley and Sons, 1985.

Appendix A

Capacitor Process Flow

Step #	Type	Description		
1.	Diffusion	Grow Stress Relief Oxide (40 nm)		
		Temp °C	*Time min*	*Gas*
	dsro400.set	950	100	Dry O_2
		950	30	N_2
2.	Diffusion	Deposit LPCVD Nitride (150 nm)		
	dnit1.5.set	Deposition Temperature = 800°C		
3.	Lithography(ND)	Active Area Definition		
	phfieldpkt.set			
4.	Etch	Dry Etch Nitride		
	plnit1.5k.set			
5.	Implant (4)	Field Implant		
	infieldpkt.set	*Implant Element*	*Energy (keV)*	*Dose*
		Phosphorus	40	3E12
6.	Ashing	Plasma Strip Resist		
	ash.set			

7. Diffusion
dfield5.1k.set

Field Oxidation (500 nm)

Temp °C	*Time min*	*Gas*
950	30	Dry O_2
950	175	Wet O_2
950	30	Dry O_2
950	30	N_2

8. Wet Etch
wnit1.5k.set

Wet Nitride Strip
Surface nitride removal - 7:1 buffered HF
Strip nitride in Transetch - 175°C

9. Wet Etch
wsro320.set

Stress Relief Oxide Strip
7:1 buffered HF

10. Diffusion
dgate220.set

Sacrificial Gate Oxide[1] (22 nm)

Temp °C	*Time min*	*Gas*
950	35	Dry O_2
950	30	N_2

11. Wet Etch
wgate200.set

Dummy Gate Oxide Strip
50:1 HF during RCA clean

12. Diffusion
dgate100.set

Gate Oxide[2]
Various Oxide and ROXNOX dielectrics
850°C - 950°C

13. Diffusion
dpoly5k.set

Deposit LPCVD Polysilicon (500 nm)
This step is done immediately after step #12
Deposition Temperature = 625°C

14. Lithography(NP)
phcoat.set

Remove Back Side Polysilicon

15. Etch
plpolyback5k.set

Dry Etch Back Side Polysilicon

16. Wet etch
wox5k.set

Strip Back Side Gate Oxide
7:1 buffered HF

17. Ashing
ash.set

Plasma Strip Resist

18.	Diffusion dphos8.set	Dope Polysilicon[3] ($POCl_3$)		
		Temp °C	*Time min*	*Gas*
		925	60	DryO_2 + N_2
		925	15	O_2
		925	10	N_2
19.	Wet Etch wphos1.2.set	Strip pglass[4] 7:1 buffered HF		
20.	Lithography (NP) phpolypkt.set	Pattern Polysilicon		
21.	Etch plpoly5k.set	Dry Etch Polysilicon		
22.	Ashing ash.set	Plasma Strip Resist		
23.	Diffusion dsinter.set	Final Sinter[5]		
		Temp °C	*Time min*	*Gas*
		400	15	Forming Gas

Footnotes:

1. The sacrificial gate oxide is neccesary to avoid the Kooi effect [35], which can cause thinning of the oxide at the edges of the active area.
2. Gate oxidation performed in low-pressure oxidation tube in TRL in various O_2, NH_3 and N_2 ambients, as described in Section 2.1.
3. Because the back side polysilicon and oxide are stripped before this step, both the front side polysilicon and back side substrate are simultaneously heavily doped.
4. Exposure of bare polysilicon to HF etch is kept to a minimum; overexposure seems to cause significant loss of yield.
5. The total anneal time at 400°C is 30 min, but the H_2 gas is only on for 15 min.
6. No metalization is performed for the capacitor process; contact is made directly to the heavily doped substrate with a probe, and to the back side by the probe-station chuck.

Appendix B

CMOS Process Flow

Step #	Type	Description		
1.	Diffusion	Grow Stress Relief Oxide (40 nm)		
		Temp °C	*Time min*	*Gas*
	dsro430.set	950	100	Dry O_2
		950	30	N_2
2.	Diffusion	LPCVD Nitride Deposition (150 nm)		
	dnit1.5.set	Deposition Temperature = 800°C		
3.	Lithography (1)	Well Definition		
	phnwelldec.set			
4.	Etch	Plasma Etch Nitride		
	plnit1.5.set			
5.	Implant (1)	n-Well Formation		
	inwellpkt.set	*Implant Element*	*Energy (keV)*	*Dose*
		Phosphorus	180	2E12
6.	Ashing	Plasma Strip Resist		
	ash.set			

7. Diffusion
dnwell5.1k.set

n-Well Cover Oxidation (500 nm)

Temp °C	*Time min*	*Gas*
950	30	Dry O_2
950	175	Wet O_2
950	30	Dry O_2
950	30	N_2

8. Wet Etch
wnit1.5k.set

Wet Nitride Strip
Surface nitride removal - 7:1 buffered HF
Strip nitride in Transetch -175°C

9. Implant (2)
ipwellpkt.set

p-Well Formation

Implant Element	*Energy (keV)*	*Dose*
Boron	30	1.5E12

10. Diffusion
dwell7k.set

Well Drive

Temp °C	*Time min*	*Gas*
1150	900	Dry O_2
1150	30	N_2

11. Wet Etch
wwell9k.set

Strip Surface Oxide
7:1 buffered HF

12. Diffusion
dsro430.set

Grow Stress Relief Oxide
Same as step #1

13. Diffusion
dnit1.5k.set

LPCVD Nitride Deposition
Same as step #2

14. Lithography (2)
phactivedec.set

Active Area Definition
Defined both on n-Well and p-Well

15. Etch
plnit1.5k.set

Dry Etching of Nitride

16. Lithography (3)
phpfielddec.set

Cover n-Well
Important! **DOUBLE RESIST STEP**
Resist from the previous step is left on
the wafers before spinning on the new resist

17. Implant (3) — ipfieldpkt.set

p-Well Field Threshold Adj. Implant

Implant Element	*Energy (keV)*	*Dose*
Boron	70	1E13

18. Ashing — ash.set

Plasma Ash Resist

19. Lithography (4) — phnfielddec.set

Cover p-Well

20. Implant (4) — infieldpkt.set

n-Well Field Threshold Adj. Implant

Implant Element	*Energy (keV)*	*Dose*
Phosphorus	40	3E12

21. Ashing — ash.set

Plasma Strip Resist

22. Diffusion — dfield5.1k.set

Field Oxidation

Same as step #7

23. Wet Etch — wnit1.5k.set

Nitride Strip

Same as step #8

24. Wet Etch — wsro430.set

Stress Relief Oxide Strip

7:1 buffered HF

25. Diffusion — dgate220.set

Grow Sacrificial Gate Oxide (22 nm)

Temp °C	*Time min*	*Gas*
950	35	Dry O_2
950	30	N_2

26. Lithography (5) — phnvtdec.set

n-Channel Punchthrough and Threshold Adjustment Mask

cover n-Well

27. Implant (5)

n-Channel Punchthrough and Threshold Adjustment Implants

	Implant Element	*Energy (keV)*	*Dose*
invt1pkt.set	Boron	100	6E11
invt2pkt.set	Boron	30	1.5E12

28. Ashing
ash.set

Plasma Strip Resist

29. Lithography (6)
phpvtdec.set

p-Channel Punchthrough and Threshold Adjustment Mask
cover p-Well

30. Implant (6)
ipvt1pkt.set

p-Channel Punchthrough Implant[1]

Implant Element	*Energy (keV)*	*Dose*
Phosphorus	180	6E11

31. Ashing
ash.set

Plasma Strip Resist

32. Wet etch
wgate220.set

Strip Sacrificial Gate Oxide
Dip in 50:1 HF During RCA clean
(after organic clean)
for about 3.5 min. Check sheeting.

33. Diffusion
dgate100.set

Gate Oxide[2]
Various Oxide and ROXNOX dielectrics
850°C - 950°C

34. Diffusion
dpoly5k.set

Deposit LPCVD Polysilicon (300 nm)
This step is performed immediately after step #33

35. Diffusion
dphos8.set

Dope Polysilicon ($POCl_3$)

Temp °C	*Time min*	*Gas*
925	60	DryO_2 + N_2
925	15	O_2
925	10	N_2

36. Wet Etch
wphos1.2.set

Strip pglass[3]
7:1 buffered HF

37. Lithography (7)
phpolydec.set

Pattern Polysilicon

38. Etch
plpoly5k.set

Dry Etch Polysilicon

39. Ashing
ash.set

Plasma Strip Resist

40. Lithography (8)
php+dec.set

Define PMOS S/D and p+ Regions

41. Implant (7)
ip+pkt.set

Implant PMOS S/D and p+ Regions

Implant Element	*Energy (keV)*	*Dose*
BF2	30	7E15

42. Ashing
ash.set

Plasma Strip Resist

43. Lithography (9)
phn+dec.set

Define NMOS S/D and n+ Regions

44. Implant (8)
in+pkt.set

Implant NMOS S/D and n+ Regions

Implant Element	*Energy (keV)*	*Dose*
Arsenic	90	7E15

45. Ashing
ash.set

Plasma Strip Resist

46. Diffusion
dreox.set

Reoxidation

Temp °C	*Time min*	*Gas*
900	30	Dry O_2
900	30	N_2

47. Diffusion
ddrive.set

Junction Drive

Temp °C	*Time min*	*Gas*
950	15	Dry O_2
950	15	N_2

48. Diffusion
dbpsg6k.set

Deposit BPSG (4%P, 4%B)
100 nm undoped oxide followed by
500 nm doped oxide

49. Diffusion
dflow925.set

BPSG Flow
This step is performed immediately after step #48

Temp °C	*Time min*	*Gas*
925	15	Dry O_2

50.	Lithography phcoat.set	Remove Back Side Polysilicon
51.	Wet Etch wbpsg6k.set	Strip Back Side BPSG Strip Back Side BPSG - 7:1 buffered HF
52.	Etch plpolyback5k.set	Dry Etch Back Side Polysilicon
53.	Wet Etch wbpsg6k.set	Strip Back Side Oxide Strip Back Side Oxide - 7:1 buffered HF
54.	Ashing ash.set	Plasma Strip Resist
55.	Lithography (10) phcontdec.set	Define Contacts
56.	Etch plbpsg6k.set	Dry Etch Contacts
57.	Ashing ash.set	Plasma Strip Resist
58.	Deposition mcvc1u.set	Sputter Deposition of Metal 1.0 μm AlSi
59.	Lithography (11) phmetaldec.set	Pattern Metal
60.	Etch plmetal.set	Plasma Etch Metal
61.	Ashing ash.set	Plasma Ash resist
62.	Diffusion dsinter.set	Final Sinter[4] *Temp* °C: 400; *Time min*: 5; *Gas*: Forming Gas

Further details of the CMOS process can be found in [30].

Footnotes:

1. The threshold adjust implant for the standard process described in [30] is not used here to obtain surface-channel PMOS devices rather than buried-channel devices. Buried channel devices are obtained by including the threshold-adjust implant.

2. Gate oxidation performed in low-pressure oxidation tube in TRL in various O_2, NH_3 and N_2 ambients, as described in Section 2.1.

3. Exposure of bare polysilicon to HF etch is kept to a minimum; overexposure seems to cause significant loss of yield.

4. The total anneal time at 400°C is 20 min, but the H_2 gas is only on for 5 min.

Appendix C

Deep-Submicron NMOS Process Flow

Step #	Type	Description		
1.	Diffusion	Grow Stress Relief Oxide (40 nm)		
		Temp °C	*Time min*	*Gas*
	dsro430.set	950	100	Dry O_2
		950	30	N_2
2.	Diffusion	LPCVD Nitride Deposition (150 nm)		
	dnit1.5.set	Deposition Temperature = 800°C		
3.	Lithography (1)	Active Area Definition		
	phactivertn.set	Defined both on n-Well and p-Well		
4.	Etch	Plasma Etch Nitride		
	plnit1.5.set			
5.	Implant (1)	P-Field Ion Implant		
	ipfieldpkt.set	*Implant Element*	*Energy (keV)*	*Dose*
		Boron	70	1E13
6.	Ashing	Plasma Ash Resist		
	ash.set			

7. Diffusion
dnwell5.1k.set

Field Oxidation

Temp °C	*Time min*	*Gas*
950	30	Dry O_2
950	175	Wet O_2
950	30	Dry O_2
950	30	N_2

8. Wet Etch
wnit1.5k.set

Nitride Strip
Same as step #8

9. Wet Etch
wsro430.set

Stress Relief Oxide Strip
7:1 buffered HF

10. Diffusion
dgate220.set

Grow Sacrificial Gate Oxide (22 nm)

Temp °C	*Time min*	*Gas*
950	35	Dry O_2
950	30	N_2

11. Implant (2)

n-Channel Punchthrough and Threshold Adjustment Implants

	Implant Element	*Energy (keV)*	*Dose*
invt1pkt.set	Boron	100	6E11
invt2pkt.set	Boron	30	1.5E12

12. Wet etch (TRL)
wgate220.set

Strip Sacrificial Gate Oxide
Dip in 50:1 HF During RCA clean
(after organic clean)
for about 3.5 min. Check sheeting.

13. Diffusion (TRL)
dgate100.set

Gate Oxide[2]
Various Oxide and ROXNOX dielectrics
850°C - 950°C

14. Diffusion
dpoly5k.set

Deposit LPCVD Polysilicon (300 nm)
This step is performed immediately after step #13

15. Diffusion
dphos8.set

Dope Polysilicon ($POCl_3$)

Temp °C	*Time min*	*Gas*
925	60	DryO_2 + N_2
925	15	O_2
925	10	N_2

16. Wet Etch
wphos1.2.set

Strip pglass[3]
7:1 buffered HF

17. Lithography (2)
phpolyrtn.set

Pattern Polysilicon

18. Etch
plpoly5k.set

Plasma Thin Resist[3]
1 min, 100 W, 200 mTorr, O_2 Plasma Etch
Removes 0.6 μm (drawn) line.

19. Etch
plpoly5k.set

Dry Etch Polysilicon

20. Ashing
ash.set

Plasma Strip Resist

21. Implant (3)
in+pkt.set

Implant S/D Regions

Implant Element	*Energy (keV)*	*Dose*
Arsenic	90	7E15

22. Diffusion
dreox.set

Reoxidation

Temp °C	*Time min*	*Gas*
900	30	Dry O_2
900	30	N_2

23. Diffusion
ddrive.set

Junction Drive

Temp °C	*Time min*	*Gas*
950	15	Dry O_2
950	15	N_2

24. Diffusion
dbpsg6k.set

Deposit BPSG (4%P, 4%B)
100 nm undoped oxide followed by
500 nm doped oxide

25. Diffusion
dflow925.set

BPSG Flow
This step is performed immediately after step #24

Temp °C	*Time min*	*Gas*
925	15	Dry O_2

26. Lithography
phcoat.set

Remove Back Side Polysilicon

27.	Wet Etch wbpsg6k.set	Strip Back Side BPSG Strip Back Side BPSG - 7:1 buffered HF		
28.	Etch plpolyback5k.set	Dry Etch Back Side Polysilicon		
29.	Wet Etch wbpsg6k.set	Strip Back Side Oxide Strip Back Side Oxide - 7:1 buffered HF		
30.	Ashing ash.set	Plasma Strip Resist		
31.	Lithography (3) phcontrtn.set	Define Contacts		
32.	Etch plbpsg6k.set	Dry Etch Contacts		
33.	Ashing ash.set	Plasma Strip Resist		
34.	Deposition mcvc1u.set	Sputter Deposition of Metal 1.0 μm AlSi		
35.	Lithography (4) phmetalrtn.set	Pattern Metal		
36.	Etch plmetal.set	Plasma Etch Metal		
37.	Ashing ash.set	Plasma Ash resist		
38.	Diffusion dsinter.set	Final Sinter[4]		
		Temp °C	*Time min*	*Gas*
		400	5	Forming Gas

Footnotes:

1. The threshold adjust implant for the standard process described in [30] is not used here to obtain surface-channel PMOS devices rather than buried-channel devices. Buried channel devices are obtained by including the threshold-adjust

implant.

2. Gate oxidation performed in low-pressure oxidation tube in TRL in various O_2, NH_3 and N_2 ambients, as described in Section 2.1.

3. Plasma thinning of resist is performed using the method described by Chung et al. [31].

4. Exposure of bare polysilicon to HF etch is kept to a minimum; overexposure seems to cause significant loss of yield.

5. The total anneal time at 400°C is 20 min, but the H_2 gas is only on for 5 min.

Appendix D

Poisson Random Processes

The purpose of this appendix is to introduce some general properties of the ideal Poisson process used to model the random telegraph signal, and to present the calculation of the autocorrelation function of the random telegraph signal. These results are used in Chapter 5 to model the noise spectra of single electron traps.

D.1 Statistics

A Poisson process is a process which has discrete changes in state which occur randomly with some rate λ transitions per unit time. The probability $\mathcal{P}(\delta t)$ that a change of state will occur in some infinitesimal interval of time δt is proportional to the transition rate λ, and is proportional to the length of the time interval, i.e.,

$$\mathcal{P}(\delta t) = \lambda \delta t. \tag{D.1}$$

The probability that more than one transition will occur in the infinitesimal time interval δt is negligible. Summing up these transition rates, the probability distribution for the time until the next change in state can be written as

$$\mathcal{P}(t) = \lambda e^{-\lambda t}. \tag{D.2}$$

As presented in Chapter 5, both the capture and emission times for a random telegraph signal have probability distributions which are exponentially distributed, sug-

gesting that they may be modeled as ideal Poisson processes.

Given that a Poisson random process has a rate constant given by λ, the mean time between changes in state τ can be written as

$$\tau = \frac{1}{\lambda}. \tag{D.3}$$

Calculating the sample mean time between changes in state $<\tau>_{sample}$ for a time data capture allows an estimate of the rate of the random Poisson process to be obtained by

$$\lambda_{est} = \frac{1}{<\tau>_{sample}}. \tag{D.4}$$

A random Poisson process also has the interesting property that the variance of the mean time between changes in state is equal to

$$\sigma^2 = \frac{1}{\lambda}. \tag{D.5}$$

D.2 Autocorrelation Function

The autocorrelation function is a quantity relating the state of a system at any time s, to the state of the system at some different time r. The autocorrelation function R is defined as

$$R(s,r) = <I(s)I(r)> - <I(s)>^2, \tag{D.6}$$

where the brackets denote the expected value of the enclosed quantity, and $I(s)$ is the state variable of the system in question. Making the assumption that the random process is stationary allows one to write the autocorrelation function as the function of a single time variable [42]. Defining the variable τ as

$$\tau = s - r, \tag{D.7}$$

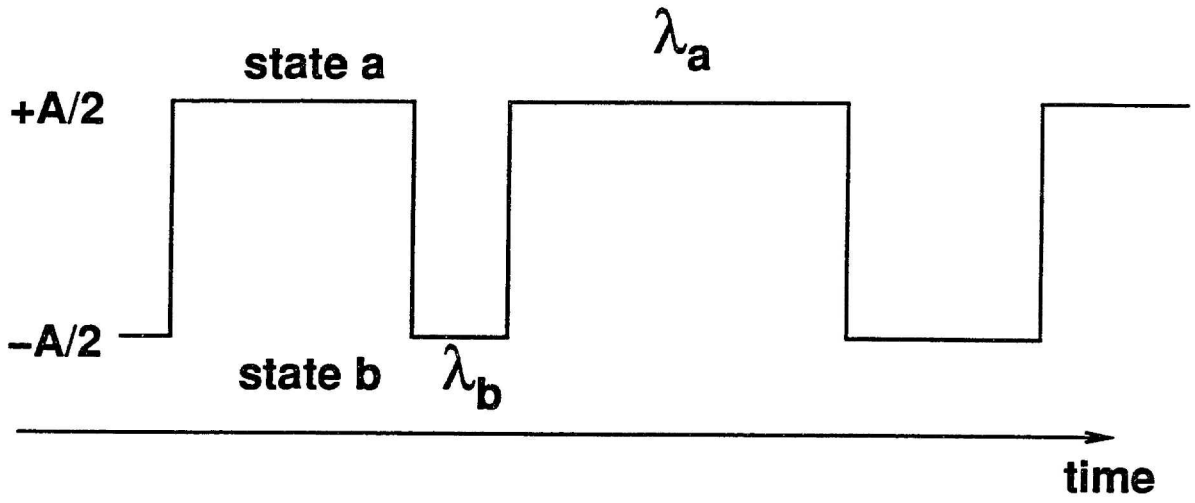


Figure D.1: Diagram of random telegraph signal. The rate of transition from state a to state b is given by λ_a, and the rate of transition from state b to state a is given by λ_b.

allows us to write

$$R(s,\tau) = R(\tau) = < I(s)I(s+\tau) > - < I(s) >^2 . \tag{D.8}$$

Using this notation it is straightforward to derive the autocorrelation function of a random telegraph signal.

The derivation of the autocorrelation function presented here follows the outline given in Kirton and Uren [25]. Let us define a random telegraph process as a process which can be in either of two states, which we will call state a and state b. The rate for the system to change state out of state a is λ_a, and the rate for the system to change out of state b is λ_b. This is shown schematically in Figure D.1. For simplicity, we make the assumption that the amplitude is $+A/2$ when the system is in state a, and the amplitude is $-A/2$ when the system is in state b. The probability that the system will be in a given state at some time chosen completely at random is given by the expressions

$$\mathcal{P}_a = \frac{\lambda_b}{\lambda_a + \lambda_b},$$

$$\mathcal{P}_b = \frac{\lambda_a}{\lambda_a + \lambda_b}. \tag{D.9}$$

We can introduce the notation of the probability $\mathcal{P}_{11}(t)$ that the system will go through an even number of transitions in time t, and the probability $\mathcal{P}_{10}(t)$ that the system will go through an odd number of transitions in time t. These two quantities are related by the expression

$$\mathcal{P}_{11}(t) + \mathcal{P}_{10}(t) = 1. \tag{D.10}$$

Assuming that the system is in state a at time t_o, we can write the probability of having an even number of transitions in the time $(t + \delta t)$ following t_o as the sum of the probability of making an odd number of transitions in time interval t, followed by a single transition in time interval δt, plus the probability of making an even number of transitions in time interval t, followed by no transitions in time interval δt. This is written as

$$\mathcal{P}_{11}(t + \delta t|a) = \mathcal{P}_{10}(t|a)\lambda_b \delta t + \mathcal{P}_{11}(t|a)(1 - \lambda_a \delta t). \tag{D.11}$$

Making use of Equation (D.10), this expression can be simplified to

$$\mathcal{P}_{11}(t + \delta t|a) = (1 - \mathcal{P}_{11}(t|a))\lambda_b \delta t + \mathcal{P}_{11}(t|a)(1 - \lambda_a \delta t) =$$

$$\mathcal{P}_{11}(t|a) + [\lambda_b - (\lambda_a + \lambda_b)\mathcal{P}_{11}(t|a)]\,\delta t. \tag{D.12}$$

This in turn can be written as a differential equation

$$\lim_{\delta t \to 0} \frac{\mathcal{P}_{11}(t + \delta t|a) - \mathcal{P}_{11}(t|a)}{\delta t} = \frac{\partial \mathcal{P}_{11}(t|a)}{\partial t} = \lambda_b - (\lambda_a + \lambda_b)\mathcal{P}_{11}(t|a), \tag{D.13}$$

which can be solved in close' form to yield the expressions

$$\mathcal{P}_{11}(t|a) = \frac{\lambda_b}{\lambda_a + \lambda_b} + \frac{\lambda_a}{\lambda_a + \lambda_b} e^{-(\lambda_a + \lambda_b)t},$$

$$\mathcal{P}_{10}(t|a) = \frac{\lambda_a}{\lambda_a + \lambda_b} - \frac{\lambda_a}{\lambda_a + \lambda_b} e^{-(\lambda_a + \lambda_b)t}. \tag{D.14}$$

The corresponding transition probabilities for starting in state b are easily found by switching the a and b subscripts.

Using the quantities from the above and Equation (D.6), the autocorrelation function can be written as

$$R(\tau)=\left(\frac{A}{2}\right)^2\left[\mathcal{P}_a\left(\mathcal{P}_{11}(\tau|a)-\mathcal{P}_{10}(\tau|a)\right)-\mathcal{P}_b\left(\mathcal{P}_{11}(\tau|b)-\mathcal{P}_{10}(\tau|b)\right)-\frac{1}{4}\left(\frac{\lambda_a-\lambda_b}{\lambda_a+\lambda_b}\right)^2\right], \tag{D.15}$$

which yields the relation

$$R(\tau)=A^2\frac{\lambda_a\lambda_b}{(\lambda_a+\lambda_b)^2}e^{-(\lambda_a+\lambda_b)\tau}. \tag{D.16}$$

The autocorrelation expression above is valid only for values of τ greater than zero. The autocorrelation function also provides information about the correlation between the current state of a random process and its past values. Therefore, by symmetry, we can write the autocorrelation function for all τ as

$$R(\tau)=A^2\frac{\lambda_a\lambda_b}{(\lambda_a+\lambda_b)^2}e^{-(\lambda_a+\lambda_b)|\tau|}. \tag{D.17}$$

This equation is an equivalent way to express Equation (5.14).

D.3 Poisson Area Distribution

A random distribution of traps over a surface can be described by Poisson statistics. For this case, the Poisson random variable is dependent on the state variable area, and not on time. For the case of a Poisson surface distribution, we say that the probability of finding a single trap in a differential area element δA is proportional to the density of single traps λ_d and is proportional to the area of the differential surface element as

$$\mathcal{P}(\delta A)=\lambda_d\delta A. \tag{D.18}$$

We will can also say that if the differential surface element is made small enough, the probability of finding two traps in that surface element is identically zero. Summing up these differential surface elements, the probability of finding n traps on a surface of area A can be written as [72]

$$\mathcal{P}_n(A) = \frac{(\lambda_d A)^n}{n!} e^{-\lambda_d A}. \tag{D.19}$$

For example, the probability of finding only one trap can be written as

$$\mathcal{P}_n(A) = \lambda_d A e^{-\lambda_d A}. \tag{D.20}$$

The expected number of traps on a surface of area A can be written as

$$E(n) = \lambda_d A. \tag{D.21}$$

The variance of the number of traps on the same surface can be written as [72]

$$\mathrm{Var}(n) = \sigma_n^2 = \lambda_d A. \tag{D.22}$$

The standard deviation of a random variable, normalized by the mean value of that random variable, is useful as a measure of the average fractional deviation of a random variable from its mean value. Using the above two relations, we can write the normalized standard deviation of the number of traps on a surface of area A as

$$\frac{\sigma_n}{E(n)} = \frac{\sqrt{\lambda_d A}}{\lambda_d A} = \frac{1}{\sqrt{\lambda_d A}}. \tag{D.23}$$

This equation is very interesting in that is demonstrates that as the area of the surface in question becomes smaller and smaller, the relative variation of the number of traps on a surface becomes larger and larger.

As an example, consider a MOSFET device with a density of surface traps of $\lambda_d = 1 \times 10^9$ cm^{-2} over a surface area of 20×10 μm. The mean number of traps in

this surface can be solved to be $E(n) = 2000$. The standard deviation for this case is equal to about 45, which corresponds to a normalized variation of 2.2%. For this case, it is expected that 95% of the 20×10 μm devices will have a number of traps within about ±5% of the mean value. For a device of size 5×5 μm the expected number of traps is only $E(n) = 250$, but the normalized variation is 6.3%, meaning that 95% of the devices will be within approximately ±12.5% of the mean value. This variation is solely due to the traps being randomly distributed across the surface and as such is a fundamental limit to the controllability of the variation of device noise from device to device. It should be noted that this latter variation is consistent with the measurement variation observed when characterizing devices of size comparable to 5×5 μm.

Appendix E

Hole in Ideal Conducting Sheet

The purpose of this Appendix is to demonstrate that the effect of removing a single charge from the channel is *not* a strong function of the location of the trap in the channel region. It is demonstrated here that as long as a trap is not within a single screening distance a of the boundaries of the channel, the effect of that trap on the drain current fluctuation is independent of position in the channel region.

The effect that removing a single charge from the channel has on the drain current of a MOS device will be determined by analogy with the effect that a hole in an ideal conducting sheet has on the resistance of the conducting sheet measured at the terminals[1]. The trapping of an electron from the channel will cause the channel region local to this trap to have lower conductivity. For a lightly inverted surface, it is conceivable that the trapping of an electron could cause the channel to be depleted of mobile carriers in the vicinity of this trap. Hence the analogy to a 'hole' in a ideal conducting sheet.

[1]The term 'hole' in this appendix is used to mean a macroscopic non-conducting region, and not a hole as a quasi-particle from solid-state physics.

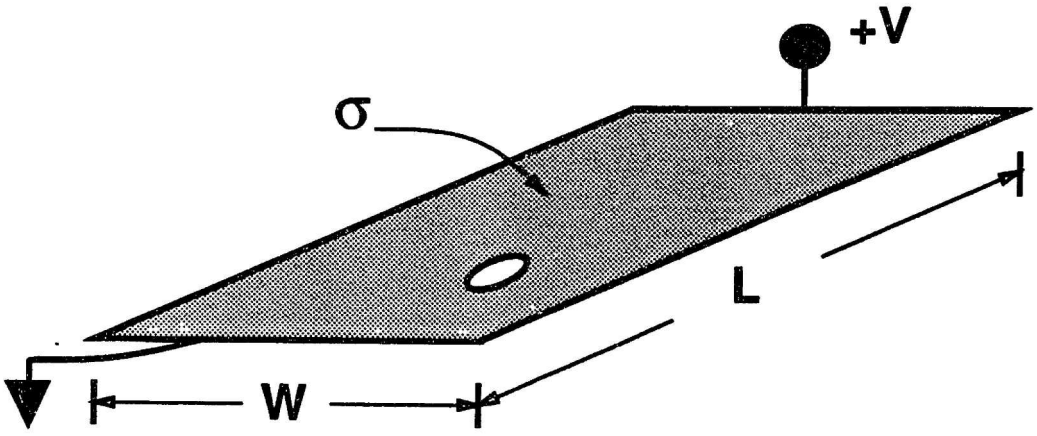


Figure E.1: Diagram of the hole in an ideal conducting sheet problem.

E.1 Hole with Zero Conductivity

A diagram of the hole in an ideal conducting sheet problem is shown in Figure E.1. The problem is that of a sheet of uniform conductivity, with a hole of radius a located at point (x_o, y_o). This problem can be solved by a superposition of a sheet current J_o with a source and sink of current located very close to each other in the $\hat{x}$-direction and centered at the point (x_o, y_o). The source and sink of current form a doublet, with the current they emit being given by the equation,

$$\vec{J}_{doublet}(r,\theta) = \frac{\mu}{4\pi r^2}\cos\theta\,\hat{r} + \frac{\mu}{4\pi r^2}\sin\theta\,\hat{\theta}, \tag{E.1}$$

with coordinates r and θ centered on the point (x_o, y_o). The amplitude of the doublet, μ, can be solved for by using the boundary condition that nc current can flow into the hole; that is, that the radially directed current at $r = a$ must be identically zero. Assuming that all of the sheet current flows in the $-\hat{x}$ direction, we can write the

sheet current as

$$\vec{J}_{sheet}(r,\theta) = -J_o \cos\theta\,\hat{r} + J_o \sin\theta\,\hat{\theta}. \tag{E.2}$$

Solving for the current passing through the boundary at the edge of the hole (at $r = a$), we have

$$(\vec{J}(r,\theta)\cdot\hat{r})\Big|_{r=a} = \frac{\mu}{4\pi a^2}\cos\theta - J_o\cos\theta = 0 \tag{E.3}$$

So we find with

$$\mu = 4\pi a^2 J_o. \tag{E.4}$$

From the preceding, it can be seen that by using a superposition of a uniform current with a doublet of current, we can satisfy the insulating boundary conditions at the edges of the hole.

Until now, I have been speaking of current flow which is directly analogous to electric field lines. The doublet of current can be directly replaced with a doublet of charge, and the uniform sheet current can be directly replaced with a uniform $-\hat{x}$-directed electric field, without changing any of the previous results, or any loss of generality. J can be converted to electric field by the relation $\vec{J} = \sigma_o \vec{E}$. I make these substitutions now because it is easier to talk about electric potentials, rather than hydrodynamic flow potentials, in solving the rest of the problem[2].

The other boundary conditions are that of an insulating boundary at $y = 0$ and $y = W$, and that of perfect conductors at the boundaries at $x = 0$ and $x = L$. The easiest way to satisfy these boundary conditions is by inspection using the method of images (see Shen and Kong [121], pg. 326, for example). To satisfy the insulating boundary conditions, the method of images requires us to place image doublets at the points $-y_o$, and at the point $2W - y_o$. These image doublets will in turn require additional image doublets be placed in relation to the opposite boundary; in this case,

[2]Note that the units of $\vec{J}$ is A/cm, and the units of σ_o is S because we are talking about sheet currents. Also recall that a point charge in a conducting medium acts as a point current source.

at $2W + y_o$ and $-2W + y_o$, respectively. These additional image doublets will in turn require additional images, *ad infinitum*. The image doublets will be located at

$$Y_I = \pm y_o, \pm(2W \pm y_o), \pm(4W \pm y_o), ..., \pm(2nW \pm y_o), ... \tag{E.5}$$

The perfect conductor boundary conditions can be satisfied by inspection in a analogous way, requiring infinite one-dimensional arrays of doublets (specified by the Y_I points above) at each of the points X_I given by

$$X_I = \pm x_o, \pm(2L \pm x_o), \pm(4L \pm x_o), ..., \pm(2mL \pm x_o), ... \tag{E.6}$$

The system of doublets we are left with is shown schematically in Figure E.2. Note

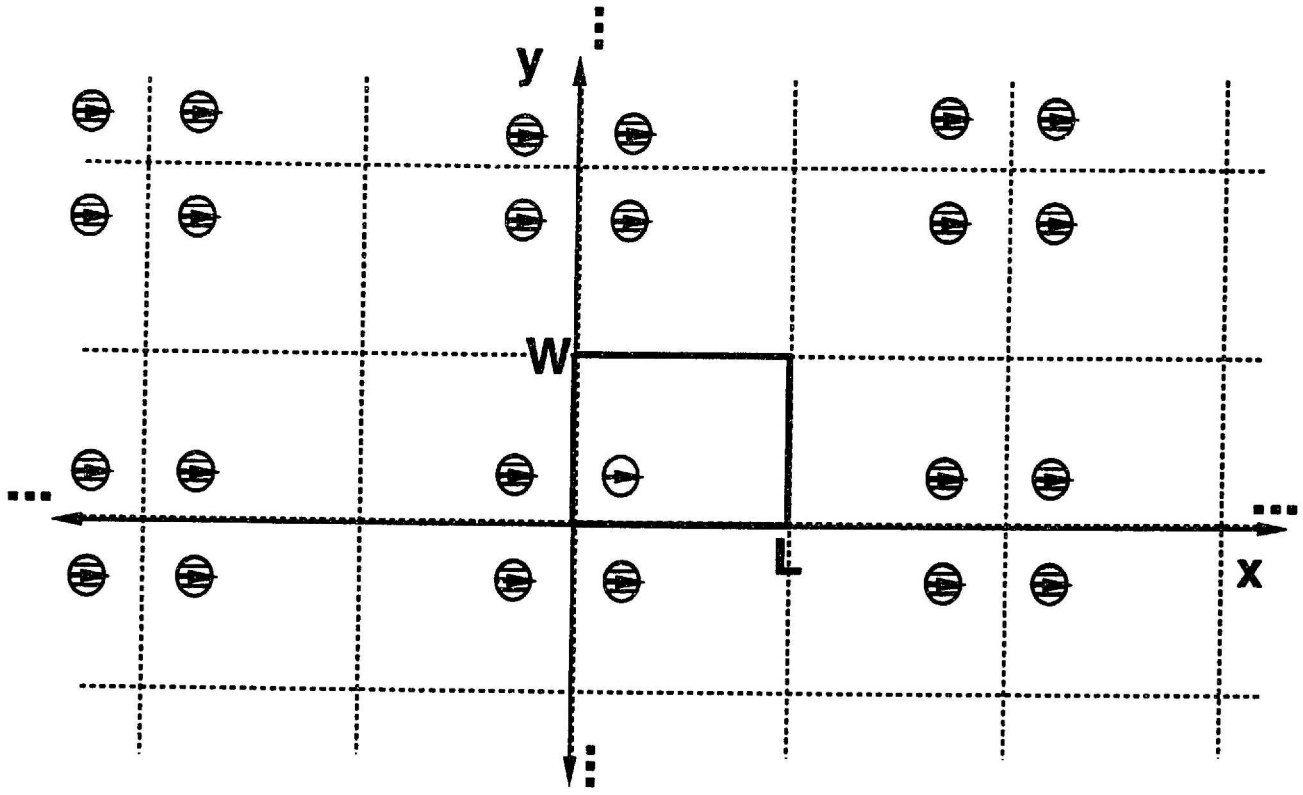


Figure E.2: Infinite array of $\hat{x}$-oriented doublets required to satisfy the insulating boundary conditions at $y = 0, W$, and the perfectly conducting boundary conditions at $x = 0, L$. The cross-hatched doublets are all images.

that these image doublets will cause a distortion in the shape of the hole from a perfect circle. This distortion is greatest when the hole is located near any of the conducting sheet boundaries. At the end of this section, it will be shown that this distortion is very small even for holes located a fraction of a hole radius a away from any of the conducting sheet boundaries.

The electric potential for a single $\hat{x}$-oriented doublet located at (x_o, y_o) is given by

$$\phi(x,y) = \frac{J_o a^2}{\sigma_o} \frac{x - x_o}{(x-x_o)^2 + (y-y_o)^2}. \tag{E.7}$$

The electric potential for the entire system is found by inspection to be

$$\Phi(x,y) = \frac{Vx}{L} + \frac{J_o a^2}{\sigma_o} \sum_{allX_I} \sum_{allY_I} \frac{x - X_I}{(x-X_I)^2 + (y-Y_I)^2}, \tag{E.8}$$

where the X_I and the Y_I are as listed in Equations (E.6) and (E.5).

We are interested in finding the change in the drain current caused by the introduction of the hole into the conducting sheet. For this purpose, we want to solve for the electric field at the perfectly conducting plate (choose the $x = 0$ terminal for simplicity). Following normalization by the effective length L, the electric field is given by

$$\begin{aligned}
\vec{E}(x,y) &= -\nabla\Phi(x,y) \\
&= -\frac{V}{L}\hat{x} - \frac{J_o a^2}{L\sigma_o} \sum_{allX_I} \sum_{allY_I} \frac{1}{L} \frac{(\frac{y-Y_I}{L})^2 - (\frac{x-X_I}{L})^2}{\left[(\frac{x-X_I}{L})^2 + (\frac{y-Y_I}{L})^2\right]^2} \hat{x} \\
&\quad + \frac{J_o a^2}{L\sigma_o} \sum_{allX_I} \sum_{allY_I} \frac{1}{L} \frac{2(\frac{y-Y_I}{L})(\frac{x-X_I}{L})}{\left[(\frac{x-X_I}{L})^2 + (\frac{y-Y_I}{L})^2\right]^2} \hat{y}
\end{aligned} \tag{E.9}$$

Using the above relation, we can solve for the total current through this conducting sheet as

$$\begin{aligned}
I &= \int_0^W dy\, \sigma_o \vec{E}(x,y) \cdot \hat{x}\Big|_{x=0} \\
&= -\frac{VW\sigma_o}{L} - \frac{J_o a^2}{L} \sum_{allX_I} \sum_{allY_I} \left[\frac{\frac{Y_I - W}{L}}{(\frac{X_I}{L})^2 + (\frac{W-Y_I}{L})^2} - \frac{\frac{Y_I}{L}}{(\frac{X_I}{L})^2 + (\frac{Y_I}{L})^2} \right].
\end{aligned} \tag{E.10}$$

Note that the $\hat{y}$-term will vanish when the dot product with $\hat{x}$ is taken. This equation can readily be seen to consist of a large signal component $VW\sigma_o/L$ and a small signal component. We can write the current from Equation (E.10) as

$$I = I_o + \delta I. \tag{E.11}$$

We will deal with only the small signal component, normalized by the large signal component, from here on. The term in brackets in Equation (E.10), when summed over all Y_I, appears on first inspection to be equal to 0 by the symmetry of the problem. The sum over all Y_I does approach 0 very slowly. Taking the term due to the real doublet, plus the first N sets of *image* terms, up to $(\pm NW + y_o)$ where N is an even number, we can show that

$$\sum_{allY_I}^{(\pm NW+y_o)} \left[\frac{\frac{Y_I - W}{L}}{(\frac{X_I}{L})^2 + (\frac{W-Y_I}{L})^2} - \frac{\frac{Y_I}{L}}{(\frac{X_I}{L})^2 + (\frac{Y_I}{L})^2} \right]$$

$$= -\frac{\frac{NW+y_o}{L}}{(\frac{NW+y_o}{L})^2 + (\frac{X_I}{L})^2} - \frac{\frac{(N+1)W-y_o}{L}}{(\frac{(N+1)W-y_o}{L})^2 + (\frac{X_I}{L})^2}. \tag{E.12}$$

All of the previous terms in the sum exactly cancel. Therefore, we can write the normalized small signal part of the current through the conducting sheet as

$$\frac{\delta I}{I_o} = \frac{a^2}{LW} \sum_{allX_I} \lim_{N\to\infty} \frac{-\frac{NW+y_o}{L}}{(\frac{NW+y_o}{L})^2 + (\frac{X_I}{L})^2} - \frac{\frac{(N+1)W-y_o}{L}}{(\frac{(N+1)W-y_o}{L})^2 + (\frac{X_I}{L})^2}. \tag{E.13}$$

From this equation, we can split the sum into two halfs, one over the $X_I = (2mL + y_o)$, and the other over $X_I = (2mL - y_o)$, for m being all of the integers between $-\infty$ and $+\infty$. I will solve for only one half of this sum, with the other half following by symmetry. In a like manner, I will only solve for the first term of Equation (E.13), with the second term following in an identical manner by symmetry. In effect, we will be solving for

$$\frac{\delta I}{4I}.$$

Inserting the above X_I's into the first term of Equation (E.13) and multiplying numerator and denominator by $[L/(2NW+y_o)]^2$, we are left with

$$\frac{\delta I}{4I_o} = \frac{a^2}{LW} \sum_{m=-\infty}^{+\infty} \lim_{N\to\infty} \frac{-\frac{L}{2NW+y_o}}{1+(\frac{2mL+y_o}{2NW+y_o})^2} \tag{E.14}$$

It is easy to see that from one value of m to the next, the value of the term in the sum varies by a insignificant amount in the limit of large N. Thus we can approximate the value of the term in the sum as an integral

$$\frac{\delta I}{4I} \simeq \frac{a^2}{LW} \lim_{N\to\infty} \sum_{m=-\infty}^{+\infty} \frac{1}{2L} \int_{(2mL+y_o)-L}^{(2mL+y_o)+L} \frac{-\frac{L}{2NW+y_o}}{1+(\frac{X}{2NW+y_o})^2}\, dX, \tag{E.15}$$

where X is an arbitrary integration variable. We see that in the limit of large N ($2NW \gg L$), the approximation is exactly correct. From this equation it is easy to make the final simplification to the integral equation

$$\frac{\delta I}{4I_o} = \frac{a^2}{LW} \lim_{N\to\infty} \frac{1}{2L} \int_{-\infty}^{+\infty} \frac{-\frac{L}{2NW+y_o}}{1+(\frac{X}{2NW+y_o})^2}\, dX = \frac{a^2}{LW} \lim_{N\to\infty} (-\frac{\pi}{2}). \tag{E.16}$$

We can therefore write the perturbed part of the current as

$$\frac{\delta I}{I_o} = -\frac{2\pi a^2}{LW}. \tag{E.17}$$

Thus we see that the change in conductance by putting a hole in an ideal conducting sheet is proportional to $1/LW$ and to the area of the hole, and is independent of the position of the hole in relation to the electrodes to first order. Note that it is expected that some edge effects will be present, which by symmetry could reduce the fluctuation by as much as a factor of 2.

As was mentioned near the beginning of the section, the image doublets inserted to satisfy the conducting sheet boundary conditions will cause a slight distortion in the shape of the hole in the conducting sheet. This distortion is illustrated in Figure E.3, where we have plotted the flow lines for the case of a hole near a single

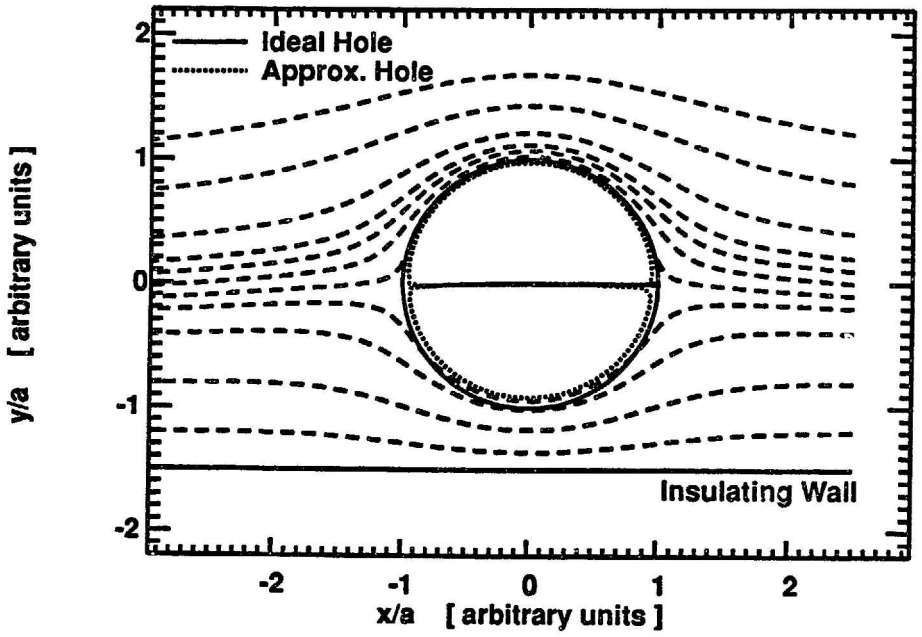


Figure E.3: Flow lines for a hole near an insulating boundary which satisfy Equation (E.9). The edge of the hole is $a/2$ away from the insulating boundary. The solid line shows the ideal hole in the conducting sheet, and the dotted line shows the effect of the image doublets to distort the shape of the ideal hole. Note that even though the hole is quite close to the insulating boundary, the distortion introduced by the image doublets is small.

insulating boundary and far away from all other boundaries. The edge of the hole for this case is $a/2$ away from the boundary. The ideal hole is shown by the solid line. The dotted lines show the boundaries of the approximation to the ideal hole which satisfies Equation (E.10). Note that the distortion in the shape of the hole is very small, even though the hole is close to the insulating boundary.

The net result of this distortion is that we have *a solution* to the hole in a conducting sheet problem for a hole or radius a' which is slightly smaller than a, but which produces the same perturbation in drain current as a hole of radius a located far away from all edges (from Equation (E.17)). Thus the magnitude of drain current fluctuations caused by a hole near any of the edges of the conducting sheet will be higher than that caused by a same-sized hole far away from any of the edges. For the case shown in Figure E.3, the perturbation in drain current is only $\sim 11\%$ higher than if the hole were located far from any edges. Because of these observations, we can say that the perturbation in drain current caused by introducing a 'hole' into a conducting sheet is relatively independent of the position of the hole.

We can quantify the change in the size of the hole due to a single image doublet by noting that if the separation between the image dipole and the actual dipole is sufficient, the field which occurs at the actual dipole due to the image dipole is relatively planar. As is shown from Figure E.3, the deviation from planarity does not produce significant distortion in the shape of the hole from perfectly circular even for the hole located very close to one of the boundaries. The error in estimating hole area can be then estimated using Equation (E.4) to be

$$\left(\frac{a'}{a}\right)^2 = \frac{1}{1+\frac{a^2}{4y_o^2}}. \tag{E.18}$$

For the case of Figure E.3, the error in estimating the hole area is $\sim$11%.

E.2 Hole with Finite Conductivity

Using superposition, we can readily write the solution to the problem if the hole is a region of finite conductivity σ_1. If we add the constant current I_1 due to a uniform conducting sheet of conductivity σ_1 to Equation (E.11), we have

$$I = I_o + \delta I + I_1. \tag{E.19}$$

The sheet conductivity is increased to $\sigma_{total} = \sigma_o + \sigma_1$, and the large signal component of I is increased to

$$I_o' = I_o + I_1 = \frac{\sigma_o + \sigma_1}{\sigma_o} I_o. \tag{E.20}$$

Using this new value for the large signal component of I, and noting that the change in conductivity between the conducting sheet and the hole is

$$\delta\sigma = (\sigma_o + \sigma_1) - \sigma_1 = \sigma_o, \tag{E.21}$$

we can rewrite Equation (E.17) above as

$$\frac{\delta I}{I_o'} = -\frac{2\pi a^2}{LW} \frac{\delta\sigma}{\sigma_{total}}, \tag{E.22}$$

which is similar to the result reported by Simoen et al. [70].

Appendix F

Quantum Subband Simulation

A simulation of the inversion layer was performed following the method outlined by Stern and Howard [75]. This method has been the subject of numerous work since, as outlined in the review article by Ando et al. [79].

F.1 Outline of Problem

This simulation was performed assuming the effective mass approximation. The solution to the inversion layer wave functions is found in a self-consistent way. First, the wave functions are obtained from the Schrödinger equation assuming a certain potential function V(z) in the inversion layer. Next, the amount of charge which results from these wave functions is calculated and used to produce a new value of the potential function in the inversion layer by the use of the Poisson equation. This new value of the potential function is then used to calculate a new estimate of the wave functions using the Schrödinger equation and so on. This procedure is repeated to converge to a self-consistent value of the potential function V(z) and wave functions which satisfy both the Schrödinger and the Poisson equations. The notation and physical constants used in this simulation are the same as used by Stern [80]. Many-body effects in the inversion layer are expected to be increasingly important at higher

bias levels [79]. These effects are neglected, but it is believed that the simulation results are still qualitatively valid at the high bias levels.

The inversion layer carriers are confined in the $\hat{z}$-direction near the surface by the depletion region electric field, which forms a potential well at the surface. In the directions parallel to the interface (the $\hat{x}$- and $\hat{y}$-directions), the inversion layer carriers are described by plane wave states. The inversion layer wave function can therefore be written as

$$\psi_i(x,y,z) = \zeta_i(z)e^{jk_x x + jk_y y}, \tag{F.1}$$

where k_x and k_y are the wave-vector components for motion parallel to the surface, and $\zeta_i(z)$ is the envelope function resulting from the confinement of electrons in the narrow potential well. The corresponding energy levels for this wave function take the form

$$E = E_i + \frac{\hbar^2}{2m_{||}}(k_x^2 + k_y^2), \tag{F.2}$$

where $m_{||}$ is the effective mass for motion parallel to the interface.

We can solve for the values of $\zeta_i(z)$ and E_i for the case of one electron in the inversion layer by solving the Schrödinger equation

$$-\frac{\hbar^2}{2m_\perp}\nabla^2\zeta_i(z) + V(z)\zeta_i(z) = E_i\zeta_i(z), \tag{F.3}$$

where $m_\perp$ is the effective mass for motion perpendicular to the interface (the $\hat{z}$-direction). Each of the solutions E_i of Equation (F.3) represents the ground-state energy of a sub-band of allowed states. Due to the two dimensional symmetry of the problem, each subband will have a density of states which is uniform in energy above the ground state. The series of solutions E_i, $i = 0, 1, 2, ...$, for a given conduction-band valley is called a ladder of subband solutions.

The conduction band of silicon has six equivalent minima or valleys in a Brillouin zone. These minima are shown schematically as Fermi-surfaces in Figure 5.9, which

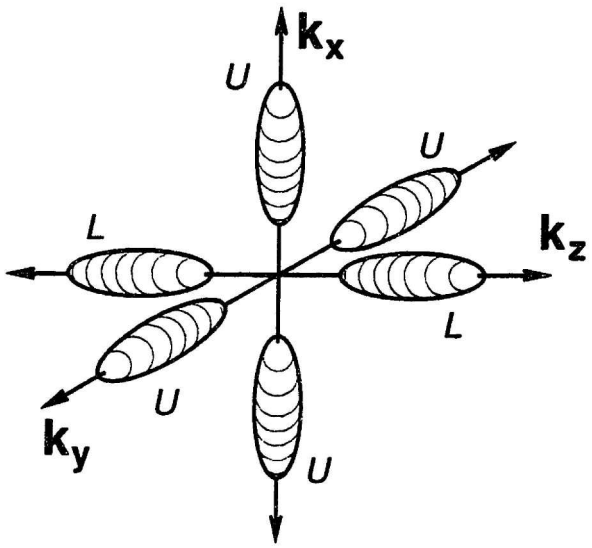


Figure F.1: Fermi surfaces of a Brillouin zone for silicon.

is reproduced here as Figure F.1. The two valleys parallel to the $\hat{z}$-direction (labeled L in Figure F.1) are degenerate with each other, have the highest mass for motion in the $\hat{z}$-direction (i.e. the highest value of $m_\perp$ in Equation (F.3) above), and so will have the lowest energy levels. The solutions of Equation (F.3) above for these two valleys is said to be the lower-ladder of subband solutions and have the indices $i = 0, 1, 2, ...$ The other four conduction band valleys (labeled U in Figure F.1) are degenerate with each other, have the lowest mass for motion in the $\hat{z}$-direction, and so have the highest energy levels. The solutions of Equation (F.3) above for these four valleys is called the upper-ladder of subband solutions, and uses the indices $i = 0', 1', 2', ...$

Given a set of wave function solutions and the value of the Fermi-energy E_f, the carrier concentration in the ith subband N_i can be found using Fermi-Dirac statistics as

$$N_i = \left(\frac{n_{vi} m_{di} kT}{\pi \hbar^2} \right) F_o \left[(E_f - E_i)/kT \right], \tag{F.4}$$

where $F_o(x) = \ln(1 + e^x)$, n_{vi} is the valley degeneracy, and m_{di} is the density-of-states effective mass per valley [80]. Using these values for the density of carriers in each

subband, the total inversion plus depletion charge density $qN(z)$ can be found to be

$$qN(z) = -qN_d(z) - \sum_{\text{all } i,i'} qN_i|\zeta_i(z)|^2, \tag{F.5}$$

where $N_d(z)$ is the density of ionized donors in the depletion region. Note that the envelope function is properly normalized such that

$$\int_0^{+\infty} |\zeta_i(z)|^2 dz = 1. \tag{F.6}$$

The value of potential which corresponds to this charge distribution can be solved for using the Poisson equation

$$\frac{\partial^2 \phi(z)}{\partial z^2} = \frac{qN(z)}{\epsilon_{Si}}, \tag{F.7}$$

and noting that $V(z) = -q\phi(z)$. The zero of potential for the entire system in the simulation is chosen to be at the interface, such that $V(0) = 0.0$ eV.

These are the complete set of equations needed to arrive at a self-consistent solution to the inversion layer wave functions.

F.2 Numerical Methods

As long as the inversion charge produces a negligible perturbation in the potential well, the above set of equations can be solved in closed form. For most practical applications, the amount of inversion charge will produce a significant change in the potential in the inversion layer, and numerical methods are required.

The above continuous one-dimensional equations are converted to vector form to obtain a numerical solution. The conversion to vector form is accomplished by transforming the one-dimensional continuous variables $f(z)$ to a one-dimensional m-element array $f[n]$. In vector form the solution to the above equations is a set of eigenvectors $\vec{\zeta}_i$, the nth element of which is denoted as $\zeta_i[n]$. To accomplish this

conversion, we use the transformations

$$\begin{aligned} z &\Longleftrightarrow n \cdot du, \\ \zeta_i(z) &\Longleftrightarrow \zeta_i[n], \\ \frac{\partial^2}{\partial z^2}\zeta_i(z) &\Longleftrightarrow \frac{\zeta_i[n-1] - 2\zeta_i[n] + \zeta_i[n+1]}{(du)^2}. \end{aligned} \tag{F.8}$$

The Schrödinger's equation can be written in vector form as

$$\left(\underline{\overline{H_i}} + \underline{\overline{I}}\vec{V}\right)\vec{\zeta_i} = E_i\vec{\zeta_i}, \tag{F.9}$$

where $\underline{\overline{H_i}}$ is the Hamiltonian matrix for the ith subband, $\underline{\overline{I}}$ is the identity matrix, $\vec{V}$ is the potential vector, and $\vec{\zeta_i}$ is the ith eigenvector. The Hamiltonian matrix used for the simplest case of an infinite-oxide-barrier boundary conditions is written as

$$\underline{\overline{H_i}} = -\frac{\hbar^2}{2m_{\perp,i}(du)^2}\begin{bmatrix} -2 & \boxed{0} & 0 & 0 & & \cdots & & 0 \\ 1 & -2 & 1 & 0 & & \cdots & & 0 \\ 0 & 1 & -2 & 1 & & \cdots & & 0 \\ 0 & 0 & 1 & -2 & & \cdots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & & & \vdots \\ 0 & & \cdots & & -2 & 1 & 0 & 0 \\ 0 & & \cdots & & 1 & -2 & 1 & 0 \\ 0 & & \cdots & & 0 & 1 & -2 & 1 \\ 0 & & \cdots & & 0 & 0 & 1 & -2 \end{bmatrix}. \tag{F.10}$$

Using this equation, the boundary condition for the infinite oxide barrier is enforced by setting $\zeta_i[0] = 0$, and setting the $2nd$ element in the top row of the Hamiltonian to be equal to zero.

The initial potential-vector $\vec{V}$, the initial wave functions eigenvectors $\vec{\zeta_i}$, and the initial energy-level/eigenvalues E_i are solved for assuming no inversion layer charge. Then an initial value of the Fermi level E_f is assumed which results in a small density of inversion layer charge and a self-consistent solution for this Fermi level is obtained by the following method:

1. Using the Fermi level E_f, the charge density vector $q\vec{N}$ is determined.

2. Using this charge density, a new potential vector $\vec{V}_{new}$ is determined as the solution to Poisson's equation. Subtracting the initial potential $\vec{V}$ from this yields a difference potential,

$$\vec{V}_{diff} = \vec{V}_{new} - \vec{V}. \tag{F.11}$$

3. A fraction of the difference potential is added to the old potential to obtain

$$\vec{V}_{try} = \vec{V} + f\vec{V}_{diff}. \tag{F.12}$$

   The value of f must be made less than one to avoid problems of numerical convergence.

4. The Schrödinger equation is solved using the potential vector $\vec{V}_{try}$ to obtain a new set of eigenvectors $\vec{\zeta}_i$ and a new set of eigenvalues E_i. Note that solutions are found for both ladders of subbands.

The above steps are repeated until the magnitude of the difference vector $\vec{V}_{diff}$ approaches zero. In the program implementation, a small threshold value is used to signal when the magnitude of $\vec{V}_{diff}$ is close to zero. When this occurs, a solution for that particular Fermi level has been converged to. The value of the Fermi level is then incremented by a small amount, and the above steps are repeated to converge to a solution for that new Fermi level. This process of incrementing the Fermi level and converging to a solution is repeated until the inversion charge density has reached the desired target amount.

F.3 Eigenvalue Problems

We have decomposed the problem into finding the eigenvalues and eigenvectors of

$$\left(\underline{\overline{H_i}} + \underline{\overline{I}}\vec{V}\right)\vec{\zeta}_i = \underline{\overline{A}}\vec{\zeta}_i = E_i\vec{\zeta}_i. \tag{F.13}$$

The program uses different inverse power methods to solve the eigenvalue problem. The following descriptions of the inverse power methods are taken from Strang [122]. The basic idea behind the inverse power method is that in general any vector $\vec{x}$ can be written as a sum of eigenvectors $\vec{e_i}$ each with eigenvalue λ_i since the eigenvectors form a complete orthonormal basis

$$\vec{x} = c_1\vec{e_1} + c_2\vec{e_2} + c_3\vec{e_3} + \cdots + c_n\vec{e_n}. \tag{F.14}$$

Dividing $\vec{x}$ by the Hamiltonian matrix $\underline{\overline{A}}$ will yield a new vector

$$\underline{\overline{A}}^{-1}\vec{x} = c_1\lambda_1^{-1}\vec{e_1} + c_2\lambda_2^{-1}\vec{e_2} + c_3^{-1}\lambda_3\vec{e_3} + \cdots + c_n\lambda_n^{-1}\vec{e_n}. \tag{F.15}$$

Repeating this process will yield a vector which converges to the eigenvector with the smallest eigenvalue. Assuming that $\lambda_1 < \lambda_2, \lambda_3, \cdots, \lambda_n$, we can write this case as m gets large to be

$$\underline{\overline{A}}^{-m}\vec{x} = c_1\lambda_1^{-m}\vec{e_1} + c_2\lambda_2^{-m}\vec{e_2} + c_3\lambda_3^{-m}\vec{e_3} + \cdots \simeq c_1\lambda_1^{-m}\vec{e_1}. \tag{F.16}$$

The number of iterations m must be sufficient to reduce all the other eigenvector components into the numerical noise.

The shifted inverse method is also used in the program to solve the eigenvalue problem for eigenvalues which are higher than the lowest value. Guessing an eigenvalue λ_g and computing $(\underline{\overline{A}} - \underline{\overline{I}}\lambda_g)^{-1}\vec{x}$ will pick out the eigenvector with an eigenvalue closest to λ_g. If the guess eigenvalue is very close to the sth eigenvalue, we can write

$$(\underline{\overline{A}} - \underline{\overline{I}}\lambda_g)^{-1}\vec{x} = \frac{c_1}{\lambda_1 - \lambda_g}\vec{e_1} + \frac{c_2}{\lambda_2 - \lambda_g}\vec{e_2} + \cdots + \frac{c_s}{\lambda_s - \lambda_g}\vec{e_s} + \cdots \simeq \frac{c_s}{\lambda_s - \lambda_g}\vec{e_s}. \tag{F.17}$$

In the program the current guess-vector is normalized to a magnitude of 1 after each matrix division. It is observed in the simulation that the straight inverse power series method converges rapidly to the proper eigenvalue, but is slow to converge to

the proper eigenvector. The shifted-inverse method compliments this well. Given that one can guess the desired eigenvalue to an accuracy of within ϵ, the shifted-inverse method effectively multiplies the proper-eigenvector component of the guess-vector by $1/\epsilon$ which becomes very large if ϵ is very small! The shifted-inverse method converges rapidly to the proper eigenvector but gives no information about the proper eigenvalue. The two methods are used alternately in the program to achieve rapid convergence.

It should be noted that division of a vector by an array is very straightforward to compute. If the solution to

$$\overline{A}^{-1}\vec{x} = \vec{y} \tag{F.18}$$

is desired, the equation can be rewritten as

$$\vec{x} = \overline{A}\vec{y}. \tag{F.19}$$

With the equation in this form, it is straightforward to solve for the vector $\vec{y}$ by Gaussian elimination [122]. The program uses this algorithm in using the above power series methods.

F.4 Boundary Conditions

The boundary conditions assuming an infinite oxide barrier are imposed by forcing $\zeta_i[0] = 0$ as described above. Boundary conditions assuming a finite oxide barrier were also implemented in the simulation. For a finite oxide barrier, we can write the wave function in the oxide to be the solution to the Schrödinger equation

$$-\frac{\hbar^2}{2m_{ox}}\frac{\partial^2}{\partial z^2}\zeta_i(z) + \Phi_{ox}\zeta_i(z) = E_i\zeta_i(z), \tag{F.20}$$

where Φ_{ox} is the potential-barrier height between the conduction band of the silicon and the oxide conduction band. The wave function in the oxide can be written as

$$\zeta_i(z) = \zeta_i(0)e^{\alpha_i z}, \tag{F.21}$$

where α_i is written as

$$\alpha_i = \sqrt{\frac{2m_{ox}}{\hbar^2}(\Phi_{ox} - E_i)}. \tag{F.22}$$

Note that this expression uses the ground-state energy of a given subband to determine the wave function penetration into the oxide of carriers in that subband. This approximation should introduce only a small error into the simulation. The finite-barrier boundary conditions are implemented by requiring continuity of the value of the wave function and the value of the first derivative of the wave function at $z = 0$. The first derivative of the wave function at the interface can be written as

$$\left.\frac{\partial \zeta_i(z)}{\partial z}\right|_{z=0} = \alpha_i \zeta_i(0). \tag{F.23}$$

Using this expression to implement the finite-barrier boundary conditions in the numerical simulation results in modification of the upper corner of the Hamiltonian as

$$\underline{\overline{H_i}} = -\frac{\hbar^2}{2m_{\perp,i}(du)^2}\begin{bmatrix} -1-(du)\alpha_i & 1 & 0 & \cdots \\ 1 & -2 & 1 & \cdots \\ 0 & 1 & -2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{F.24}$$

F.5 Results

In general, only the lowest 40 subbands are considered when solving the above self-consistent solutions. A negligible amount of carriers lie in the subbands above this level at room temperature. For the simulation results presented in this thesis, the vector size m was 500 elements. The granularity factor du was set to be very close

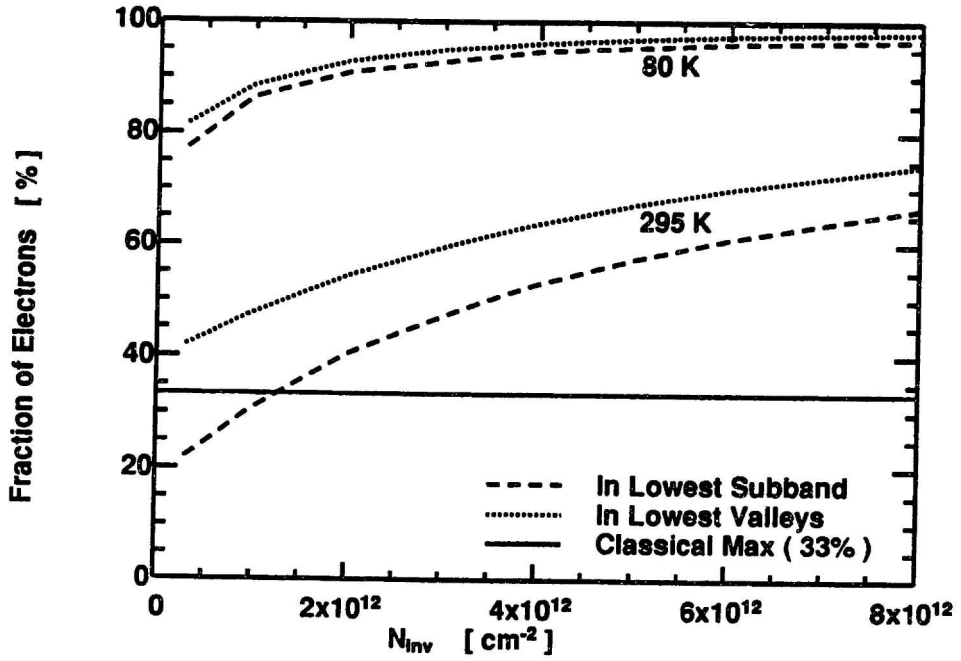
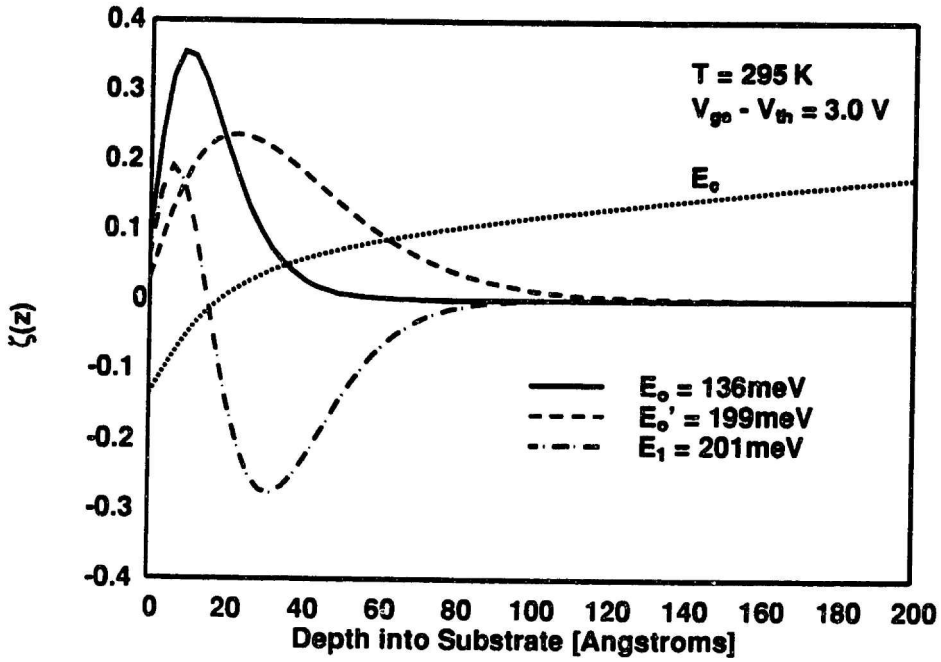


Figure F.2: Fraction of electrons in the lowest two conduction-band valleys and the in the lowest subband as a function of inversion carrier density. Note that there are more carriers in the lowest valleys than expected classically even for very low bias levels. Note that for a 10 nm dielectric, $N_{inv} = 10^{12}$ corresponds roughly to $(V_{gs} - V_{th}) = 1.0$. $T_{ox} = 10$ nm. $N_d = 10^{16}$ cm^{-3}.

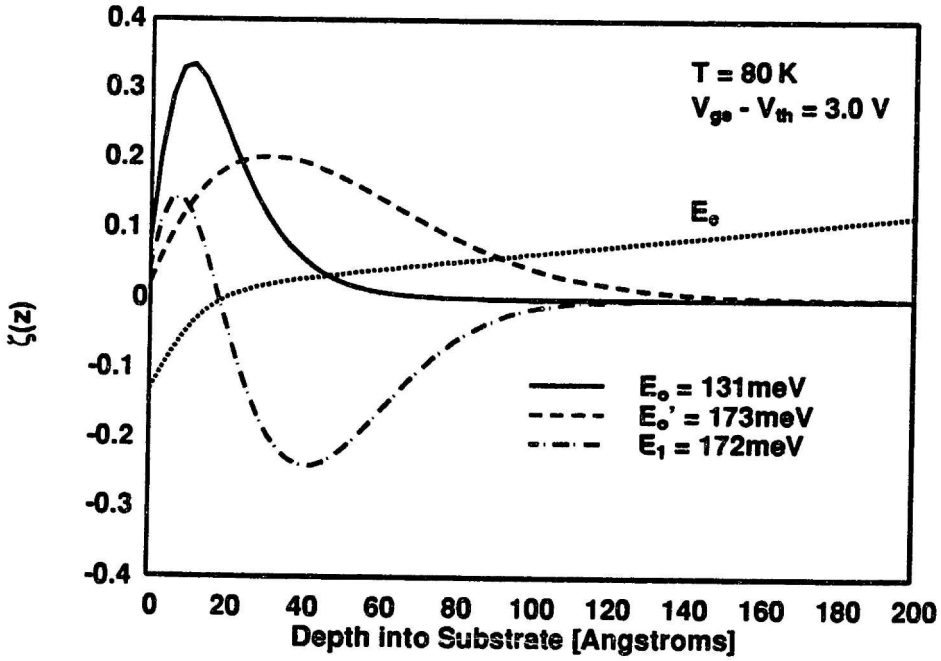
to 2 Å for all the simulations reported. The f factor used to increment the potential was generally set to 0.5 at low bias and was gradually decreased to 0.1 at the highest biases reported, to avoid convergence problems.

The most interesting results of this simulation is that quantum-effects are significant in the inversion layer at room-temperature for even low bias levels. This result is illustrated in Figure F.2 which plots the fraction of electrons in the two lowest conduction-band valleys as a function of inversion layer concentration. Classically, it is expected that one third of the carriers would lie in the lower two conduction band valleys; this result is plotted as the solid line on the figure for reference. This figure is in good qualitative agreement with the inversion layer simulation results given in [80].

Solutions of the quantum-mechanical wave function for two different temperatures in strong inversion are shown in Figure F.3. The value of the wave function is plotted against depth into the substrate. The nominal conduction band is shown for reference; this corresponds to the potential relative to the conduction band in the substrate. The nominal conduction band provides a reference for the shape of the potential well $V(z)$. The effect of the inversion charge on the inversion layer potential is particularly distinct for the plot in Figure F.3b. In this plot, nearly all (96%) of the carriers are in the lowest subband (E_0). The peak in the wave function for this subband corresponds to the point where the potential turns sharply downward.



a) Room temperature.



b) Low temperature.

Figure F.3: Inversion layer wave function at room temperature and near LN_2 temperature. The ground-state subband energies are given in the figure legend. The scale for E_c corresponds to the left axis with units of eV. Note the sharp drop in potential in the inversion layer for the low-temperature case, accentuated because most of the carriers (96%) are very near the interface in the lowest (E_0) subband. $N_d = 10^{16}$ cm^{-3}. $T_{ox} = 10$ nm.

Appendix G

Single Electron Trap Emission

The single electron trap in the dielectric can be modeled as a microstate with ground state energy E_t in thermal contact with the dielectric lattice at temperature T. For such a system, the probability that the trap will be at some excited energy E above the ground state is given classically by a Boltzmann factor [123], so we can write

$$Pr(E) = \frac{1}{kT} e^{-\frac{E-E_t}{kT}}. \tag{G.1}$$

The inversion layer is modeled using the physics of quantized-subbands [80]. In this form, the inversion layer carriers are located in a series of subbands of ground state energy E_n with density of states g_n. Note that with this inversion layer model, the density of states in a given subband is independent of energy, simplifying the calculation of the area density of carriers in a subband, N_n.

Elastic tunneling transitions are allowed from the trap at excited energy E, to the conduction band, and inelastic tunneling transitions are considered to be unlikely. The probability of tunneling from the trap at excited energy E to subband n is proportional to the density of empty states in the subband at energy E, and is also proportional to the wave-function matrix element between the initial trapped state and the final state in the conduction band [124]. For simplicity, this tunneling

probability will be written as

$$Pr(E \rightarrow n) = e_n(x, V_{gs}) \frac{g_n(1 - f(E))}{N_n}, \tag{G.2}$$

where $f(E)$ is the Fermi function of carriers in the silicon, and $e_n(x, V_{gs})$ is a generalized probability factor for a trapped carrier to tunnel to subband n.

Using these results, the emission rate for emission from a trap to subband n is written as the sum of the probability of the trap being excited to a given energy times the emission probability at the energy, i.e.

$$R_n = \int_{E_n}^{\infty} \left[\frac{1}{kT} e^{-\frac{E - E_t}{kT}} \right] e_n(x, V_{gs}) \frac{g_n(1 - f(E))}{N_n} \, dE. \tag{G.3}$$

Making the assumption that the generalized tunnel probability factor $e_n(x, V_{gs})$ varies only slowly with energy allows us to remove it from the integration. Performing the integration and taking the sum over the all subbands, the total emission rate can be written as the inverse of the mean emission time,

$$\left(\frac{1}{<\tau_e>} \right) = \sum_n R_n = e(x, V_{gs}) e^{-\frac{E_f - E_t}{kT}}, \tag{G.4}$$

where E_f is the Fermi level in the silicon. If the difference between the semiconductor Fermi level and the trap energy is large enough, a large activation energy for trap emission is obtained.

It should be noted that a trap at such a low energy should be full nearly all of the time, yet the traps we have characterized go from being nearly always full to nearly always empty depending on the gate bias.